RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2017

Section New Results

Edition: 2018-02-19

COMPUTATIONAL BIOLOGY
1. ABS Project-Team 5
2. AMIBIO Team 8
3. BEAGLE Project-Team 10
4. BIGS Project-Team 14
5. BONSAI Project-Team 16
6. CAPSID Project-Team 18
7. DYLISS Project-Team 20
8. ERABLE Project-Team 23
9. GENSCALE Project-Team 29
10. IBIS Project-Team 34
11. LIFEWARE Project-Team 39
12. MORPHEME Project-Team 43
13. PLEIADE Team 60
14. SERPICO Project-Team 63
15. TAPDANCE Team (section vide) 74
16. VIRTUAL PLANTS Project-Team 75
COMPUTATIONAL NEUROSCIENCE AND MEDICINE
17. ARAMIS Project-Team 92
18. ASCLEPIOS Project-Team 100
19. ATHENA Project-Team 114
20. BIOVISION Team 122
21. CAMIN Team 127
22. GALEN Project-Team 144
23. MATHNEURO Team 153
24. MIMESIS Team 160
25. MNEMOSYNE Project-Team 166
26. NEUROSYS Project-Team 168
27. PARIETAL Project-Team 171
28. VISAGES Project-Team 179
EARTH, ENVIRONMENTAL AND ENERGY SCIENCES
29. AIRSEA Project-Team 187
30. ANGE Project-Team 195
31. CASTOR Project-Team 201
32. COFFEE Project-Team 207
33. FLUMINANCE Project-Team 208
34. LEMON Team 215
35. MAGIQUE-3D Project-Team 217
36. SERENA Project-Team 228
37. STEEP Project-Team 233
38. TONUS Team 235

MODELING AND CONTROL FOR LIFE SCIENCES

39. BIOCORE Project-Team 239
40. CARMEN Project-Team 247
41. DRACULA Project-Team 250
42. M3DISIM Project-Team 255
43. MAMBA Project-Team 263
44. MONC Project-Team 265
45. MYCENAE Project-Team 268
46. NUMED Project-Team (section vide) 271
47. REO Project-Team 272
48. SISTM Project-Team 275
49. XPOP Project-Team 279

ABS Project-Team

5. New Results

5.1. Modeling interfaces and contacts

Keywords: docking, scoring, interfaces, protein complexes, Voronoi diagrams, arrangements of balls.

5.1.1. Novel structural parameters of Ig-Ag complexes yield a quantitative description of interaction specificity and binding affinity

Participants: F. Cazals, S. Marillet.

In collaboration with P. Boudinot (INRA Jouy-en-Josas) and M-P. Lefranc (University of Montpellier 2).

Antibody-antigen complexes challenge our understanding, as analyses to date failed to unveil the key determinants of binding affinity and interaction specificity. In this work [17], we partially fill this gap based on novel quantitative analyses using two standardized databases, the IMGT/3Dstructure-DB and the structure affinity benchmark.

First, we introduce a statistical analysis of interfaces which enables the classification of ligand types (protein, peptide, chemical; cross-validated classification error of 9.6%), and yield binding affinity predictions of unprecedented accuracy (median absolute error of 0.878 kcal/mol). Second, we exploit the contributions made by CDRs in terms of position at the interface and atomic packing properties to show that in general, VH CDR3 and VL CDR3 make dominant contributions to the binding affinity, a fact also shown to be consistent with the enthalpy - entropy compensation associated with pre-configuration of CDR3. Our work suggests that the affinity prediction problem could be solved from databases of high resolution crystal structures of complexes with known affinity.

5.1.2. Anti-interleukin-6 signalling therapy rebalances the disrupted cytokine production of B cells from patients with active rheumatoid arthritis

Participants: F. Cazals, A. Lhéritier.

In collaboration with S. Fleischer (1. Charité University Medicine Berlin, Berlin, Germany), S. Ries (2. Deutsches Rheuma-Forschungszentrum Berlin, Berlin, Germany), P. Shen (2.), G.R. Burmester (1.), T. Dörner (1.), S. Fillatreau (2., Institut Necker-Enfants Malades, Université Paris Descartes, IHP Hôpital Necker Enfants Malades).

Rheumatoid arthritis (RA) is associated with abnormal B cell-functions implicating antibody-dependent and -independent mechanisms. B cells have emerged as important cytokine-producing cells, and cytokines are well-known drivers of RA pathogenesis. To identify novel cytokine-mediated B-cell functions in RA, in this work [16], we comprehensively analysed the capacity of B cells from RA patients with an inadequate response to disease modifying anti-rheumatic drugs to produce cytokines in comparison with healthy donors (HD). RA B cells displayed a constitutively higher production of the pathogenic factors interleukin (IL)-8 and Gro-α, while their production of several cytokines upon activation via the B cell receptor for antigen (BCR) was broadly suppressed, including a loss of the expression of the protective factor TRAIL, compared to HD B cells. These defects were partly erased after treatment with the IL-6-signalling inhibitor tocilizumab, indicating that abnormal IL-6 signalling contributed to these abnormalities. Noteworthy, the clinical response of individual patients to tocilizumab therapy could be predicted using the amounts of MIP-1β and β-NGF produced by these patients' B cells before treatment. Taken together, our study highlights hitherto unknown abnormal B-cell functions in RA patients, which are related to the unbalanced cytokine network, and are potentially relevant for RA pathogenesis and treatment.

5.2. Modeling macro-molecular assemblies

Keywords: macro-molecular assembly, reconstruction by data integration, proteomics, mass spectrometry, modeling with uncertainties, connectivity inference.

5.2.1. Complexity dichotomies for the minimum F-overlay problem

Participants: D. Mazauric, R. Watrigant.

In collaboration with N. Cohen (LRI, UMR de l'Université Paris-Sud et du CNRS), F. Havet (Université Côte d'Azur, I3S, UMR de l'Université Nice Sophia et du CNRS), I. Sau (LIRMM, UMR de l'Université Montpellier et du CNRS, and Universidade Federal do Ceará, Brazil).

The *connectivity inference* problem for native mass spectrometry aims at finding the most plausible pairwise contacts between the individual subunits of a macro-molecular assembly, given the composition of overlapping oligomers. The associated combinatorial optimization problem consists in determining a minimal-cardinality set of contact (edges) such that all the subunits of each oligomer must be "connected" (each oligomer must induce a connected graph). We studied in [18] the general inference problem that consists of considering more general properties on oligomers. For this new problem, we are given a list of possible topologies (graphs) for each oligomer and we aim at minimizing the total number of contacts between subunits. In terms of graphs, we are given a family of subgraphs that can match the structure of the oligomers. These new constraints reflect biophysical properties: a subunit has a limited number of neighbors (bounded maximum degree of the subgraphs), selected contacts are already known (a given subgraph contained in the complex), etc. We prove that the problem is NP-complete (no polynomial time algorithm, unless P = NP) for almost all cases.

5.3. Modeling the flexibility of macro-molecules

Keywords: protein, flexibility, collective coordinate, conformational sampling dimensionality reduction.

No new result on this topic in 2017.

5.4. Algorithmic foundations

Keywords: Computational geometry, computational topology, optimization, data analysis.

Making a stride towards a better understanding of the biophysical questions discussed in the previous sections requires various methodological developments discussed below.

5.4.1. Extracting the groupwise core structural connectivity network: bridging statistical and graph-theoretical approaches

Participant: D. Mazauric.

In collaboration with N. Lascano (Universidad de Buenos Aires, Argentina, Université Côte d'Azur, and Inria Sophia Antipolis - Méditerranée, EPI ATHENA), G. Gallardo (2. Université Côte d'Azur and Inria Sophia Antipolis - Méditerranée, EPI ATHENA), D. Wassermann (2).

Finding the common structural brain connectivity network for a given population is an open problem, crucial for current neuro-science. Recent evidence suggests there is a tightly connected network shared between humans. Obtaining this network will, among many advantages, allow us to focus cognitive and clinical analyses on common connections, thus increasing their statistical power. In turn, knowledge about the common network will facilitate novel analyses to understand the structure-function relationship in the brain. In [19], we present a new algorithm for computing the core structural connectivity network of a subject sample combining graph theory and statistics. Our algorithm works in accordance with novel evidence on brain topology. We analyze the problem theoretically and prove its complexity. Using 309 subjects, we show its advantages when used as a feature selection for connectivity analysis on populations, outperforming the current approaches.

5.4.2. Maximum flow under proportional delay constraint

Participant: D. Mazauric.

In collaboration with P. Bonami (LIF, UMR d'Aix-Marseille Université et du CNRS, and IBM ILOG CPLEX, Madrid), Y. Vaxès (LIF, UMR d'Aix-Marseille Université et du CNRS).

Network operators must satisfy some Quality of Service requirements for their clients. One of the most important parameters in telecommunication networks is the end-to-end delay of a unit of flow between a source node and a destination node. Given a network and a set of source destination pairs (connections), we consider in [14] the problem of maximizing the sum of the flow under proportional delay constraints. In this paper, the delay for crossing a link is proportional to the total flow crossing this link. If a connection supports non-zero flow, then the sum of the delays along any path corresponding to that connection must be lower than a given bound. The constraints of delay are on-off constraints because if a connection carries zero flow, then there is no constraint for that connection. The difficulty of the problem comes from the choice of the connections supporting non-zero flow. We first prove a general approximation ratio using linear programming for a variant of the problem. We then prove a linear time 2-approximation algorithm when the network is a path. We finally show a Polynomial Time Approximation Scheme when the graph of intersections of the paths has bounded treewidth.

5.4.3. Comparing two clusterings using matchings between clusters of clusters

Participants: F. Cazals, D. Mazauric, R. Tetley, R. Watrigant.

Clustering is a fundamental problem in data science, yet, the variety of clustering methods and their sensitivity to parameters make clustering hard. To analyze the stability of a given clustering algorithm while varying its parameters, and to compare clusters yielded by different algorithms, several comparison schemes based on matchings, information theory and various indices (Rand, Jaccard) have been developed. In this work [20], we go beyond these by providing a novel class of methods computing meta-clusters within each clustering– a meta-cluster is a group of clusters, together with a matching between these. Let the intersection graph of two clusterings be the edge-weighted bipartite graph in which the nodes represent the clusters, the edges represent the non empty intersection between two clusters, and the weight of an edge is the number of common items. We introduce the so-called D-family-matching problem on intersection graphs, with D the upper-bound on the diameter of the graph induced by the clusters of any meta-cluster. First we prove NP-completeness results and unbounded approximation ratio of simple strategies. Second, we design exact polynomial time dynamic programming algorithms for some classes of graphs (in particular trees). Then, we prove spanning-tree based efficient algorithms for general graphs. Our experiments illustrate the role of D as a scale parameter providing information on the relationship between clusters within a clustering and in-between two clusterings. They also show the advantages of our built-in mapping over classical cluster comparison measures such as the variation of information (VI).

5.4.4. The SBL

Participants: F. Cazals, T. Dreyfus.

Software in structural bioinformatics has mainly been application driven. To favor practitioners seeking off-the-shelf applications, but also developers seeking advanced building blocks to develop novel applications, we undertook the design of the Structural Bioinformatics Library (SBL), a generic C++/python cross-platform software library targeting complex problems in structural bioinformatics. Its tenet is based on a modular design offering a rich and versatile framework allowing the development of novel applications requiring well specified complex operations, without compromising robustness and performances.

The SBL involves four software components (1–4 thereafter) [15]. For end-users, the SBL provides ready to use, state-of-the-art (1) applications to handle molecular models defined by unions of balls, to deal with molecular flexibility, to model macro-molecular assemblies. These applications can also be combined to tackle integrated analysis problems. For developers, the SBL provides a broad C++ toolbox with modular design, involving core (2) algorithms, (3) biophysical models and (4) modules, the latter being especially suited to develop novel applications. The SBL comes with a thorough documentation consisting of user and reference manuals, and a bugzilla platform to handle community feedback.

The SBL is available from http://sbl.inria.fr

See also the section **New Software and Platforms**.

AMIBIO Team

5. New Results

5.1. New circular RNAs identified in Pyrococcus abyssi

We contributed a new method for the detection of circRNAs, which we validated on simulated data, and used to analyze the transcriptome of *Pyrococcus abyssi*, an archae living at high depth and temperature [1]. Using this method, which was shown to produce less false positives than previous computational approaches, we analyzed data produced in collaboration with LOB (Ecole Polytechnique), and detected roughly a hundred of novel candidates circular RNAs. Moreover, we provided evidence, on a large scale, that the protein *Pab1020* acts as a ligase, and interacts with some of these circular RNAs, shedding new light on the mechanisms underlying the circularization process.

5.2. Minimal absent words

Minimal absent words are words that do not occur but whose proper factors all occur in the sequence. In a collaboration with King's College, several algorithms, we have designed algorithms to search for minimal absent words in external memory [8], and *in-line*, using a sliding window [13] (parallelization, external memory,...) that outperform previous solutions and achieve near-optimal speed up. This opens new scenarios in the applications of minimal absent words in computational biology, including phylogeny or evolution. For instance, it was shown that there exist three minimal words in Ebola virus genomes which are absent from human genome. As two strings coincide iff they have the same set of minimal absent words, an interesting side result is to solve in optimal time the pattern matching problem using *negative information*.

5.3. Kinematics-inspired algorithms for macromolecular modeling

At a geometric level, RNA is much more flexible than protein, and undergoes smooth transitions between its various conformations. Such transitions are difficult to observe, but can be predicted using algorithms inspired by kinematics and motion-planning. With our partners at Stanford, we designed and implemented such an algorithm within the KGS library [8] to morph between two RNA conformations while keeping distance constraints induced by base pairs and, more importantly, avoiding clashes. In a more preliminary work, we also used similar approaches to automagically fit multi-conformer ligand models into electron density maps [16].

5.4. RNA design

In a paper published in Algorithmica [6], we have shown that our previous results [30] hold for more sophisticated energy models where base-pairs are associated with arbitrary energy contributions. This result, which required a complete overhaul of our previous proofs (e.g. using arguments based on graph coloring), allows us to foresee an extension of (at least some of) our results to state-of-the-art models, such as the Turner energy model.

In collaboration with Danny Barash's group at Ben-Gurion university (Israel), we contributed a review of existing tools and techniques for RNA design, which was published in *Briefings in Bioinformatics* series [3].

Finally, in a paper [14] recently accepted for a presentation at the prestigious RECOMB'18 conference, we revisited the problem of generating at random an RNA sequence which is simultaneously compatible with a set of target secondary structures. This problem was previously addressed by our collaborators at the TBI Vienna/Univ. Leipzig, using an exponential-time algorithm. We established the #*P*-hardness of the problem, and its inapproximibility in general. However, the problem is still amenable to an efficient parametrization, and we proposes an FTP algorithm named `RNARedPrint` based on the tree decomposition for the random generation, to which we adapted a multidimensionnal Boltzmann sampling technique in order to gain (probabilistic) control over secondary features such as the *GC%*, the relative free-energy of the various structures...

5.5. Game theory and macromolecular modeling

Initially based on a very coarse representation of RNA, we refined our model of RNA folding as a game, using on-lattice coordinates and statistical potentials for the utility function. The resulting algorithm was implemented in the subsequent version of the `GARN` [2] software.

The final year of Amélie Héliou's PhD led to theoretical developments in game theory, mainly obtained in collaboration with J. Cohen (LRI, Univ. Paris-Sud). First, the quasi-exponential convergence, under reasonable assumptions, of the HEDGE algorithm was demonstrated in collaboration with the POLARIS team in Grenoble [11]. Moreover, in a paper accepted at NIPS'17 [12], we addressed the learning of Nash equilibria. In this context, we established the convergence with high probability of no-regret learning in the bandit and semi-bandit settings.

5.6. RNA kinetics using non-redundant sampling

RNA kinetics is arguably the next frontier in RNA 2D bioinformatics. In particular, computational methods for studying the kinetics of RNA beyond 150nts are hindered by the combinatorial explosion of the conformation space. In an effort to circumvent such an effect, we have proposed a sampling approach that explicitly target local minima of the energy function. Our sampling algorithm, jointly proposed with H. Touzet (Bonsai, Inria Lille & CrisTaL, Univ. Lille I) and accepted for a presentation at the ISMB/ECCB'17 conference in Prague [15], uses non-redundant sampling principles to avoid an excessive concentration of samples within low local minima.

5.7. New insight from SHAPE probing data

Existing computational methods for structure prediction are typically hindered by their assumption of a single structure, and their assumption of orthogonal signals stemming from different reagents. To overcome these limitations, we contributed an integrative approach combining stochastic sampling and structural clustering [17] (journal version pending). In collaboration with ENS Lyon/Univ. Lyon I and Univ. Paris-Descartes, we used this method to model the structure of the HIV-1 gag open-reading frame [4].

We also addressed the problem of binning sets of NGS reads arising from the simultaneous probing, using the SHAPEmap protocol, of variants produced by a error-prone PCR. We proposed a variant of the Expectation-Maximization algorithm [10] to jointly infer maximum-likelihood origins for reads and mutational profiles for each variant.

BEAGLE Project-Team

6. New Results

6.1. The fossil record of microbes augmented by an order of magnitude

Participants: Eric Tannier
Biodiversity has always been predominantly microbial and the scarcity of fossils from bacteria, archaea and microbial eukaryotes has prevented a comprehensive dating of the tree of life. We have shown that patterns of lateral gene transfer deduced from the analysis of modern genomes encode a novel and abundant source of information about the temporal coexistence of lineages throughout the history of life. We constructed and used new phylogenetic methods to reconstruct the history of thousands of gene families and demonstrate that dates implied by gene transfers are consistent with estimates from relaxed molecular clocks in bacteria, archaea and eukaryotes. An inspection of discrepancies between transfers and clocks and a comparison with mammal fossils show that gene transfer in microbes is potentially as informative for dating the tree of life as the geological record in macroorganisms.
Main publications: [30]

6.2. Phylogenetics of dependence, and dependence of phylogenies

Participants: Wandrille Duchemin, Eric Tannier
Standard phylogenetics use DNA or protein sequences, along with probabilistic models of substitutions, which are Markov processes on trees. The big default of this methodology is to assume a common evolution of all sites inside a gene, and a total independence with other genes. This model does not capture the essence of living things, which is made of dependencies and interactions. We made several methodological developments to take into account these dependencies, by improving the gene tree species tree reconciliation methods and reconstructing phylogenies of relations between genes.
Publications: [16], [32], [19]

6.3. Beware batch culture: Seasonality and niche construction predicted to favor bacterial adaptive diversification

Participants: Charles Rocabert, Carole Knibbe, Guillaume Beslon
The evolution of stable bacterial cross-feeding interactions is often considered as the first step toward bacterial speciation in sympatry. It is thus important to study the conditions favoring the emergence and the stabilization of cross-feeding interactions in well-mixed environments. Experimental evolution in laboratory, where fast organisms are replicated for thousands of generations in controlled conditions, provides important insights on this question. Indeed cross-feeding is commonly observed in batch cultures or in chemostat. However, the reasons why cross-feeding interactions become stable and lead to monophyletic ecotypes remain unclear. Because laboratory experiments are a long and costly process, we explored this question by evolving digital organisms in artificial systems mimicking the conditions of wet experiments.

Models of digital evolution helped a lot to decipher the evolution of cross-feeding interactions. However, the evolution of real microorganisms implies the interaction of a wide range of biological structures and levels, while those models often include only two or three levels, limiting their ability to mimic real experiments. In this work, we developed a new multi-scale model of digital evolution, integrating a complex and realistic genotype-to-phenotype mapping (including a metabolic network) and a complex environment (that links the organism's metabolic networks together, opening the possibility for cross-feeding). This model has been developed under the European project EvoEvo, and has been inspired by previous models developed by the Beagle team, and in the Theoretical Biology and Bioinformatics group at Utrecht University. By mimicking laboratory experiment setups and running simulations for tens of thousands of generations, we were able to recover ecological dynamics similar to those found in real experiments.

In batch culture, like in the Long Term Evolution Experiment (LTEE), it is accepted that the seasonality generated by the serial transfers triggers the maintenance of cross-feeding interactions on the long-term by favoring niche construction and specialization. In chemostat, cross-feeding interactions are observed and seem stable for a few hundreds of generations but the reasons of their stability remain unclear. Thanks to our model, we were able to observe stable cross-feeding interactions reproducing the same properties as those observed in the LTEE. We then showed that seasonal conditions found in batch cultures are essential for the maintenance of stable ecotypes on the long-term, since it produces conditions for niche construction and stable cross-feeding. In chemostat conditions, the absence of seasonality and competitive exclusion precludes any stabilization of emerging cross-feeding interactions. Finally, we proposed to consider a cross-feeding interaction to be stable only if interacting ecotypes undergo independent periodic selection events on the long-term. Stable cross-feeding interactions could then be considered as premises to speciation in sympatry.

This work is the result of an enriching collaboration between the Beagle team (Charles Rocabert, Carole Knibbe, Guillaume Beslon), and microbiologists from the TIMC-IMAG in Grenoble (Jessika Consuegra, Dominique Schneider). It has been published in the renowned journal PLoS Computational Biology in January 2017. This work is of interest for the fields of evolutionary biology, microbiology but also for computer science. Indeed our findings also suggest that digital evolution is a useful tool to study bacterial evolution, and that the use of models integrating a complex genotype-to-phenotype mapping and complex interactions between digital organisms and their environment is important to accurately study real biological systems thus appealing for further fruitful transdisciplinary collaborations.
Publication: [21].

6.4. Evolution of phenotypic noise

Participants: Charles Rocabert, Guillaume Beslon, Carole Knibbe
The phenotype of an organism is a complex non-linear cascade of developmental, physiological and regulatory processes, formalized by the concept of genotype-to-phenotype map. An increasing number of experimental studies demonstrate the existence of phenotypic noise, which can be finely tuned by the genotype-to-phenotype map, and that phenotypic noise can be adaptive.

In stabilizing selection, when the population is at a fitness optimum, phenotypic noise is deleterious and minimized by evolution. Nevertheless, phenotypic noise can be positively selected when the population is exposed to stressful conditions. It was thus suggested that during an adaptation event, phenotypic noise would increase in directional evolution, and then be reduced when the selection becomes stabilizing. In 1930, R.A. Fisher suggested with its so-called Fisher's Geometric Model (FGM) that organisms adapting to a new environment experience a "cost of complexity", where beneficial mutations become increasingly harder to fix when the number of phenotypic characters increases. Predictions made on the evolution of phenotypic noise are mostly based on single trait observations. Is there also a cost of complexity on the phenotypic noise?

To address this question, we extended the FGM by adding an evolvable phenotypic noise. First, using a simple form of noise, affecting similarly every phenotypic character, we show that a cost of complexity indeed makes phenotypic noise deleterious in directional evolution. Second, we extended the FGM with a fully evolvable noise, allowing evolution on noise amplitudes on each character, as well as on noise correlations between characters. In directional evolution, we show that phenotypic noise evolves towards a flattened shape, with elevated noise in the direction of the optimum, and minimized noise in all other directions. In this case, the noise becomes advantageous again, even with many characters. Non-isotropic phenotypic noise thus facilitates evolution towards the fitness optimum, and significantly reduces the cost of complexity. Our results show that such non-isotropic phenotypic noise could be exploited by evolution, and suggest further experiments to assess the functional nature of phenotypic noise.

This result is currently under review for the Evolution journal. It is the result of an enriching collaboration between the Beagle team (Charles Rocabert, Guillaume Beslon, Carole Knibbe), and the Dracula team (Samuel Bernard). Although the results are grounded in theory and mathematical modeling, they provide stringent conditions for noise to be beneficial, which are experimentally testable. We believe the results to be of wide interest for researchers working on phenotypic evolution. By deciphering the conditions in which phenotypic noise evolves towards specific patterns, our work may also contribute to a better understanding of drug resistance and cancer cells proliferation, and also to the growing field of predictive biology.

6.5. Impact of group size and social composition on group vocal activity and acoustic network in a social songbird

Participants: Marie Fernandez, Hédi Soula
In social species individuals living in the same group may synchronize activities such as movements, foraging or antipredator vigilance. Vocalizations are behaviors that can be coordinated between individuals, but simultaneous vocalizations in groups have mostly been considered as noise that does not bear any information. Indeed, little is known about the structure and function of vocal communications involving a network of individuals. Zebra finches, Taeniopygia guttata, are social, monogamous songbirds that form lifelong pair bonds. In the wild, they are typically found in small groups and they gather in 'social' trees where they produce vocalizations. Here we investigated in the laboratory the influence of group size and composition on general vocal activity and synchrony, as well as the influence of pair bond and spatial location on the finer characteristics of dyads' vocal interactions. We used a set-up that locked the birds at fixed spatial positions of our choosing to control the proximity network and allowed us to match most of the vocalizations with specific individuals. We used an in-house software suite that automatically detects vocalizations from hours of passive recording. We found that zebra finch groups synchronized their general vocal activity with waves of collective vocalizations, which depended on both the size and the composition of the group. The acoustic network was shaped by pair bonds at different timescales.
Publication: [17].

6.6. Dopamine-endocannabinoid interactions mediate spike-timing dependent potentiation in the striatum

Participants: Hugues Berry, Alexandre Foncelle, Ilya Prokin
Synaptic long-term plasticity underlies multiple forms of learning and memory in the brain. In most systems, its bidirectionality - depression (LTD) and potentiation (LTP) allows adaptive adjustment of the synaptic weight depending on the activity. Endocannabinoids (eCBs), one of the most widespread neurotransmitter systems, are very well established as depressing neuronal communication but recent experimental evidence challenges this depression-only vision. Our previous work in collaboration with L. Venance's lab at CIRB, College de France, Paris (experimental neuroscience) has combined experimental and mathematical modeling approaches to identify in the basal ganglia the existence of an eCB-mediated spike-timing dependent LTP (eCB-tLTP) induced by a low number of paired stimulations [50], [49]. However, the regulation and control mechanisms of eCB-tLTP remained unknown. Using the same combination of experimental and modelling approaches, we have now discovered that dopamine controls eCB-tLTP. The dopamine system is a key actor of associative learning in the basal ganglia and a pivotal system in several pathologies including Parkinson's disease. We identified that eCB-tLTP depends on dopamine and involves the activation of D2R dopamine receptors located presynaptically in corticostriatal glutamatergic afferents. We moreover show that dopamine control of eCB-tLTP is of pathological significance since it is impaired in a rodent model of Parkinson's disease and rescued by chronic L-DOPA treatment in those animals. Combining our experimental finding with a realistic mathematical model of the underlying signaling pathway, we could describe the mechanisms accounting for this endocannabinoid-dopamine regulation of eCB-tLTP.
This result is currently under review in the journal Nature Communications.

6.7. Subspace clustering based on medians using evolutionary algorithms

Participants: Sergio Peignier, Christophe Rigotti, Jonas Abernot, Guillaume Beslon

Subspace clustering is a data mining task that searches for objects sharing similar features, and at the same time looks for the subspaces where these similarities appear. For this reason subspace clustering is recognized as more general and complicated than standard clustering, since it needs to detect these relevant subspaces. Taking advantage of the expertise of the team in evolution in silico, we previously showed that evolutionary algorithms are promising approaches to address this problem. Another important clustering task is the K-medians one, where objects are grouped around medians, leading to cluster centers more robust to noise and outliers. In order to take advantage of these benefits within the subspace clustering process itself, we developed a new evolutionary algorithm, KymeroClust, that builds cluster centers that are medians in subspaces. This algorithm takes advantage of an evolvable representation of the genotypes to adapt the numbers of clusters produced and the subspace dimensionalities. It is based on new bio-inspired mutation operators to evolve the cluster centers as medians and is able to handle streaming data. KymeroClust has been compared to the main subspace clustering methods and turns out to be very competitive both in terms of cluster quality and runtime, while requiring an easier parameter setting.

Publications: [12], [24]

BIGS Project-Team

6. New Results

6.1. Stochastic modelling

Participants: T. Bastogne, P. Vallois, S. Wantz-Mezieres, L. Batista, A. Gégout-Petit

Because of the observation of longitudinal data for each subject in medicine, we have to care about the random effect due to the subject and to choose adapted models like mixed effect models [39], [40]. We recently improved this methodology for the analysis of data collected in vivo for growth tumor for the biopharmaceutical company Transgene. The problem was to measure the differential effect of treatments (different molecules and doses) on the dynamics of the tumor taking into account the effect of censoring [10].

In the framework of the esca-illness of vines, we developed different spatial models and spatio-temporal models for different purposes: (1) study the distribution and the dynamics of esca vines in order to tackle the aggregation and the potential spread of the illness (2) propose a spatio-temporal model in order to capture the dynamics of cases and measure the effects of environmental covariates. For purpose (1), we propose different tests based on the join count statistics [6].

6.2. Estimation and control for Markov Processes

Participants: R. Azais, F. Bouguet, T. Bastogne

We have developed statistical inference techniques for estimating the jump rate of PDMPs (piecewise-deterministic Markov processes) [2] which is an essential step to build relevant application models. In [2], we state a new characterization of the jump rate when the transition kernel only charges a discrete subset of the state space and deduce from it a competitive nonparametric technique for estimating this feature of interest. Our methodologies have been illustrated on numerical examples and real data. We also investigated the probabilistic properties of the PDMPs [5] or more general Markov processes [31] that could be useful to study properties of estimators.

A bit more generally, we have made contributions to a variety of specific estimation problems. We considered the problem of estimation of integrals under Markov design, which has a large variety of applications, in particular in biology and climatology. In [24], we have developed and analyzed a technique for estimating the average value over space when sensors describe a Markovian trajectory; this method leads to rates that are better than the traditional "root n"-rate, where n is the sample size, and was applied to the evaluation of the average temperature of oceans.

Control of stochastic processes is also a way to optimise administration (dose, frequency) of therapy. In [8], we have presented the design and validation of a real time controller able to track a preset photobleaching trajectory by modulating the width of light impulses during the treatment sessions, which is useful in a Photodynamic therapy context. This innovative solution was validated by *in vivo* experiments that have shown a significantly improvement of reproducibility of the inter-individual photobleaching kinetics. This innovative controller is the first personalized solution able to adapt in realtime the dose of light to be applied in photodynamic therapy.

6.3. Algorithms and Estimation for graph data

Participants: R. Azais, F. Bouguet, T. Bastogne

Tree-structured data naturally appear in various fields, particularly in biology where plants and blood vessels may be described by trees. The paper [27] is devoted to the estimation of the relative scale of ordered trees that share the same layout. The theoretical study is achieved for the stochastic model of conditioned Galton-Watson trees. New estimators are introduced and their consistency is stated. A comparison is made with an existing approach of the literature. A simulation study shows the good behavior of our procedure on finite-sample sizes.

6.4. Regression and machine learning

Participants: A. Gégout-Petit, A. Muller-Gueudin, T. Bastogne, L. Batista, R. Azais, S. Ferrigno, K. Duarte, J.-M. Monnez

We consider the problem of sequential least square multidimensional linear regression using a stochastic approximation process. The choice of the stepsize may be crucial in this type of process. In order to avoid the risk of numerical explosion which can be encountered, we define three processes with a variable or a constant stepsize and establish their convergence. Finally these processes are compared to classic processes on 11 datasets, 6 with a continuous output and 5 with a binary output, for a fixed total number of observations used and then for a fixed processing time. It appears that the third-defined process with a very simple choice of the stepsize gives usually the best results [32].

We study many other regression models like survival analysis, spatio temporal models with covariates. Among the multiple regression models, we want to test, thanks to simulation methods, validity of their assumptions [25]. Tests of this kind are called omnibus test. An omnibus test is an overall test that examines several assumptions together, the most known omnibus test is the one for testing gaussianity (that examines both skewness and kurtosis).

In the purpose of selecting factors linked to the efficiency of a treatment in the context of high dimension (about 100.000 covariates), we have developed a new methodology to select and rank covariates associated to a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology imbricates successively rough selection, clustering of variables, decorrelation of variables using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking through bootstrap replications. Simulations study shows the interest of the decorrelation inside the different clusters of covariates. The methodology was applied to select covariates among genomics, proteomics covariates linked to the success of a immunotherapy treatment for lung cancer [21], [19], [20].

We also focus on the biological context of high-throughput and high-content bioassays in which several hundreds or thousands of biological signals are measured for a posterior analysis. In this experimental context, each culture well is a biological system in which the output variable is the cell proliferation, the input variable can be an electrical or a light stimulus signal and the covariate may be the type of cells, type of medium or tested compounds. The ambition is to identify a batch of several thousands of wells in a single step with the same model structure. Mixed effects models are largely used in regression but up to now they have rarely been used in the field of dynamical system identification. Our approach aims at developing a new solution based on an ARX (Auto Regressive model with eXternal inputs) model structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameter [13], [10].

BONSAI Project-Team

7. New Results

7.1. Metagenomics

Reconstruction of phylogenetic marker genes. Accurate identification of organisms present within a community is essential to understanding the structure of an ecosystem. However, current HTS technologies generate short reads, such as Illumina reads, which makes it a difficult task. One possibility is to focus on assembly of taxonomic markers of interest, such as 16S ribosomal RNA. The PhD thesis of P. Pericard proposed an algorithm that is specifically dedicated to this problem. The method implements a stepwise process based on construction and analysis of a read overlap graph, which is built using read alignments (produced by SortMeRNA) and is decomposed into relevant connected components extracted from a compressed representation of the graph. It is able to recover full length 16S sequences with high precision assemblies ($\leq 0.1\%$ error rate). This work is published in the reference journal in the field [23] and the resulting software, MATAM, was released this spring. It is currently being tested in several labs [0]. This work received the Best Oral Presentation Award from the SFBI [0] this year [29].

Metagenomics assembly. Another important task that could help taxonomic assignment is to reconstruct uncultured microbial strains and species for which the genome sequence is fully unknown. To this end, metagenomics mainly borrows techniques from classical genomics, i.e. from *de novo* assembly of isolate genomes. We built upon continuous methodological advances with our genomic assembler Minia, adding new data structures such as the minimal perfect hash function [26] and the compressed graph representation. We participated in 2015 in the CAMI metagenomic reconstruction challenge [0]. This challenged gathered a total of 17 international groups, and Minia performed among the top assembly methods. This result is reported in an article to appear in Nature Methods from the CAMI consortium [25]. We further presented a poster at RECOMB 2017.

Targeted metagenomics. Within the PhD thesis of L. Siegwald, we have participated to the design of a comprehensive evaluation protocol to compare computational pipelines to analyze 16S amplicons, and have studied the impact of different variables on the biological interpretation of results. This study included the following tools: CLARK, Kraken, Mothur, Qiime and One Codex. It has been the subject of an invited keynote at the international workshop Recent Computational Advances in Metagenomics (RCAM 2017) [0].

7.2. Nonribosomal peptides

We further investigate the NRPs produced by *Burkholderia*, focusing on the identification of new compounds implicated in biocontrol and pharmaceutical [19].

New functionalities have been added to Norine to query the SMILES field either by the query form and the REST service. We also continue our curating of Norine data by improving peptide annotations and validation submissions of new peptides.

7.3. High-throughput V(D)J repertoire analysis

Researches on high-throughput V(D)J repertoire analysis started in the group in 2012. We have developed Vidjil, a web platform dedicated to the analysis of lymphocyte populations. Starting from DNA sequences, uploaded by the user, Vidjil identifies and quantifies lymphocyte populations and provides an interactive visualization.

[0]Tests of MATAM at MEDIS (INRA-Université Clermont Auvergne) for gene capture, Labgem (Genoscope) where it is on tracks to be integrated into the PathoTRACK-MicroScope platform dedicated to the human intestinal microbiome, and the Australian Centre for Ancient DNA (University of Adelaïde) for oral microbiome research.

[0]SFBI: Société Française de Bioinformatique

[0]CAMI challenge: https://data.cami-challenge.org/

[0]RCAM 2017: http://maiage.jouy.inra.fr/?q=fr/rcam2017

Seven European hospitals are now using Vidjil for their daily clinical practice. This year we published our experience of the minimal residual disease follow-up for acute lymphoblastic leukemia using Vidjil [24]. This is a first step towards using high-throughput sequencing and Vidjil for all the follow-up of the patients. We also participated to a joint publication with the EuroClonality-NGS consortium (see below).

Finally, we are working on transferring activities on platform development and user support. After meetings with several partners, we selected the Inria Foundation. The VidjilNet consortium (http://www.vidjil.net) will be launched in January 2018 within the InriaSoft action of the Foundation and will hire two engineers. VidjilNet will first gather hematology labs of French hospitals working on diagnosis and follow-up of acute lymphoblastic leukemia, and will be then extended to labs working on other pathologies as well as foreign labs.

7.4. RNA-Seq software benchmarking

Plenty of methods have been devised to analyze RNA-Seq data. Due to this large choice, it is a difficult task to determine what software is the best suited for a given question. To help in solving this problem, with colleagues at IRMB and in the SeqOne start-up in Montpellier, we devised a flexible benchmarking pipeline [18].

This pipeline is intended to be flexible enough to deal either with simulated or real data and to evaluate software on many possible aspects (mapping, splice detection, fusion detection, variant calling, and also on the post-analysis aspects such as gene quantification).

7.5. RNA folding landscape

Kinetics is key to understand many phenomena involving RNAs, such as co-transcriptional folding and riboswitches. Exact out-of-equilibrium studies induce extreme computational demands, leading state-of-the-art methods to rely on approximated kinetics landscapes, obtained using sampling strategies that strive to generate the key landmarks of the landscape topology. However, such methods are impeded by a large level of redundancy within sampled sets. Such a redundancy is uninformative, and obfuscates important intermediate states, leading to an incomplete vision of RNA dynamics.

Within the context of ANR RNAlands, we introduced RNANR, a new set of algorithms for the exploration of RNA kinetics landscapes at the secondary structure level. RNANR considers locally optimal structures, a reduced set of RNA conformations, in order to focus its sampling on basins in the kinetic landscape. Along with an exhaustive enumeration, RNANR implements a novel non-redundant stochastic sampling, and offers a rich array of structural parameters. Our tests on both real and random RNAs reveal that RNANR allows to generate more unique structures in a given time than its competitors, and allows a deeper exploration of kinetics landscapes [27].

7.6. Large-scale sequencing data indexing

Petabytes of DNA and RNA sequencing data are currently stored in online databases. It is currently possible to access these databases in two ways: 1) metadata queries, such as organism, instrument type, etc, and 2) download raw data. Due to the sheer size of the data, the web servers do not offer the possibility to search for sequences inside datasets. Such an operation would be invaluable to biology investigators, for example to determine which experiments contain an organism of interest, high expression of a certain transcript, a certain mutation, etc. Prior work exists for indexing sequencing data (Bloom Filter Tries, Sequence Bloom Trees), yet the performance remains prohibitive (either high memory usage, or several days for performing certain queries).

We proposed a new formalism, the Allsome Sequence Bloom Trees [28]. It improves upon Sequence Bloom Trees in terms of construction time (by 50%) and query time (by 40-85%), and also permits dataset-vs-dataset searches. The method has been tested by indexing a subset of 2,652 RNA-seq human experiments from the Sequence Read Archive. Allsome Sequence Bloom Trees pave the way towards "Google" searches of petabytes of sequencing data.

CAPSID Project-Team

7. New Results

7.1. Drug Targeting and Adverse Drug Side Effects

Identifying new molecular targets using comparative genomics and knowledge of disease mechanisms is a rational first step in the search for new preventative or therapeutic drug treatments [47]. We are mostly concerned with three global health problems, namely fungal and bacterial infections and hypertension. Through on-going collaborations with several Brasilian laboratories (at University of Mato Grosso State, University of Maringá, Embrapa, and University of Brasilia), we previously identified several novel small-molecule drug leads against *Trypanosoma cruzi*, a parasite responsible for Chagas disease [72]. With the University of Maringá, we subsequently found several active molecules against the flavoenzyme TRR1 in *Candida albicans*, and two manuscripts are in preparation. We also proposed several small-molecule inhibitors against *Fusarium graminearum*, a fungal threat to global wheat production [47], [31]. Two further manuscripts on this topic are currently in preparation. Concerning hypertension, we continued our collaboration with Prof. Catherine Llorens-Cortes at Collège de France to study the interaction between the apelin receptor (a transmembrane protein important for blood pressure regulation) and the aminopetidase A enzyme [15].

It is well known that many therapeutic drug molecules can have adverse side effects. However, when patients take several combinations of drugs it can be difficult to determine which drug is responsible for which side effect. In collaboration with Adrien Coulet (Orpailleur team co-supervisor of Gabin Personeni) and Prof. Michel Dumontier (Biomedical Informatics Research Laboratory, Stanford), we developed an approach which combines multiple ontologies such as the Anatomical Therapeutical Classification of Drugs, the ICD-9 classification of diseases, and the SNOMED-CT medical vocabulary together with the use of Pattern Structures (an extension of Formal Concept Analysis) in order to extract association rules to analyse the co-occurrence of adverse drug effects in patient records [57], [56]. A paper describing this work has been published in the Journal of Biomedical Semantics [20].

7.2. Docking Symmetrical Protein Structures

Many proteins form symmetrical complexes in which each structure contains two or more identical copies of the same sub-unit. We recently developed a novel polar Fourier docking algorithm called “Sam” for automatically assembling symmetrical protein complexes. A journal article describing the Sam algorithm has been published [8]. An article describing the results obtained when using Sam to dock several symmetrical protein complexes from the “CASP/CAPRI” docking experiment has also been published [18]. This study showed that many of the models of protein structures built by members of the “CASP” fold prediction community are “dockable” in the sense that Sam is able to find acceptable docking solutions from amongst the CASP models.

7.3. Multiple Flexible Protein Structure Alignments

Comparing two or more proteins by optimally aligning and superposing their backbone structures provides a way to detect evolutionary relationships between proteins that cannot be detected by comparing only their primary amino-acid sequences. The latest version of our “Kpax” protein structure alignment algorithm can flexibly align pairs of structures that cannot be completely superposed by a single rigid-body transformation, and can calculate multiple alignments of several similar structures flexibly [9]. In collaboration with Alain Hein of the INRA lab “Agronomie et Environnement”, we used Kpax to help study the structures of various “Cyp450” enzymes in plants [21]. In collaboration with Emmanuel Levy of the Weizmann Institute, we used Kpax to superpose and compare all of the symmetrical protein complexes in the Protein Databank in order to verify or remediate their quaternary structure annotations. A manuscript describing this work has been published in Nature Methods [16].

7.4. Large-Scale Annotation of Protein Domains and Sequences

Many protein chains in the Protein Data Bank (PDB) are cross-referenced with Pfam domains and Gene Ontology (GO) terms. However, these annotations do not explicitly indicate any relation between EC numbers and Pfam domains, and many others lack GO annotations. In order to address this limitation, as part of the PhD thesis project of Seyed Alborzi, we developed the CODAC approach for mining multiple protein data sources (i.e. SwissProt, TremBL, and SIFTS) in order to associate GO molecular function terms with Pfam domains, for example. We named the software implementation "GO-DomainMiner". This work was first presented at IWBBIO 2017 [23]. A full paper has been submitted to a special issue of *BMC Bioinformatics*, and is now in review. In collaboration with Maria Martin's team at the European Bioinformatics Institute (EBI), we combined the CODAC approach with a novel combinatorial association rule based approach called "CARDM" for annotating protein sequences. When applied to the large Uniprot/TrEMBL sequence database of 63 million protein entries, CARDM predicted over 24 million EC numbers and 188 million GO terms for those entries. A journal paper in collaboration with the EBI on comparing the quality of these predicted annotations with other state of the art annotation methods is in preparation, and a poster was presented at ISMB-ECCB-2017 [24].

7.5. Distributed Protein Graph Processing

The huge number of protein sequences in protein databases such as UniProtKB calls for rapid procedures to annotate them automatically. We are using existing protein annotations to predict the annotations of new or non-reviewed proteins. In this context, we developed the "DistNBLP" method for annotating protein sequences using a graph representation and a distributed label propagation algorithm. DistNBLP uses the BLADYG framework [12] to process protein graphs on multiple compute nodes by applying a neighbourhood-based label propagation algorithm in a distributed way. We applied DistNBLP in the recent "CAFA 3" (critical Assessment of Protein Function Annotation) community experiment to annotate new protein sequences automatically. This work was presented as a poster at ISMB/ECCB-2017 [22]. We are also interested in feature selection for subgraph patterns. In collaboration with the LIMOS laboratory at Université Clermont Auvergne we also developed a scalable approach using MapReduce for identifying sub-graphs having similar labels in very large graphs [17].

DYLISS Project-Team

7. New Results

7.1. Data integration and pre-processing with semantic-based technologies

Participants: Olivier Dameron, Xavier Garnier, Yann Rivault, Anne Siegel, Denis Tagu.

Interoperable infrastructure and implementation of a health data model for remote monitoring of chronic diseases with comorbidities In the context of **telemedecine**, we worked on a numerical application for monitoring patients with chronic diseases. We have developed a system based on a formal ontology that integrates the alert information and the patient data extracted from the electronic health record in order to better classify the importance of alerts. A pilot study was conducted on atrial fibrillation alerts. The results suggest that this approach has the potential to significantly reduce the alert burden in telecardiology [101], [100]. In 2017, we proposed an architecture supporting data exchange in the context of multiple chronic diseases [*O. Dameron, Y. Rivault*] [27].

AskOmics, a web tool to integrate and query biological data using semantic web technologies The software AksOmics has been adapted to two types of scientific topics important in agronomical and environmental sciences: plant genomic data and insect pest genomic data. With *AskOmics*, plant genomicists (from academic and private labs from the Rapsodyn project - Investment for the future) working on the rapeseed (Brassica napus) are able to tackle the understanding of which gene copy is active or repressed in key developmental processes in relation with seed quality and oil production, in the frame of plant breeding. Additionally, entomologists use this tool to extract valuable knowledge on the way insect pests such as aphids are able to rapidly disseminate on crops, in the frame of free-pesticide methods for plant protection. *AskOmics* has been presented to the international community of insect genomics (i5k: http://i5k.github.io/) by web-seminars and *AskOmics* developers have been invited at international workshops. For facilitating AskOmics's adoption by end-user, it has recently been integrated within the Galaxy workflow engine [*O. Dameron, X. Garnier, A. Siegel, D. Tagu*] [28], [29], [23]

7.2. Data and knowledge integration based on combinatorial optimization

Participants: Meziane Aite, Lucas Bourneuf, Marie Chevallier, Damien Eveillard, Clémence Frioux, Jeanne Got, Julie Laniau, François Moreews, Jacques Nicolas, Anne Siegel.

A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [59] was valuable for figuring out the main gene-regulators of the response of porks to several diets [*F. Moreews, A. Siegel*] [18]

FCA in a Logical Programming Setting for Visualization-oriented Graph Compression We have explored the underlying idea of lossless **network compression** to address the problem of uncertainty in biological networks built from predictions, to help to visualize the networks and to classify their nodes in accordance with available annotations [119]. Network compression has been used with success in Dresden (M. Schroeder) with a heuristic approach called Power Graph analysis building abstract graphs where nodes are clusters of nodes in the initial graph and edges represent bicliques between two sets of nodes. First encouraging results have been presented (best paper award) showing that it is possible to mimic the Power Graph behaviour while opening the possibility to achieve better compression levels compared to alternative compression schema. [*L. Bourneuf, J. Nicolas*] [24]

Metabolic network completion and analysis We released the application paper of the tool *Meneco*, a tool dedicated to the topological gap-filling of genome-scale draft metabolic networks. The tool reformulates gap-filling as a qualitative combinatorial optimization problem, omitting constraints raised by the stoichiometry, and solves this problem using Answer Set Programming. Run on an artificial test set of 10,800 degraded *Escherichia coli* networks, we evidenced that *Meneco* outperforms the stoichiometry-based tool *Gapfill* in terms of precision. In addition, *Meneco* reports 10 times less putative reactions than MILP-based tool *Fastgapfill* for an equivalent precision. This is a strong advantage for manual curation post-processing, since curating 50 to 80 reactions is still possible whereas manually-curating 800 reactions is out-of-range. Meneco was applied to the reconstruction and understanding of a pathogeneic strain of salmon. [*C. Frioux, J. Got, A. Siegel*] [21], [16]

Toward the study of metabolic functions in communities of organisms In [21], we provided a first example on how to use topological metabolic modeling to assess the complementarity between two members of an algal ecosystem. Since this study, we generalized the selection of subcommunities of interest and propose likely interactions that could occur between seaweeds and their associated bacteria. A focus has also been done on plant microbiota and the reasons underlying the organization of the community. Altogether, these on-going works enable a better understanding of holobiont organizations and functioning. [*M. Aite, M. Chevallier, C. Frioux, J. got, A. Siegel, C. Trottier*] [21], [31], [30]

Hybrid Metabolic Network Completion In order to improve the precision of gap-filling approaches, we introduced a hybrid approach to formally reconcile existing stoichiometric and topological approaches to network completion in a unified formalism. An hybrid ASP encoding based on MILP constraint propagator was developed. It relies upon the theory reasoning capacities of the ASP system Clingo to solve the resulting logic program with linear constraints over reals. For short, this technology made it possible to combine the best of the combinatorial problem solver Clingo with the MILP solver CPlex. Run on the artificial test set of 10,800 degraded *Escherichia coli* networks introduced in [21], our approach yielded greatly superior results than obtainable from purely qualitative or MILP approaches. [*C. Frioux, A. Siegel*] [26], [19]

Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks Whenever flux or graph-based criteria are used to study metabolic networks, these analyses are generally centered on the outcome of the network and considers all metabolic compounds to be equivalent in this respect. We generalized the concept of essentiality to metabolites and introduced the concept of the phenotypic essential metabolite (PEM) which influences the growth phenotype according to sustainability, producibility or optimal-efficiency criteria. The exhaustive study of phenotypic essential metabolites in six genome-scale metabolic models suggests that the combination and the comparison of graph, stoichiometry and optimal flux-based criteria allow some features of the metabolic network functionality to be deciphered by focusing on a small number of compounds. [*C. Frioux, J. Laniau, A. Siegel*] [19]

7.3. Systems biology

Participants: Jérémie Bourdon, Jean Coquet, Victorien Delannée, Jacques Nicolas, Anne Siegel, Nathalie Théret, Pierre Vignet.

A modeling approach to evaluate the balance between bioactivation and detoxification of MeIQx in human hepatocytes Heterocyclic aromatic amines (HAA), including MeIQx, are environmental and food contaminants that are potentially carcinogenic for humans. Using a computational approach, we developed a numerical model for MeIQx metabolism that predicts the MeIQx biotransformation into detoxification or bioactivation pathways according to the concentration of MeIQx. Our results demonstrate that CYP1A2 is a key enzyme in the system that regulates the balance between bioactivation and detoxification. This highlights the importance of complex regulations of enzyme competitions that should be taken into account in any multi-organ model [*V. Delannée, A. Siegel, N. Théret*] [17]

caspo: a toolbox for automated reasoning on the response of logical signaling networks families The accompanying paper of the complete family of modules introduced in the caspo software was published in 2017 (see software section for details) [*A. Siegel*] [22]

Identifying Functional Families of Trajectories in Biological Pathways by Soft Clustering: Application to TGF-β Signaling At a dynamical level, in [40], reaction-based and regulatory information was transpoed in a unified formalism of enriched Petri Nets (discrete dynamical systems), namely 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. This allowed integrating the 137 human signaling maps from the Pathway Interaction Database (PID) into a single unified large-scale dynamic model. Simulation and model checking analyses evidence that 15,934 different sets of molecules are able to regulate 159 of TGF-β target genes (TGF-β is a multifunctional cytokine that regulates mammalian cell development, differentiation, and homeostasis). Further analysis of these 15,934 sets of molecules by biological experts is obviously impractical. Our study identified five clusters of sets of molecules for which enrichment analysis highlighted the over-represented molecules as well as the specific biological processes they are associated with. These results are biologically-relevant and consistent with the pleiotropic nature of TGF-β [*J. Coquet, N. Théret, O. Dameron*] [25]

A Logic for checking the probabilistic steady-state properties of reaction networks. We have constructed a probabilistic analog to flux balance analysis of reaction networks to enable a formal verification of logical constrains about the stationary regime of a system by using information from experimental variances and co-variances. This is mainly based on a stationary analysis of the probabilistic dynamics relying on a Bernoulli approximation of a reaction network. The analysis requires solving non linear optimization problems [*J. Bourdon, A. Siegel*] [20]

7.4. Sequence and structure annotation

Participants: Catherine Belleannée, François Coste, Jacques Nicolas.

Better scoring schemes for the recognition of functional proteins by protomata The machine learning algorithm included in *Protomata-learner* learns weighted automata representing both functional families from the sequences of amino acids, and the possible disjunctions between members. We investigated alternative sequence weighting strategies and null-models. We introduced a normalization of the score, and a method to assess the significance of scores, to simplify the prediction. Preliminary results show a good improvement of the prediction power of the computed models. [*F. Coste*] [36]

Detection of mutated primers and impact on targeted metagenomics results In targeted metagenomics, an initial task is the detection in each sequence of the primers used for amplifying the targeted region. The selected sequences are then trimmed and clustered in order to inventory the species present in the sample. Common pratices consist in retaining only the sequences with perfect primers (i.e. non-mutated by sequencing error). In the context of a study characterizing the biodiversity of tropical soils in unicellular eukaryotes, we have implemented the search for mutated primers, using the grammatical pattern matching tool Logol, and shown that retrieving sequences with mutated primers has a significant impact on targeted metagenomics results, as it makes possible to detect more species (7% additional OTUs in our study). [*C. Belleannée*] [34].

First landscape of binding to chromosomes for a domesticated mariner transposase in the human genome. In order to study the diversity of genomic targets of the SETMAR protein in two colorectal cell lines, a first task was to massively discover the Made1 80-bp transposon element in the human genome. For that, we used our Logol grammar-like approach to look for non perfect Made1 instances. In Logol, a pattern can be divided into several sub-patterns. The Made1 model took advantage of this feature to strengthen the most conserved regions. Cumulating this search with the Blast alignment search permitted to significantly increase the Made1 annotation in the human genome.[*C. Belleannée*] [33]

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2017.

We tried to organise these following the six main axes of research of the team. Clearly, in some cases, a result obtained overlaps more than one axis. In such case, we chose the one that could be seen as the main one concerned by such results.

We did not indicate here the results on more theoretical aspects of computer science if it did not seem for now that they could be relevant in contexts related to computational biology. Actually, we do believe those on scheduling [25], [24], and on text [37], graph [4], [32], [34], [5], [36], [35] or general algorithmic problems notably related to performance issues [23], [28] could in the future become more specifically relevant for life sciences (biology or ecology). We did not indicate either work that was done a few years ago by members who were in ERABLE but whose associated publication appeared only this year [27].

Notice that the theoretical results related to problems closely resembling questions that have already been addressed by us in computational biology and that we present below concern not only cross-fertilising issues among different computational approaches, and we therefore extended the title of this axis for the purpose of presenting such results, for now purely theoretical.

A few other results of 2017 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised, or the work represented a survey.

6.2. Identifying the molecular elements

Motif tries for pattern discovery. In [14], the motif trie data structure was introduced to improve the extraction of recurring patterns in sequences. Such extraction concerned maximal patterns with at most k don't care symbols and at least q occurrences, according to a given maximality notion. The motif trie was applied to this problem, also showing how to build it efficiently. This led to the first algorithm that attains a stronger notion of output-sensitivity, where the cost for an input sequence of n symbols is proportional to the actual number of occurrences of each pattern, which is at most n (much smaller in practice). This avoids the best-known cost of $O(nc)O(nc)$ per pattern, for a constant $c > 1$, which is otherwise impractical for massive sequences with a large value of n.

Identification of genome and alternative splicing variants in RNA-seq data. The team's work on identifying alternative splicing and other genome variants such as SNPs (Single Nucleotide Polymorphism), indels, etc., started around 2010. This has concerned mostly RNA-seq data also for the variants investigated.

Both DNA and RNA-seq data analysis using so-called NGS (Next Generation Sequencing) is a domain of research that has been active for decades now, with many open questions remaining despite such long and intense activity. One is the case of non-model organisms, but actually there is another major problem that has not been solved, at least in any really satisfying way since the premises of genome sequencing. This is the problem of repeats. Notice however that repeats are not just "problems to be avoided", but have a strong biological interest in themselves, notably those related to transposable elements. Various papers of the team in 2017, notably [13], [19], [22], [1], were concerned with the study of such elements.

As concerns non-model organisms, the team extended a method it had previously developed, called KISSPLICE, to identify, quantify and annotate SNPs without any reference genome, using RNA-seq data only. The paper (Lopez-Maestre *et al.*, *Nucleic Acids Research*, 44(19):e148, 2016) appeared at the end of 2016. There we showed that individuals can be pooled prior to sequencing if not enough material is available from one individual. Using pooled human RNA-seq data, we clarified the precision and the recall of our method and discussed them with respect to others which use a reference genome or an assembled transcriptome. We then validated experimentally the predictions of our method using RNA-seq data from two non-model species. The method can be used for any species to annotate SNPs and predict their impact on the protein sequence. It enables to test for the association of the identified SNPs with a phenotype of interest. One of the phenotypes explored was related to the dependence of the insect *Asobara tabida* on its endosymbiont *Wolbachia*.

The methodological part of the work above relied in part on a number of more theoretical results, related to algorithmics and more specifically focused on the problem of repeats [21]. The most theoretical recent work of the team, accepted at the 43rd International Workshop on Graph-Theoretic Concepts in Computer Science (WG) in 2017 [30], proposed the notion of a bubble generator set, *i.e.* of a polynomial-sized subset of bubbles from which all the others can be obtained, also in polynomial time, through the application of a specific symmetric difference operator. This is further described in the last axis (Axis 6).

Genome and haplotype assembly. Fully assembling the genome sequence of an organism remains an important and challenging task. Genome scaffolding (*i.e.* the process of ordering and orientating contigs) of *de novo* assemblies usually represents the first step in most genome finishing pipelines. The team started by developing an algorithm (called MEDUSA) for such task (Bosi *et al.*, *Bioinformatics*, 31(15):2443-2451, 2015). It exploited information obtained from a set of (draft or closed) genomes from related organisms to determine the correct order and orientation of the contigs. It formalised the scaffolding problem by means of a combinatorial optimisation formulation on graphs and implements an efficient constant factor approximation algorithm to solve it. In contrast to the majority of the scaffolders, it did not require either prior knowledge on the input dataset (usually of micro-organisms) or the availability of paired-end read libraries. MEDUSA however presented limitations both in the construction of the scaffolding graph for large genomes, and in the subsequent assembly. The first aspect has been recently greatly improved by a method developed in collaboration with researchers (among which Alex di Genova) from Chile. This work led to the software FAST-SG already publicly available, and to a first publication that is in revision.

6.3. Inferring and analysing the networks of molecular elements

Metabolic impact of a change of conditions. The increasing availability of metabolomics data enables to better understand the metabolic processes involved in the immediate response of an organism to environmental changes, where the latter can be related to the presence of other species. 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 infer how such metabolites are interconnected. The team introduced a method that enables to organise the data from any metabolomics experiment into what we initially called *metabolic stories* when we were working with a simpler, graph representation of metabolism, and which have now become *metabolic hyperstories* as more accurate directed hypergraphs representations are considered. Each (hyper)story corresponds to a possible scenario explaining the flow of matter between the metabolites of interest. The initial work on a graph representation led to the GOBBOLINO + TOUCHÉ software (Milre *et al.*, *Bioinformatics*, 30(1):61-70, 2014). Two newer works working with directed hypergraphs were presented in the PhD of Alice Julien-Laferrière (defended in 2016). Two papers are currently in preparation. They led to the software TOTORO (which uses a qualitative measurement of concentrations in two steady-states) and KOTOURA (which infers quantitative changes of the reactions) which are both already publicly available.

Metabolic network reconstruction and comparison for understanding virulence. The respiratory tract of swines is colonised by several bacteria among which are three Mycoplasma species: *Mycoplasma flocculare*, *Mycoplasma hyopneumoniae* and *Mycoplasma hyorhinis*. While colonisation by *M. flocculare* is virtually asymptomatic, *M. hyopneumoniae* is the causative agent of enzootic pneumonia and *M. hyorhinis* is present in cases of pneumonia, polyserositis and arthritis. The genomic resemblance among these three Mycoplasma species combined with their different levels of pathogenicity is an indication that they have unknown mechanisms of virulence and differential expression, as for most mycoplasmas. We performed whole-genome metabolic network reconstructions for the three mycoplasmas, as well as cultivation tests and metabolomic experiments through nuclear magnetic resonance spectroscopy (NMR) (Ferrarini *et al.*, *BMC Genomics*, 17(1):353, 2016). We were able to infer from such reconstructed networks that the lack of pathogenicity of *M. flocculare* if compared to the highly pathogenic *M. hyopneumoniae* may be related to its incapacity to produce cytotoxic hydrogen peroxide. A second, more experimentally oriented-paper is currently under revision.

6.4. Modelling and analysing a network of individuals, or a network of individuals' networks

On unrooted and root-uncertain variants of several well-known phylogenetic network problems Genetic hybridisation is the process individuals from genetically distinct populations that are able to interbreed and this produce a hybrid.

The hybridisation number problem refers to finding the minimum number of hybridisation events necessary to explain conflicts among several evolutionary trees. It requires to embed a set of binary rooted phylogenetic trees into a binary rooted phylogenetic network such that the number of nodes with in-degree two is minimised. However, from a biological point of view accurately inferring the root location in a phylogenetic tree is notoriously difficult and poor root placement can artificially inflate the hybridisation number. To this end, a number of relaxed variants of this problem were studied in [29]. We started by showing that the fundamental problem of determining whether an unrooted phylogenetic network displays (*i.e.* embeds) an unrooted phylogenetic tree, is NP-hard. On the positive side, we showed that this problem is FPT in reticulation number. In the rooted case, the corresponding FPT result is trivial, but here a more subtle argumentation was required. Next, we showed that the hybridisation number problem for unrooted networks (when given two unrooted trees) is equivalent to the problem of computing the tree bisection and reconnect distance of the two unrooted trees. We then considered the "root uncertain" variant of the hybridisation number. Here we are free to choose the root location in each of a set of unrooted input trees such that the hybridisation number of the resulting rooted trees is minimised. On the negative side, we showed that this problem is APX-hard. On the positive side, we showed that it is FPT in the hybridisation number, via kernelisation, for any number of input trees.

Phylogenetic tree reconciliation. Phylogenetic tree reconciliation consists in a mapping of one tree (usually the symbiont tree) to the other (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. Indeed, the problem is not just that if we proposed a single solution, there is a good chance we would miss the "true" answer, but also that we would lose the capacity to verify whether there exist some characteristics that are common to enough of the solutions to increase our confidence in the "story" such reconciliation tells of the past.

When the ERABLE team started addressing this issue, only two algorithms existed that attempted such enumeration; in one case (software CoRe-Pa) not all possible solutions were produced while in the other (software Notung) not all cost vectors were handled. We then introduced a polynomial-delay algorithm, called Eucalypt, for enumerating all optimal reconciliations, and showed that in general many solutions exist (Donati *et al.*, *Algorithms for Molecular Biology*, 10(1):11, 2015). Some might not be time-feasible. However, we further showed that, among the many solutions that are usually found, in the majority of the cases, at least some will be time-feasible, and we provided a polynomial algorithm to test for time-feasibility.

We also considered a restricted version of the model where host switches are allowed to happen only between species that are within some fixed distance along the host tree. This restriction allows 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 optimal solution is found.

More recently, we defined two equivalence relations that enable to identify many reconciliations with a single one, thereby reducing their number. These results were published in a paper which was accepted at CIBB 2017 and will appear in the *LNCI-LNCS* proceedings of the conference (published after CIBB). Extensive experiments indicated that the number of output solutions greatly decreases in general. By how much clearly depends on the constraints that are given as input. An extended journal version of this work that includes its theoretical part will be submitted at the beginning of 2018. Other forms of grouping (or clustering) solutions are also being explored that rely instead on defining a distance between two different reconciliations. Two approaches are being investigated, one in collaboration with a researcher in Italy (paper in preparation), and the other with researchers in the UK (one paper submitted and one in preparation).

Improving the biological realism of coevolutionary models. The host-symbiont coevolutionary models developed so far needed also to be improved. The realism we wished to add to such models was for now the possibility to handle the case of multiple associations of a symbiont. Among the few previous works that allowed for this, all presented some limitation either in terms of the model or of the algorithm developed. Handling such multiple associations requires to introduce an event that was little or not formally considered in the literature. This is the event of *spread*, which precisely corresponds to the invasion of different hosts by a same symbiont. In this case, as when spreads are not considered, the optimal reconciliations obtained will depend on the choice made for the costs of the events. The need to develop statistical methods to assign the most appropriate ones therefore remained also of actuality. This is one of the problems we addressed in the PhD of Laura Urbini that was defended in October 2017. Two types of spread were in fact introduced: vertical and horizontal. The first corresponds to the case where the evolution of the symbiont "freezes" while the symbiont continues to be associated with a host and with the new species that descend from this host. The second includes both an invasion, of the symbiont which remains with the initial host but at the same time gets associated with ("invades") another one incomparable with the first, and a double freeze (in relation to the evolution of the host with which it was initially associated and in relation to the evolution of the second one it "invaded"). Two papers addressing distinct aspects related to the spread problem with different approaches are in preparation and will be submitted before the end of 2017 or beginning of 2018.

Estimating the frequency and expansion process of an infection We addressed the question of how often an infection occurs and of whether its expansion reached an equilibrium using as model *Wolbachia*. *Wolbachia* is a bacterial genus that infects about half of all arthropods, with diverse and extreme consequences ranging from sex-ratio distortion and mating incompatibilities to protection against viruses. These phenotypic effects, combined with efficient vertical transmission from mothers to offspring, satisfactorily explain the invasion dynamics of *Wolbachia* within species. However, beyond the species level, the lack of congruence between the host and symbiont phylogenetic trees indicates that *Wolbachia* horizontal transfers and extinctions do happen and underlie its global distribution.

In [3], we inferred recent acquisition/loss events from the distribution of *Wolbachia* lineages across the mitochondrial DNA tree of 3600 arthropod specimens, spanning 1100 species from Tahiti and the surrounding islands. We showed that most events occurred within the last million years, but are likely attributable to individual level variation (*e.g.*, imperfect maternal transmission) rather than to population level variation (*e.g.*, *Wolbachia* extinction). At the population level, we estimated that mitochondria typically accumulate 4.7% substitutions per site during an infected episode, and 7.1% substitutions per site during the uninfected phase. Using a Bayesian time calibration of the mitochondrial tree, these numbers translate into infected and uninfected phases of approximately 7 and 9 million years. Infected species thus lose *Wolbachia* slightly more often than uninfected species acquire it, supporting the view that its present incidence, estimated here slightly below 0.5, represents an epidemiological equilibrium.

6.5. Going towards control

Quantitative synthetic biology. Synthetic biology has boomed since the early 2000s when it started being shown that it was possible to efficiently synthetise compounds of interest in a much more rapid and effective way by using other organisms than those naturally producing them. However, to thus engineer a single organism, often a microbe, to optimise one or a collection of metabolic tasks may lead to difficulties when attempting to obtain a production system that is efficient, or to avoid toxic effects for the recruited microorganism. The idea of using instead a microbial consortium has thus started being developed in the last decade. Establishing which consortium is best for the production of a given compound or set thereof remains however a great challenge. The team introduced an initial model and a method, called MULTIPUS, that enable to propose a consortium to synthetically produce compounds that are either exogenous to it, or are endogenous but where interaction among the species in the consortium could improve the production line (Julien-Laferrière *et al.*, *Scientific Reports*, 6, 2016).

Since the work on MULTIPUS, the team has been considering quantitative approaches for synthetic biology. We thus explored the concept of multi-objective optimisation in the field of metabolic engineering when both continuous and integer decision variables are involved in the model. In particular, we proposed multi-objective models, initially for a single species, to suggest reaction deletion strategies, and also to deal with situations where several functions must be optimised simultaneously, such as the maximisation of bioproducts while minimising toxicity (Hartmann *et al.*, *BMC Systems Biology*, see https://www.ncbi.nlm.nih.gov/pubmed/29268790, just accepted and not yet visible in Hal-Inria). We compared our results with those obtained by using the well-known bi-level optimisation model of OPTKNOCK, and studied two multi-objective optimisation problems arising from the metabolic engineering of microorganisms. One of them, using Yeast, has been validated experimentally. The work is submitted. The team has then started expanding it to communities (Master Thesis of Irene Ziska who is continuing into a PhD).

6.6. Health

Rare Diseases. Splicing is an essential step in the process leading to gene expression because it not only removes the introns from the primary transcripts, but also generates a combination of mature transcripts through the differential inclusion/exclusion of exons and sometimes retention of introns. Some pathologies are associated to such abnormal splicing. This is the case of the Taybi-Linder Syndrome (TALS), a very rare malformative syndrome with autosomal recessive transmission, belonging to the group of microcephalic dwarfism and responsible for death usually before the age of 2 years. This pathology was recently found to be caused by mutations in RNU4ATAC, a small nuclear RNA, which is an essential component of the minor spliceosome. We started a collaboration with the group of Pr. P. Edery (who first identified this alteration in 2012) with the objective to establish a comprehensive catalog of splice alterations in several cohorts of TALS patients. In this work, we take advantage of our reference-free assembly approach of transcripts (KISSPLICE) in order to detect new splicing alterations and to identify the associated deregulated signalling genes and pathways.

Cancer. Alain Viari has continued to develop a strong interaction with clinicians concerned with cancer, notably of the breast and in the early human embryo. A number of papers have appeared in 2017 that describe this work [7], [6], [10], [11], [12], [16], [17], [18]. We highlight here just two.

The first [7] refers to breast cancer. Mismatch repair (MMR)-deficient cancers have been discovered to be highly responsive to immune therapies such as PD-1 checkpoint blockade, making their definition in patients, where they may be relatively rare, paramount for treatment decisions. In the study published in [7], we utilised patterns of mutagenesis known as mutational signatures, which are imprints of the mutagenic processes associated with MMR deficiency, to identify MMR-deficient breast tumours from a whole-genome sequencing dataset comprising a cohort of 640 patients. We identified 11 of 640 tumours as MMR deficient, but only 2 of 11 exhibited germline mutations in MMR genes or Lynch Syndrome. Two additional tumours had a substantially reduced proportion of mutations attributed to MMR deficiency, where the predominant mutational signatures were related to APOBEC enzymatic activity. Overall, 6 of 11 of the MMR-deficient

cases in this cohort were confirmed genetically or epigenetically as having abrogation of MMR genes. However, IHC analysis of MMR-related proteins revealed all but one of 10 samples available for testing as MMR deficient. Thus, the mutational signatures more faithfully reported MMR deficiency than sequencing of MMR genes, because they represent a direct pathophysiologic readout of repair pathway abnormalities. As whole-genome sequencing continues to become more affordable, it could be used to expose individually abnormal tumours in tissue types where MMR deficiency has been rarely detected, but also rarely sought.

The second [18] concerns early human embryo. Somatic cells acquire mutations throughout the course of an individual's life. Mutations occurring early in embryogenesis are often present in a substantial proportion of, but not all, cells in postnatal humans and thus have particular characteristics and effects. Depending on their location in the genome and the proportion of cells they are present in, these mosaic mutations can cause a wide range of genetic disease syndromes and predispose carriers to cancer. They have a high chance of being transmitted to offspring as de novo germline mutations and, in principle, can provide insights into early human embryonic cell lineages and their contributions to adult tissues. Although it is known that gross chromosomal abnormalities are remarkably common in early human embryos, our understanding of early embryonic somatic mutations is very limited. In this work, whole-genome sequences of normal blood from 241 adults was used to identify 163 early embryonic mutations. It was estimated that approximately three base substitution mutations occur per cell per cell-doubling event in early human embryogenesis and these are mainly attributable to two known mutational signatures. The mutations were then used to reconstruct developmental lineages of adult cells and demonstrate that the two daughter cells of many early embryonic cell-doubling events contribute asymmetrically to adult blood at an approximately 2:1 ratio. This study provided insights into the mutation rates, mutational processes and developmental outcomes of cell dynamics that operate during early human embryogenesis.

6.7. Cross-fertilising different computational approaches and other theoretical results

Bubble generator.
As mentioned earlier, a theoretical recent work of the team related to NGS analysis was accepted at the 43rd International Workshop on Graph-Theoretic Concepts in Computer Science (WG) in 2017 [30]. It introduced what was called a bubble generator.

Bubbles are pairs of internally vertex-disjoint (s, t)-paths with applications in the processing of DNA and RNA data. For example, enumerating alternative splicing events in a reference-free context can be done by enumerating all bubbles in a de Bruijn graph built from RNA-seq reads. However, listing and analysing all bubbles in a given graph is usually unfeasible in practice, due to the exponential number of bubbles present in real data graphs. In [30], we proposed a notion of a bubble generator set, *i.e.* a polynomial-sized subset of bubbles from which all the others can be obtained through the application of a specific symmetric difference operator. This set provides a compact representation of the bubble space of a graph, which can be useful in practice since some pertinent information about all the bubbles can be more conveniently extracted from this compact set. Furthermore, we provide a polynomial-time algorithm to decompose any bubble of a graph into the bubbles of such a generator in a tree-like fashion.

GENSCALE Project-Team

7. New Results

7.1. Data Structure

7.1.1. Minimal perfect hash function

Participants: Antoine Limasset, Guillaume Rizk, Pierre Peterlongo.

Minimal perfect hash functions are fundamental objects used in many applications. Existing algorithms and implementations that build such functions have in practice some upper bounds on the number of input elements they can handle, due to high construction time and/or memory usage. We propose a simple algorithm having very competitive construction times, memory usage and query times compared to state of the art techniques [27]. We provide a parallel implementation called BBHash. It is capable of creating a minimal perfect hash function of 10^{10} elements in less than 1 hour and 4 GB of memory. To the best of our knowledge, this library is also the first that has been successfully tested on 10^{12} input elements. Source code: https://github.com/rizkg/BBHash

7.1.2. Quasi-dictionary

Participants: Camille Marchet, Antoine Limasset, Pierre Peterlongo.

Indexing massive data sets is extremely expensive for large scale problems. In many fields, huge amounts of data are currently generated, however extracting meaningful information from voluminous data sets, such as computing similarity between elements, is far from being trivial. It remains nonetheless a fundamental need. In this context, we proposed a probabilistic data structure based on a minimal perfect hash function for indexing large sets of keys. This structure out-competes the hash table for construction, query times and for memory usage, in the case of the indexation of a static set. To illustrate the impact of algorithms performances, we provided two applications based on similarity computation between collections of sequences, and for which this calculation is an expensive but required operation. In particular, we showed a practical case in which other bioinformatics tools failed to scale up the tested data set or provide lower recall quality results [43].

7.2. Algorithms & Methods

7.2.1. Short Read Correction

Participants: Antoine Limasset, Pierre Peterlongo.

We proposed a new method to correct short reads using de Bruijn graphs, and we implemented it as a tool called Bcool. As a first step, Bcool constructs a corrected compacted de Bruijn graph from the reads. This graph is then used as a reference and the reads are corrected according to their mapping on the graph. We showed that this approach yields a better correction than $k$mer-spectrum techniques, while being scalable, making it possible to apply it to human-size genomic datasets and beyond [41].

7.2.2. Long transcriptomic read clustering

Participants: Camille Marchet, Pierre Peterlongo.

This contribution tackles the problem of clustering RNA reads in clusters representing all variants of each gene, in a *de novo* way i.e. without any reference sequences. Such problem is not new as is, but the latest, Third Generation Sequencing (TGS) data redefine it. Reads can now span full-length transcripts but at the price of very high error rates, mostly insertions and deletions. This makes difficult or impossible to use tools designed for previous sequencing data. Still, the property to obtain whole RNA molecules through reads is very promising to better describe a transcriptome. In this work, we targeted the need to extract relevant information from a TGS transcriptome, even when no reference is available. In collaboration with Jacques Nicolas from the Inria/IRISA Dyliss team, we propose a novel algorithm in the community detection framework, based on the clustering coefficient. In addition we propose an implementation of this algorithm in the tool CARNAC-LR and a pipeline for the processing of transcriptome data. We validated our tool on real data from mouse and showed that it could be accurate and precise even for lowly expressed genes. We showed that our approach can be complementary to a mapping in the case a reference exists, and that a straightforward use of CARNAC-LR enables to quickly assess the genes'e expression levels [42].

7.2.3. Statistically Significant Discriminative Patterns Search

Participants: Hoang Son Pham, Dominique Lavenier.

Identifying multiple SNPs combinations associated with diseases such as cancers or diabetes is a central goal of human genetics. Recently, discriminative pattern mining algorithms have been investigated to tackle genome-wide association studies (GWAS). We designed an algorithm, called SSDPS, to discover groups of items which have significant difference of frequency in case-control datasets. The algorithm directly uses relative risk measures such as risk ratio, odds ratio and absolute risk reduction combined with confidence intervals as anti-monotonic properties to efficiently prune the search space. The algorithm discovers a complete set of discriminative patterns with regard to given thresholds or applies heuristic strategies to extract the largest statistically significant discriminative patterns in a given dataset. Experimental results on both synthetic datasets and three real variant datasets (Age-Related Macular Degeneration, Breast Cancer and Type 2 Diabetes) demonstrate that the SSDPS algorithm effectively detects multiple SNPs combinations in an acceptable execution time.

7.2.4. Reference free SNP detection in RAD-seq data

Participants: Jeremy Gauthier, Claire Lemaitre, Pierre Peterlongo.

We developed an original method for reference-free variant calling from Restriction site associated DNA Sequencing (RAD-Seq) data. RAD-seq is a technique widely employed in the evolutionary biology field. Based on the variant caller DiscoSnp, DiscoSnp-RAD explores the De Bruijn Graph built from all the read datasets to detect SNP and Indels. Tested on simulated and real datasets, DiscoSnp-RAD identifies thousands of variants suitable for different population genomics analyses. Furthermore, DiscoSnp-RAD stands out from other tools due to his completely different principle, making it significantly faster, in particular on large datasets [39].

7.2.5. Global Optimization for Scaffolding and Completing Genome Assemblies

Participants: Sebastien Francois, Rumen Andonov, Dominique Lavenier.

We developed a method for solving genome scaffolding as a problem of finding the longest simple path in a graph defined by the contigs that satisfies a maximal number of additional constraints encoding the insert-size information [26]. Then we solved the resulting mixed integer linear program to optimality using the Gurobi solver. We tested our algorithm on a benchmark of chloroplast genomes and showed that it outperforms other widely-used assembly solvers by the accuracy of the results.

7.2.6. Identification and characterization of long non-coding RNA

Participant: Fabrice Legeai.

We participated in the development and validation of the tool FeelNC (collaboration with IGDR group). This is a tool allowing the identification of long non coding RNA (lncRNA) from RNASeq reads with or without a reference genome. Contrary to other tools, it does not depend on the comparison with protein databanks, which usually require lots of computations, but used a machine learning approach based on a Random Forest model trained with general features such as multi k-mer frequencies and relaxed open reading frames. We delivered a module that allows to characterize the relationships of each long non coding RNA with the other genes in its genomics close environment, giving insights about the putative impact of the lncRNAs to the regulation of these genes [23], [24].

7.2.7. Characterizing repeat-associated subgraphs in de Bruijn graphs

Participant: Camille Marchet.

The main problem in genome assembly, namely repeats, is also present in transcriptomic data. They are dealt with using various heuristics in the de Bruijn Graph framework (dBG). In this work, we introduce a formal model for representing high copy-number and low-divergence repeats in RNA-seq data in dBG and infer the definition of repeat-associated subgraphs. We show that the problem of identifying such subgraphs in a dBG is NP-complete. Then we place ourselves in the case of local assembly of alternative splicing and show that such subgraphs can be avoided implicitly. Thus, more alternative splicing events can be enumerated than with previous approaches. Finally we show that this exploration of DBG explorations can improve de novo transcriptome evaluation methods [16].

7.3. Parallelism

7.3.1. Variant detection using processing-in-memory technology

Participants: Charles Deltel, Dominique Lavenier.

The concept of Processing-In-Memory aims to dispatch the computer power near the data. Together with the UPMEM company (http://www.upmem.com/), which is currently developing a DRAM memory enhanced with computing units, we investigate the parallelization of the detection of mutations on the human genome. Traditionnaly, this process is split into 2 steps: a mapping step and a variant calling step. Here, thanks to the high processing power of this new type of memory, the mapping step can nearly be done at the disk transfer rate, allowing the variant calling step to be done simultaneously on the host processor. The implementation is currently on going. First performance evaluations indicate speed-up of one or two order of magnitude compared to purely software implementation.

7.4. Bioinformatics Analysis

7.4.1. Study of marine plankton holobionts

Participants: Camille Marchet, Pierre Peterlongo.

We derived from the quasi-dictionary (described in previous section) a tool called Short Read Connector (SRC), able to find pairs of similar reads intra or inter read sets. We used SRC in meta-transcriptomics context to identify the actors of a symbiosis and help the assembly [44], [31]. The framework is the study of marina holobionts (host and its community of symbionts) for which few is known about the actors. In order to retrieve the functions that characterize such holobionts, RNA-seq reads from the sequencing of the whole holobiont are assembled *de novo*. Such assembly is prone to produce chimeras. Thus SRC is used to index sequences (reads, EST, assembled genes...) known to be close to the host and symbionts of the holobiont. Then, thanks to SRC's ability to find similarity between sequences even at a large scale, by querying reads of the holobiont we identify those similar to the host or symbionts. We report four categories: host, symbiont, shared and unassigned that can be assembled in a parallel way. As a first step we validate the SRC+assembly approach by comparing our result to literature with two known holobionts with eukaryote hosts (*Orbicella faveolata, Xestospongia muta*). We show that our approach can compare to previous results. In a second step we lean on a protist (Collodaria) holobiont for which the actors are poorly known. No assembled sequences exist in the literature so we compare the pipeline SRC+assembly to a sole assembly pipeline. Our main achievement is to highly reduce (up to $\sim$40%) the number of chimeras in the assembly compared to the sole assembly pipeline.

7.4.2. Pea aphid metagenomics

Participants: Cervin Guyomar, Fabrice Legeai, Claire Lemaitre.

We worked on a framework adapted to the study of genomic diversity and evolutionary dynamics of the pea aphid symbiotic community from an extensive set of metagenomics datasets. The framework is based on mapping to reference genomes and whole genome SNP-calling. We explored the genotypic diversity associated to the different symbionts of the pea aphid at several scales : across host biotypes, amongst individuals of the same biotype, or within individual aphids. Thorough phylogenomic analyses highlighted that the evolutionary dynamics of symbiotic associations strongly varied depending on the symbiont, reflecting different histories and possible constraints [40], [30].

7.4.3. Assembly and comparison of two genomes of highly polyphagous lepidopteran pests

Participants: Fabrice Legeai, Claire Lemaitre.

In this study, two genomes of an agronomical important lepidopteran pest, the noctuid moth *Spodoptera frugiperda*, were sequenced and compared, giving significant insights to the mechanisms involved in host-plant adaptation and speciation of this organism. In particular, we described the large expansion of gustatory receptors and detoxification genes among this polyphagous pest compared to other specialist Lepidoptera, and emphasizes the role of these 2 gene families in the evolution of one of the world's worst agricultural pests. We also provided the genome assemblies, gene annotations and whole genome alignments of both strains, and the comparison of both to a reference moth genome (*Bombyx mori*). For these purposes, several original methods were developed i) to correct genome assembly errors due to the high level of heterozygosity and ii) to extract structural variant calls from whole genome alignments [15].

7.4.4. Benchmark of de novo read dataset compression tools

Participants: Gaetan Benoit, Dominique Lavenier, Claire Lemaitre.

In this book chapter, we review the different approaches and their tools developed so far to compress sequencing data files. We detail the algorithms for each of the three main types of data contained in such files for each read : the header, the DNA and the quality sequences. We also provide a thorough benchmark of the numerous available tools on various sequencing datasets, evaluating the compression ratio as well as the running time and memory usage performances [33].

7.4.5. Genomics of the agro-ecosystems pests

Participants: Fabrice Legeai, Claire Lemaitre.

Within a large international network of biologists, GenScale has contributed to various projects for identifying important components such as protein coding or non coding genes involved in the adaptation of major agricultural pests to their environment. We provided or participated to the assembly and the annotation of 4 new aphids [17], [22], and 5 parasitic wasps. Following specific agreement or policy, these new genomes and annotations are available for a restricted consortium or a large community through the BioInformatics platform for Agro-ecosystems Arthropods (http://bipaa.genouest.org/is). Moreover our engagement in the agronomical pest genomics led to our contribution to other projects such as epigenetics and chromatin structure analysis [18], or the analysis of population genetics data for identifying hotspots of selection in the nematode *Globodera pallida* genome [14].

7.4.6. Comparison of approaches for finding alternative splicing events in RNA-seq

Participant: Camille Marchet.

In this work we compared an assembly-first and a mapping-first approach to analyze RNA-seq data and find alternative splicing (AS) events. Assembly-first approach enables to identify novel AS events and to detect events in paralog genes that are hard to find using mapping because of the multi-mapping results. On the other hand, the mapping-first approach is more sensitive and detects AS events in lowly expressed genes, and is also able to find AS events with exons containing transposable elements. In addition we support these results with experimental validation. We showed that in order to extensively study the alternative splicing via RNA-seq data and retrieve the most candidates, both approaches should be led. We provide a pipeline consituted of parallel local *de novo* assembly executed by KisSplice and mapping using a novel mapping workflow called FaRLine [37].

7.4.7. Microbial communities interaction between plant and their bioagressors

Participants: Susete Alves Carvalho, Fabrice Legeai, Claire Lemaitre, Pierre Peterlongo, Dominique Lavenier.

GenScale actively collaborates with the INRA group 'plant-microbial communities interactions' (IGEPP, Rennes) that analyze the interaction between plant, their associated microbial communities and different bioagressors. The ambition of the project is to understand the link between the taxonomic biodiversity of the microbiota and their functional diversity in relation with plant physiology and plant-bioagressors interactions. For this last point, an integrated metatranscritomic approach is developped. Beside wet lab and sequence productions, bioinformatics tools are needed and meta-transcriptomic pipelines analysis arecurrently developped based on the GenScale expertise.

7.5. Challenges

7.5.1. Participation to CAMI: de-novo metagenomics assembly competition

Participants: Charles Deltel, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

In metagenome analysis, computational methods for assembly, taxonomic profiling and binning are key components facilitating downstream biological data interpretation. However, a lack of consensus about benchmarking datasets and evaluation metrics complicates proper performance assessment. In this context, we participated to CAMI (Critical Assessment of Metagenome Interpretation), specifically on the assembly section with the Minia pipeline. The CAMI challenge aimed to benchmark programs on datasets of unprecedented complexity and realism. Benchmark metagenomes were generated from 700 newly sequenced microorganisms and 600 novel viruses and plasmids, including genomes with varying degrees of relatedness to each other and to publicly available ones and representing common experimental setups. Across all datasets, our assembly programs performed well for species represented by individual genomes, while performance was substantially affected by the presence of related strains [20].

IBIS Project-Team

6. New Results

6.1. Qualitative modeling of gene regulatory networks in food-borne pathogens

Bacteria are able to respond to a variety of environmental stresses, which poses food safety problems when these bacteria are food-borne pathogens. Addition of salt, one of the most ancient and common way of preserving food, subjects the bacteria to an osmotic stress to which some may survive. However, the molecular mechanisms of adaptation in food-born pathogens are largely unknown. As a first step towards better understanding these adaptation processes on the molecular level, Delphine Ropers and Aline Métris from the Institute for Food Research in Norwich (UK), invited researcher in IBIS last year, developed a qualitative model of the osmotic stress response in the model bacterium *Escherichia coli*. The qualitative dynamics of the network has been analyzed using the tool GENETIC NETWORK ANALYZER (GNA). The model has allowed to reproduce the behavior of *E. coli* cells adapting to an osmotic stress by including the regulatory mechanisms involved in the process. This work has been published in the *International Journal of Food Microbiology* [21]. It paves the way to modelling stress responses of other foodborne pathogens like *Salmonella* to stresses relevant for the food industry, for which much less is known.

6.2. Analysis of fluorescent 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.

Valentin Zulkower, former PhD student in IBIS, developed novel methods for the analysis of reporter gene data obtained in microplate experiments, 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. The linear inversion methods, published in *Bioinformatics* in 2015 [12], in have been implemented in the Python package WELLFARE and integrated in the web application WELLINVERTER. Funded by a grant from the Institut Français de Bioinformatique (IFB), Yannick Martin has improved WellInverter by developing a parallel computational architecture with a load balancer to distribute the analysis queries over several back-end servers, a new graphical user interface, and a plug-in system for defining high-level routines for parsing data files produced by microplate readers from different manufacturers. This has resulted in a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. This web service has been deployed on the IFB cloud and on an Inria server, accompanied by extensive user documentation, online help, and a tutorial. A paper on WELLINVERTER is in preparation.

While the use of microplate readers results in population-level measurements of gene expression, for many applications it is mandatory to monitor gene expression over time on the level of individual cells. Several developments in the past decade have enormously extended the capabilities to achieve this, in particular the combination of fluorescence time-lapse microscopy for precisely quantifying gene expression in single cells and microfluidics technology for cultivating bacteria in confined spatial compartments and under well-controlled experimental conditions. One of the most wide-spread microfluidics devices is the so-called mother machine shown in Figure 5 . A major problem is that functional software for image analysis (segmentation, tracking, lineage reconstruction, ...) adapted to the requirements of mother machine applications are still rare. IBIS has therefore collaborated with the BEAGLE project-team for the adaptation of their tool FLUOBACTRACKER

to the analysis of time-lapse movies of fluorescent reporter expression and bacterial growth in microfluidics devices. This collaboration, which has also involved the SERPICO project-team, was supported by the Technology Transfer and Innovation department of Inria, in the framework of the Inria Hub program, and has allowed the hiring of Cyril Dutrieux as a software engineer in IBIS.

6.3. Analysis of dynamic metabolomics data

An important step in the study of intracellular metabolism is the quantification of growth rates as well as uptake and excretion rates of metabolites in growing cellular populations. Traditional approaches are based on steady-state experiments, where time-invariant growth rates and exchange fluxes are measured in different experimental conditions. Technological advances in metabolomics have made it possible to monitor the concentration of extracellular metabolites over time, thus paving the way for the study of metabolism in transient conditions. Recovering time-varying exchange and growth rates from time-lapse metabolimics data is a key aspect of this challenge.

We have investigated the reconstruction of exchange reaction and growth rates from time-lapse measurements of external metabolite concentrations and population growth. In particular we have focused on the case of exhaustion of specific substrates, entailing sudden metabolic reorganization of the cell such as diauxie shifts in *E. coli*. Such discontinuities in the metabolic dynamics make data analysis and rate recosntruction particularly challenging but also information-rich. We have developed a Bayesian method that explicitly accounts for these sudden changes and the correlated adaptation of growth in order to accurately estimate time-varying exchange reaction and growth rates, and tested the method on real data from batch and fed-batch cultures of *E. coli* and *L. lactis* obtained at INRA/INSA Toulouse. The method is based on a time-inhomogeneous Gaussian process characterization of the rate dynamics, and Kalman smoothing techniques for the solution of the regularized estimation problem. Method and results were presented at the joint 2017 ISMB-ECCB conference, and published in the corresponding special issue of *Bioinformatics* [17]. The software implementing the method in Matlab is available at https://team.inria.fr/ibis/rate-estimation-software/, and has also been used for the data analysis in another joint publication with INRA/INSA Toulouse [20]. Further developments of the method are under consideration.

6.4. Models of carbon metabolism in bacteria

Adaptation of bacterial growth to changes in environmental conditions, such as the availability of specific carbon sources, is triggered at the molecular level by the reorganization of metabolism and gene expression: the concentration of metabolites is adjusted, as well as the concentration and activities of enzymes, the rate of metabolic reactions, the transcription and translation rates, and the stability of proteins and RNAs. This reprogramming of the bacterial cell is carried out by i) specific interactions involving regulatory proteins or RNAs that specifically respond to the change of environmental conditions and ii) global regulation involving changes in the concentration of RNA polymerase, ribosomes, and metabolite pools that globally affect the rates of transcription, translation, and degradation of all RNAs and proteins. While these phenomena have been well studied in steady-state growth conditions, recent works by IBIS members and collaborators support the view that regulatory mechanisms of growth adaptation are best observed in dynamical conditions.

A first study concerns the second messenger cAMP in *E. coli* and its role in carbon catabolite repression, the mechanism by which bacterial cells select their preferred carbon source for growth. Studies performed in steady-state conditions have questioned the importance of cAMP, leading to a controversy on its physiological role, more than fifty years after its discovery. In a recently submitted journal paper, reporting work started during the PhD thesis of Valentin Zulkower and continued over the past two years, we argue that in order to properly assess the role of cAMP one should shift the focus from steady-state to dynamical conditions. We show, by a combination of fluorescent reporter gene assays and quantitative modeling, that a transient peak in the expression of cAMP-dependent genes leads to the accumulation of proteins necessary for growth on a variety of alternative carbon sources. In the long run, the expression of genes cognate to the alternative carbon source present in the environment is maintained by dedicated positive feedback circuits. Our results thus demonstrate that carbon catabolite repression and diauxic growth need to be understood from a dynamical perspective within the context of a hierarchical regulatory network.

A quantitative description and understanding of this complex network, cutting across metabolism, gene expression, and signalling, can be accessed through mathematical modelling only. In collaboration with Andreas Kremling, professor at TU München and former visiting scientist in the IBIS project-team, Hans Geiselmann, Delphine Ropers and Hidde de Jong developed an ensemble of variants of a simple core model of carbon catabolite repression. The model variants, with two substrate assimilation pathways and four intracellular metabolites only, differ from one another in only a single aspect, each breaking the symmetry between the two pathways in a different manner. Interestingly, all model variants are able to reproduce the data from a reference diauxic growth experiment. For each of the model variants, we predicted the behaviour in two new experimental conditions. When qualitatively comparing these predictions with experimental data, a number of models could be excluded while other model variants are still not discriminable. The best-performing model variants are based on inducer inclusion and activation of enzymatic genes by a global transcription factor, but the other proposed factors may complement these well-known regulatory mechanisms. The model ensemble, which was described in a journal paper recently submitted for publication, offers a better understanding of the variety of mechanisms that have been proposed to play a role in carbon catabolite repression, but is also useful as an educational resource for systems biology.

The same focus on the dynamics of physiological processes has shaped a project on the post-transcriptional control of carbon central metabolism in *E. coli*. In the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique INRA-Inria, the collaboration of Delphine Ropers with Muriel Cocaign-Bousquet and Brice Enjalbert at INRA/INSA Toulouse has demonstrated the key role played by the post-transcriptional regulatory system CSR in growth transitions. In a multi-scale analysis of several wild-type and mutant strains of the CSR system, a variety of experimental data have been acquired in relevant conditions, including growth parameters, gene expression levels and metabolite pools. Data integration through the estimation of fermentation fluxes and flux balance analysis, using the method described above (Section 6.3), have elucidated the role of post-transcriptional regulation in the dynamics of glycogen storage and consumption, as well as the key role of the latter compound for bacterial fitness, through the regulation of intracellular energy levels. A paper summarizing the work has been published in *mBio* [20].

The collaboration with INRA/INSA de Toulouse is continued in the context of the PhD thesis of Thibault Etienne, funded by an INRA-Inria PhD grant, with the objective of developing models able to explain how cells coordinate their physiology and the functioning of the transcription, translation, and degradation machineries following changes in the availability of carbon sources in the environment.

6.5. 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 engineered) control of cell populations and even of single cells.

Work in IBIS on 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 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 identifiability, single-cell state estimation and control. Other problems related with single-cell modelling and extracellular variability are considered in eukaryotic cells through external collaborations.

Concerning identification of intrinsic noise dynamics in single cells, previous results on the contribution of stochasticity to parameter identifiability have been revisited in the context of reconstruction of unknown

networks. For the case of population snapshot meaurements, where the dynamics of the population statistics are observed by simple time-lapse experiments, we performed an analytical study of the additional information provided by variance measurements for the reconstruction of unknown first-order kinetics. Based on simulated example, we showed that a tremendous improvement in network reconstruction is achieved relative to the utilization of population-average statistics alone, as addressed by deterministic modelling. These exciting yet preliminary results were published in the form of a paper in the proceedings of the *IFAC World Congress* [22] and will be further developed.

Reconstruction of promoter activity statistics from reporter gene population snapshot data has been further investigated, leading to a full-blown spectral analysis and reconstruction method for reporter gene systems. Building upon reults in previous conference papers, in the context of the ANR project MEMIP (Section 8.2), we have characterized reporter systems as noisy linear systems operating on a stochastic input (promoter activity), and developed an inversion method for nonparametric estimation of promoter dynamics, namely the autocovariance function, from the considered readouts. These theoretical and simulation results have been submitted for journal publication and are also available as an arXiv pre-print. The method will be further developed and applied to real data and case studies.

Modelling of heterogeneity in isogenic cell populations is also an active research direction. Still in the context of MEMIP, in collaboration with the INBIO team, we are considering generalizations of our achievements on Mixed-Effects modelling and inference on yeast, in order to account for different sources of noise and lineage effects. As an offspring of this work, a study of inter-individual variability of *E. coli* gene expression and growth rate in growth arrest-and-restart experiments has been carried out with BIOCORE. Results obtained so far are part of the PhD thesis of Stefano Casagrande.

6.6. Modelling bacterial growth

Various mathematical approaches have been used in the literature to describe the networks of biochemical reactions involved in microbial growth. With various levels of detail, the resulting models provide an integrated view of these reaction networks, including the transport of nutrients from the environment and the metabolism and gene expression allowing the conversion of these nutrients into biomass. The models hence bridge the scale between individual reactions to the growth of cell populations. In a review article published in the *Journal of the Royal Society Interface* [18], several IBIS members as well as colleagues from the BIOCORE project-team, discuss various models of microbial growth that are, at first sight, quite diverse. They have a different scope and granularity, make different simplifications, use different approaches to obtain predictions from the model structure and have their origin in different fields. In the review we derive a general framework for the kinetic modelling of microbial growth from a few basic hypotheses on the systems of biochemical reactions underlying microbial growth. Additional simplifying assumptions lead to the several families of approximate models of microbial growth found in the literature, including self-replicator models of bacterial growth developed by Nils Giordano in his PhD thesis and published in *PLoS Computational Biology* last year [5]. This reveals how the models are related on a deeper level and provides a sound basis for further modelling studies.

Analysing the dynamics of some of the network models mentioned above becomes quickly intractable, when mathematical functions are for instance given by complex algebraic expressions resulting from the mass balance of biochemical reactions. In a paper published in the *Bulletin of Mathematical Biology* [16], Edith Grac, former post-doc in Ibis, Delphine Ropers, and Stefano Casagranda and Jean-Luc Gouzé from the BIOCORE project-team, have studied how monotone system theory and time-scale arguments can be used to reduce high-dimension models based on the mass-action law. Applying the approach to an important positive feedback loop regulating the expression of RNA polymerase in *E. coli*, made it possible to study the stability of the system steady states and relate the dynamical behaviour of the system to observations on the physiology of the bacterium *E. coli*.

6.7. Growth control in bacteria and biotechnological applications

The ability to experimentally control the growth rate is crucial for studying bacterial physiology. It is also of central importance for applications in biotechnology, where often the goal is to limit or even arrest growth. Growth-arrested cells with a functional metabolism open the possibility to channel resources into the production of a desired metabolite, instead of wasting nutrients on biomass production. The objective of the RESET project, supported in the framework of the Programme d'Investissements d'Avenir (Section 8.2), is to develop novel strategies to limit or completely stop microbial growth and to explore biotechnological applications of these approaches.

A foundation result for growth control in bacteria was published in the journal *Molecular Systems Biology* at the end of 2015 [6]. In that publication, we described an engineered *E. coli* strain where the transcription of a key component of the gene expression machinery, RNA polymerase, is under the control of an inducible promoter. By changing the inducer concentration in the medium, we can adjust the RNA polymerase concentration and thereby switch bacterial growth between zero and the maximal growth rate supported by the medium. The publication also presented a biotechnological application of the synthetic growth switch in which both the wild-type *E. coli* strain and our modified strain were endowed with the capacity to produce glycerol when growing on glucose. Cells in which growth has been switched off continue to be metabolically active and harness the energy gain to produce glycerol at a twofold higher yield than in cells with natural control of RNA polymerase expression. Remarkably, without any further optimization, the improved yield is close to the theoretical maximum computed from a flux balance model of *E. coli* metabolism. This work has been continued in several directions in the context of the RESET project by Célia Boyat. Moreover, extending work on self-replicator models of bacterial growth, we have studied the production of metabolites by means of the growth switch from an optimal control perspective, in a paper that is currently being prepared for publication.

In a review published in *Trends in Microbiology* this year [19], we have put the scientific results mentioned above in a broader context. As illustrated by the synthetic growth switch, reengineering the gene expression machinery allows modifying naturally evolved regulatory networks and thereby profoundly reorganizing the manner in which bacteria allocate resources to different cellular functions. This opens new opportunities for our fundamental understanding of microbial physiology and for a variety of applications. We describe how recent breakthroughs in genome engineering and the miniaturization and automation of culturing methods have offered new perspectives for the reengineering of the transcription and translation machinery in bacteria as well as the development of novel *in vitro* and *in vivo* gene expression systems. In our paper, we review different examples from the unifying perspective of resource reallocation, and discuss the impact of these approaches for microbial systems biology and biotechnological applications.

LIFEWARE Project-Team

7. New Results

7.1. Strong Turing completeness of continuous CRNs

Participants: François Fages, Guillaume Le Guludec (former Member), Sylvain Soliman.

When seeking to understand how computation is carried out in the cell to maintain itself in its environment, process signals and make decisions, the continuous nature of protein interaction processes forces us to consider also analog computation models and mixed analog-digital computation programs. However, recent results in the theory of analog computability and complexity obtained by Pouly and Bournez, establish fundamental links between analog and digital computing. In [8] and [10], we derive from these results the strong (uniform computability) Turing completeness of chemical reaction networks over a finite set of molecular species under the differential semantics, solving a long standing open problem. Furthermore we derive from the proof a compiler of mathematical functions into elementary chemical reactions. We illustrate the reaction code generated by our compiler on trigonometric functions, and on various sigmoid functions which can serve as markers of presence or absence for implementing program control instructions in the cell and imperative programs. This makes it possible to start comparing our compiler-generated circuits to the natural circuit of the MAPK signaling network, which plays the role of an analog-digital converter in the cell with a Hill type sigmoid input/output functions.

7.2. Influence networks compared with reaction networks

Participants: François Fages, Thierry Martinez (former Member), David Rosenblueth (former Member), Sylvain Soliman, Denis Thieffry.

Biochemical reaction networks are one of the most widely used formalism in systems biology to describe the molecular mechanisms of high-level cell processes. However modellers also reason with influence diagrams to represent the positive and negative influences between molecular species and may find an influence network useful in the process of building a reaction network. In [11], we introduce a formalism of influence networks with forces, and equip it with a hierarchy of Boolean, Petri net, stochastic and differential semantics, similarly to reaction networks with rates. We show that the expressive power of influence networks is the same as that of reaction networks under the differential semantics, but weaker under the discrete semantics. Furthermore, the hierarchy of semantics leads us to consider a (positive) Boolean semantics without test for absence, that we compare with the (negative) Boolean semantics with test for absence of gene regulatory networks a` la Thomas. We study the monotonicity properties of the positive semantics and derive from them an algorithm to compute attractors in both the positive and negative Boolean semantics. We illustrate our results on models of the literature about the p53/Mdm2 DNA damage repair system, the circadian clock, and the influence of MAPK signaling on cell-fate decision in urinary bladder cancer.

7.3. Machine learning influence networks from data

Participants: Arthur Carcano, François Fages, Jérémy Grignard, Sylvain Soliman.

Automating the process of model building from experimental data is a very desirable goal to palliate the lack of modellers for many applications. However, despite the spectacular progress of machine learning techniques in data analytics, classification, clustering and prediction making, learning dynamical models from data time-series is still challenging. In [7], we investigate the use of the Probably Approximately Correct (PAC) learning framework of Leslie Valiant as a method for the automated discovery of influence models of biochemical processes from Boolean and stochastic traces. We show that Thomas' Boolean influence systems can be naturally represented by k-CNF formulae, and learned from time-series data with a number of Boolean activation samples per species quasi-linear in the precision of the learned model, and that positive Boolean influence systems can be represented by monotone DNF formulae and learned actively with both activation samples and oracle calls. We consider Boolean traces and Boolean abstractions of stochastic simulation traces, and show the space-time tradeoff there is between the diversity of initial states and the length of the time horizon, together with its impact on the error bounds provided by the PAC learning algorithms. We evaluate the performance of this approach on a model of T-lymphocyte differentiation, with and without prior knowledge, and discuss its merits as well as its limitations with respect to realistic experiments.

7.4. Shaping bacterial population behavior through computer-interfaced control of individual cells

Participant: Jakob Ruess.

Bacteria in groups vary individually, and interact with other bacteria and the environment to produce population-level patterns of gene expression. Investigating such behavior in detail requires measuring and controlling populations at the single-cell level alongside precisely specified interactions and environmental characteristics. In [1], we present an automated, programmable platform that combines image-based gene expression and growth measurements with on-line optogenetic expression control for hundreds of individual Escherichia coli cells over days, in a dynamically adjustable environment. This integrated platform broadly enables experiments that bridge individual and population behaviors. We demonstrate: (i) population structuring by independent closed-loop control of gene expression in many individual cells, (ii) cell–cell variation control during antibiotic perturbation, (iii) hybrid bio-digital circuits in single cells, and freely specifiable digital communication between individual bacteria. These examples showcase the potential for real-time integration of theoretical models with measurement and control of many individual cells to investigate and engineer microbial population behavior.

7.5. Balancing a genetic toggle switch by real-time feedback control and periodic forcing

Participants: Gregory Batt, Jean-Baptiste Lugagne, Melanie Kirch (former Member), Agnes Köhler (former Member), Sebastian Sosa Carrillo.

Cybergenetics is a novel field of research aiming at remotely pilot cellular processes in real-time with to leverage the biotechnological potential of synthetic biology. Yet, the control of only a small number of genetic circuits has been tested so far. Here we investigate the control of multistable gene regulatory networks, which are ubiquitously found in nature and play critical roles in cell differentiation and decision-making. Using an in silico feedback control loop, we demonstrate that a bistable genetic toggle switch can be dynamically maintained near its unstable equilibrium position for extended periods of time [2]. Importantly, we show that a direct method based on dual periodic forcing is sufficient to simultaneously maintain many cells in this undecided state. These findings pave the way for the control of more complex cell decision-making systems at both the single cell and the population levels, with vast fundamental and biotechnological applications.

7.6. Abstracting the dynamics of biological pathways using information theory: a case study of apoptosis pathway

Participants: Gregory Batt, François Bertaux (former Member), Sucheendra Palaniappan (former Member).

Quantitative models are increasingly used in systems biology. Usually, these quantitative models involve many molecular species and their associated reactions. When simulating a tissue with thousands of cells, using these large models becomes computationally and time limiting. In our paper, we propose to construct abstractions using information theory notions [3]. Entropy is used to discretize the state space and mutual information is used to select a subset of all original variables and their mutual dependencies. We apply our method to an hybrid model of TRAIL-induced apoptosis in HeLa cell. Our abstraction, represented as a Dynamic Bayesian Network (DBN), reduces the number of variables from 92 to 10, and accelerates numerical simulation by an order of magnitude, yet preserving essential features of cell death time distributions.

7.7. Long-term tracking of budding yeast cells in brightfield microscopy: CellStar and the Evaluation Platform

Participants: Gregory Batt, Artémis Llamosi (former Member).

With the continuous expansion of single cell biology, the observation of the behaviour of individual cells over extended durations and with high accuracy has become a problem of central importance. Surprisingly, even for yeast cells that have relatively regular shapes, no solution has been proposed that reaches the high quality required for long-term experiments for segmentation and tracking (S&T) based on brightfield images. In this contribution, we present CellStar, a tool chain designed to achieve good performance in long-term experiments [5]. The key features are the use of a new variant of parametrized active rays for segmentation, a neighbourhood-preserving criterion for tracking, and the use of an iterative approach that incrementally improves S&T quality. A graphical user interface enables manual corrections of S&T errors and their use for the automated correction of other, related errors and for parameter learning. We created a benchmark dataset with manually analysed images and compared CellStar with six other tools, showing its high performance, notably in long-term tracking. As a community effort, we set up a website, the Yeast Image Toolkit, with the benchmark and the Evaluation Platform to gather this and additional information provided by others.

7.8. Sensitivity estimation for stochastic models of biochemical reaction networks in the presence of extrinsic variability

Participant: Jakob Ruess.

Determining the sensitivity of certain system states or outputs to variations in parameters facilitates our understanding of the inner working of that system and is an essential design tool for the de novo construction of robust systems. In cell biology, the output of interest is often the response of a certain reaction network to some input (e.g., stressors or nutrients) and one aims to quantify the sensitivity of this response in the presence of parameter heterogeneity. We argue that for such applications, parametric sensitivities in their standard form do not paint a complete picture of a system's robustness since one assumes that all cells in the population have the same parameters and are perturbed in the same way. In the published contribution, we consider stochastic reaction networks in which the parameters are randomly distributed over the population and propose a new sensitivity index that captures the robustness of system outputs upon changes in the characteristics of the parameter distribution, rather than the parameters themselves [4]. Subsequently, we make use of Girsanov's likelihood ratio method to construct a Monte Carlo estimator of this sensitivity index. However, it turns out that this estimator has an exceedingly large variance. To overcome this problem, we propose a novel estimation algorithm that makes use of a marginalization of the path distribution of stochastic reaction networks and leads to Rao-Blackwellized estimators with reduced variance.

7.9. Recombinase-based genetic circuit optimization by integer linear programming

Participant: François Fages.

The rapid advancements of synthetic biology show promising potential in biomedical and other applications. Recently, recombinases were proposed as a tool to engineer genetic logic circuits with long-term memory in living and even mammalian cells. The technology is under active development, and the complexity of engineered genetic circuits grows continuously. However, how to minimize a genetic circuit composed of recombinase-based logic gates remain largely open. In [12], we formulate the problem as a cubic-time assignment problem and solved by a 0/1-ILP solver to minimize DNA sequence length of genetic circuits. Experimental results show effective reduction of our optimization method, which may be crucial to enable practical realization of complex genetic circuits.

7.10. Coupled models of the cell cycle and circadian clock

Participants: François Fages, Sylvain Soliman, Pauline Traynard (former Member).

Experimental observations have put in evidence autonomous self-sustained circadian oscillators in most mammalian cells, and proved the existence of molecular links between the circadian clock and the cell cycle. Some mathematical models have also been built to assess conditions of control of the cell cycle by the circadian clock, with applications to cancer chronotherapy optimization. However, recent studies in individual NIH3T3 fibroblasts have shown an unexpected acceleration of the circadian clock together with the cell cycle when the culture medium is enriched with growth factors, and the absence of such acceleration in confluent cells. In order to explain these observations , we have studied a possible entrainment of the circadian clock by the cell cycle through a regulation of clock genes around the mitosis phase. We developed a computational model in Biocham with a formal specification of the observed behavior in quantitative temporal logic to investigate the conditions of entrainment in period and phase. We showed that either the selective activation of RevErb-α or the selective inhibition of Bmal1 transcription during the mitosis phase, allowed us to fit the experimental data on both period and phase, while a uniform inhibition of transcription during mitosis seems incompatible with the phase data. In [6], we presented those results and some further predictions of the bidirectional model with a coupling in both directions.

MORPHEME Project-Team

5. New Results

5.1. DIC (differential-interference-contrast) microscopy

Participants: Lola-Xiomara Bautista Rozo, Laure Blanc-Féraud.

This work is made in collaboration with Simone Rebegoldi, Marco Prato and Luca Zanni are in the Dipartimento di Scienze Fisiche, Informatiche e Matematiche, Universita di Modena e Reggio Emilia, Modena, Italy.

he DIC (differential-interference-contrast) microscopy states the problem of image phase reconstruction which is ill-posed (under-determinated) and non-convex optimization problem. We have worked on the phase reconstruction from color images by optimization of a non linear least-squares-like discrepancy term regularized with a total variation functional. We have considered two different penalties, the first one being the total variation (TV) functional which is suitable for piecewise constant images, while the second is the hypersurface (HS) potential, which is a smooth generalization of the TV able to reconstruct both sharp and smooth variations of the unknown phase. Since the latter choice leads to the minimization of a smooth functional, we developed a limited memory gradient method, in which suitable adaptive steplength parameters are chosen to improve the convergence rate of the algorithm. As concerns the TV–based model, we addressed the minimization problem by means of a recently proposed linesearch–based forward–backward method able to handle the nonsmoothness of the TV functional. Numerical tests show that in the case of smooth TV minimization funcitonal, the performance of the limited memory gradient method is much better than those of the conjugate gradient approaches proposed in the literature, in terms of number of function/gradient evaluations and, therefore, computational time. In the case of TV functional, despite the difficulties due to the presence of a nondifferentiable term, also the linesearch–based forward–backward method proposed in this case is able to provide reconstructed images with a computational cost comparable to that of the gradient methods, thus leaving to a potential user freedom to choose the desired regularizer without losing in efficiency.

This work has been done during the PhD thesis of Lola Bautista defended in June 2017 [1]. It has been published in the journal Inverse Problems in 2017 [4].

5.2. Towards a continuous relaxation of the $\ell_2 - \ell_0$ constrained problem

Participants: Gilles Aubert, Arne Henrik Bechensteen, Laure Blanc-Féraud.

We focus on the problem of minimizing the least-squares loss function under the constraint that the reconstructed signal is at maximum k-sparse. This is called the ℓ_2-ℓ_0 constrained problem. The minimization problem is of interest in signal processing, with application to compressed sensing, source separation and super-resolution imaging.

This problem has previously been relaxed, among other methods, by using the convex ℓ_1 norm instead of the ℓ_0 norm, but depending on the specific problem the global minimizer may not be the same.

The goal of our work is to propose a continuous exact relaxation of the ℓ_2-ℓ_0 constrained problem. The initial problem is non-continuous and is therefore from an algorithmic point of view difficult to minimize. A continuous exact relaxation has the same global minimizers as the initial problem, and a local minimizer of the relaxation is a local minimizer of the initial problem, with possible less local minimizers than the initial problem. Solving the initial ℓ_2-ℓ_0 constrained problem is equivalent, in the sense of the global minimizers, to solving the continuous relaxed form. Furthermore, a continuous exact relaxation provides better properties for the objective function in terms of minimization, because of the continuity and the number of local minimizers.

Based on the recent works of Marcus Carlson [17] we propose a continuous exact relaxation of the ℓ_2-ℓ_0 constrained problem S_γ, with an algorithm to minimize the function.

In order to increase the quality of the optimization, we have to chose “the best” exact relaxation. Inspired by the work by Emmanuel Soubies [23] we have computed the convex hull of the initial problem for a special case. The penalty term obtain, f_{cr} may be a continuous relaxation with respect to the initial problem, with fewer local minimizers than the initial problem and the relaxation S_γ (see figure 1). This has to be proven.

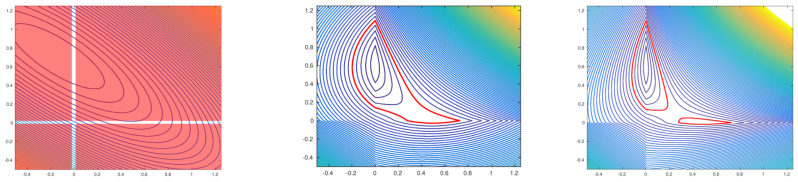


Figure 1. From the left to the right. The initial problem, the relaxation using f_{cr} and the relaxation using S_γ. The level lines of the relaxations are illustrated with a common level line marked in red.

The work will be presented at Mathematical Image Analysis 2018 conference in Berlin on the form of a poster.

5.3. Reconstruction of mosaic of microscopic images

Participants: Kévin Giulietti, Eric Debreuve, Grégoire Malandain.

This work takes place within the ANR PhaseQuant.

In microscopy imaging, a trade-off has to be made between a high resolution, that enables to see details, and the width of the field of view, that enables to see many objects. Such a trade-off is avoided by mosaicing, which consists in the acquisition of several images, say $N \times N$, with a small overlap between images. This way, an image with a N larger field of view can be reconstructed with the same resolution than a single microscopic image.

Such an imaging protocol is available on many microscopy software. Basically, displacements of the table on which lies the material to be imaged are programmed, and used to reconstruct the mosaic. However, it appears (at the overlapping areas), that a residual offset is still present. The cause of this has not be identified so far: this may be due to small geometric mis-alignement in the imaging device, or to the command of the micrometer table.

We thus investigate the stability of this residual offset with respect to time and to the image position within the mosaic.

5.4. Detection of cytoneme

Participants: Christelle Requena, Xavier Descombes.

This work is made in collaboration with Pascal Thérond, Tamas Matusek and Caterina Novelli (iBV). It is supported by the ANR project HMOVE.

Cellular communication is one of the most important processes for understanding and controlling morphogenesis (the set of laws that determine the structure of tissues and organs during embryonic development) necessary for the development of an organism. This is an important issue in the field of developmental biology and it has recently been shown that the exchange of information between cells is controlled by long cellular extensions called "cytonemes".

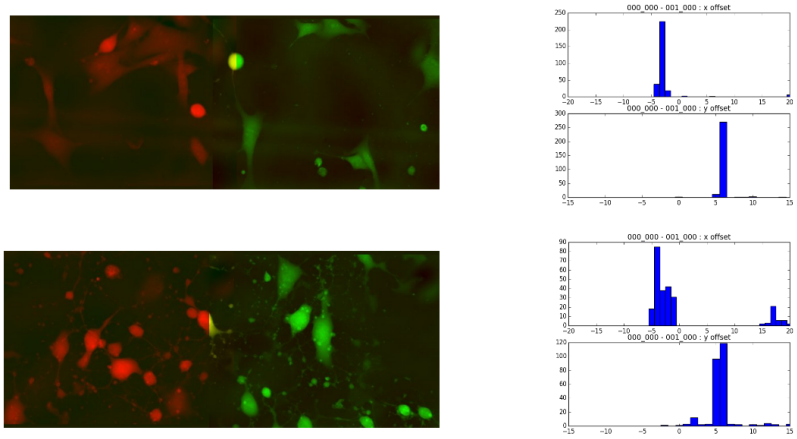


Figure 2. Example of mosaics reconstructed with two different datasets whose pairs of images were acquired at the same positions (0,0) in red and (1,0) in green. Histograms represent the offsets for all offsets overtime.

Due to the amount of information to be processed and the time required to study this information, it is essential to be able to provide image processing tools through which reliable, automatic and effective methods are proposed for these studies. In this work we have developped a pipeline for membrane extension and vesicles detection from in vivo data obtained by confocal microscopy. The vesicles are detected using a marked point process modeling. The cell extension detection embed the membrane detection using active contours and the filament detection using a tophat operator, the Frangi filter and Dijkstra algorithm. With this detection tool (exemplified in Figure 3), we have characterized a mutant population compared to a wild population of drosophila wings with respect to Hedgehog signalization. Interestingly we have shown that a significative difference appears in the cytonemes length but not in their number.

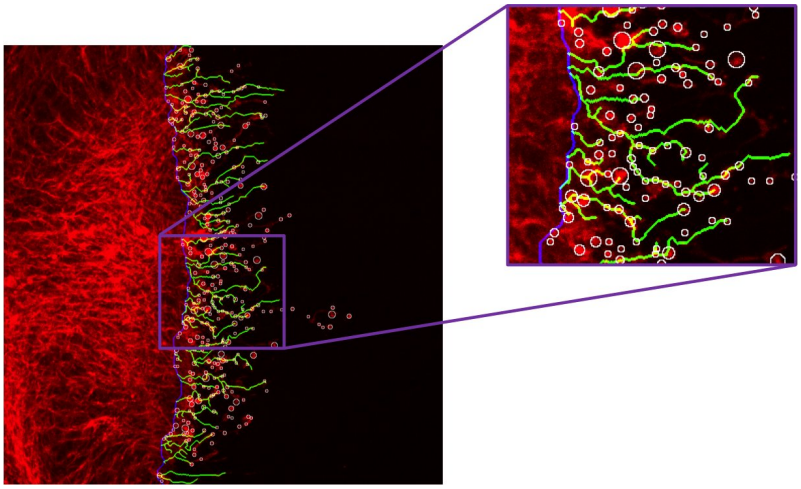


Figure 3. Cytoneme (green filaments) and Hedgehog vesicles detection (white circles).

5.5. 3D+t segmentation of single growing axons

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

Our work is motivated by the study of developmental axonal remodeling, a genetically-controlled process characterized by a degeneration step followed by a rapid regrowth of axons. Here, we focus our interest on the axonal regrowth phase, which can be studied during brain development, using the fruit fly, *Drosophila melanogaster*, as a model system.

During the regrowth, small dynamical branches can be observed: they emanate from long stable branches and have generally a short lifetime. Such small branches may contribute to rebuild the axon connectivity during the adult stage. A better knowledge of the mechanisms controlling the dynamic of these branches may contribute to a better understanding of neuronal morphogenesis. In this work, we are particularly interested in the quantification of this process, for which the extraction of both the main and second branches is required.

Neuron tracing is still a challenge in neuroinformatics. Despite the huge progresses made during the last decades, this problem is still an open question. This is exacerbated with the development of new imaging techniques, that produce more and more images with improved quality and/or resolution. Among these, live-imaging techniques are more and more prominent. Indeed, acquisitions of 3D image sequences over long periods of time, in particular, have enabled neurobiologists to follow complex processes such as the development of neuronal populations. However, they produce time series of 3D volumes, for which there does not exist dedicated tracing approach.

Apart slight movements, the dynamic changes of axons are due to growing or retracting branches. Thus, we designed a topologically constrained tracking method that first ensures that the tree structure of the axon and its branches is preserved through the time sequence, and second enables a slight displacement of the axon (within an user-specified extend), while mimicing both the retraction and the growth of branches. Results are presented in figure 4 .

5.6. Detection and characterization of mitochondrials networks

Participants: Kévin Giulietti, Xavier Descombes.

This work is made in collaboration with Frédéric Bost, Stephan Clavel, Aurélie Charazac, Celia Decondé le Butor (C3M).

We consider in this project a high content microscopy based screening focused on the effects of endocrine disruptors on prostatic cancer cells metabolism. Specifically, we developed our automatic computational tool to detect and classify mitochondrial network morphology from microscopy acquired images. The first step consists in binarizing the image and the binary pattern representing the mitochondrial network is classified in a second step. To binarize the mitochondrial network we consider the different level sets in the original image. A score is computed on each connected component of the level set pyramid depending on the contrast between the component and the neighboring background and on a shape criteria. We thus select the best scored component considering a compromise between the component contrast and a shape prior. We then run a k-mean clustering on the set defined by all the mitochondrial component extracted from the whole database. The different estimated classes are typical mitochondrial network element such as filaments or blobs. An image is then classified based on its signature defined by the number of mitochondrial element detected for each of the pre-defined classes (see Figure5). This classification scheme provides a discrimination framework based on geometrical and topological mitochondrial network properties than can differentiate for example filamentous and aggregate networks. This tool will be used for automatically specifying the effect of endocrine disruptors.

5.7. Detection and classification of neuronal extensions on fluorescence microscopy images: application to the study of metabolic diseases such as obesity or anorexia

Participants: Sarah Laroui, Eric Debreuve, Xavier Descombes.

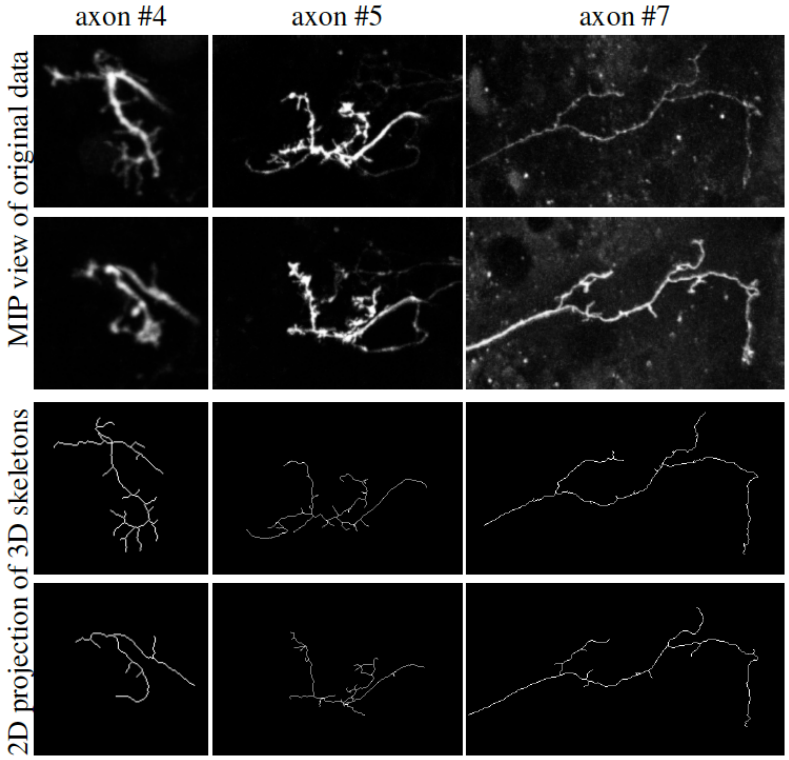


Figure 4. From top to bottom: MIP view of the first time point, MIP view of the last time point, 2D projection of the skeleton of the first time point, 2D projection of the skeleton of the last time point (series are made of 170 time points, with a 5 min time interval). Loops in skeleton projection views are projection artifacts.

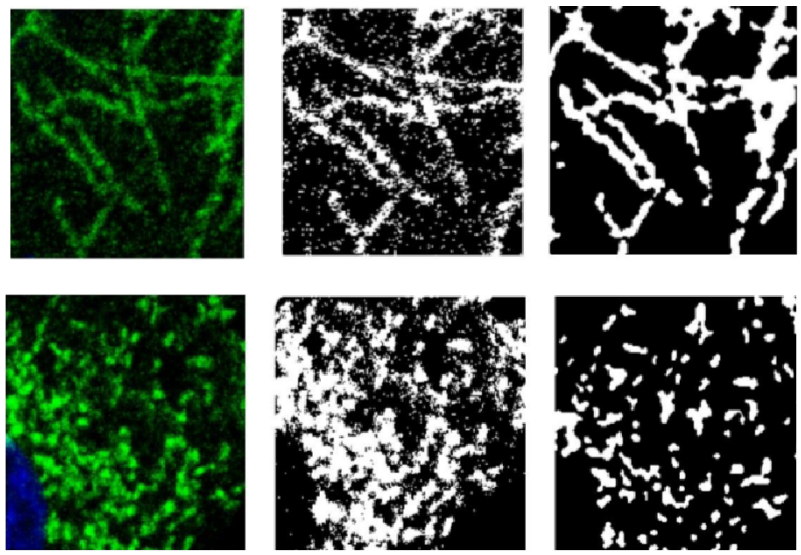


Figure 5. Detection of mitochondria filamentous/tubular (zoomed) network (top row), hyperfilamentous network (middle row) and aggregates network (bottom row). From left to right panels we have first input images (in green mitochondrial networks and in blue nuclei of the cells). Then binaries masks of mitochondrial networks using an automatic threshold. Then binaries masks resulting from our own developed method. Finally, classification of mitochondrial networks : in blue the filamentous/tubular forms, in green the hyperfilamentous form and in red the blobs forms.

This work is made in collaboration with Céline Cansell and Carole Rovere (IPMC, Sophia Antipolis).

The goal of this project is to classify 3D images of neuronal cells (astrocytes and microglia) into mice fed normally and mice fed with a high-fat diet (see Fig. 6). The distinction can be made in two different areas of interest of the hypothalamus: Median Eminence (EM) and Arcuate Nucleus (ARC).

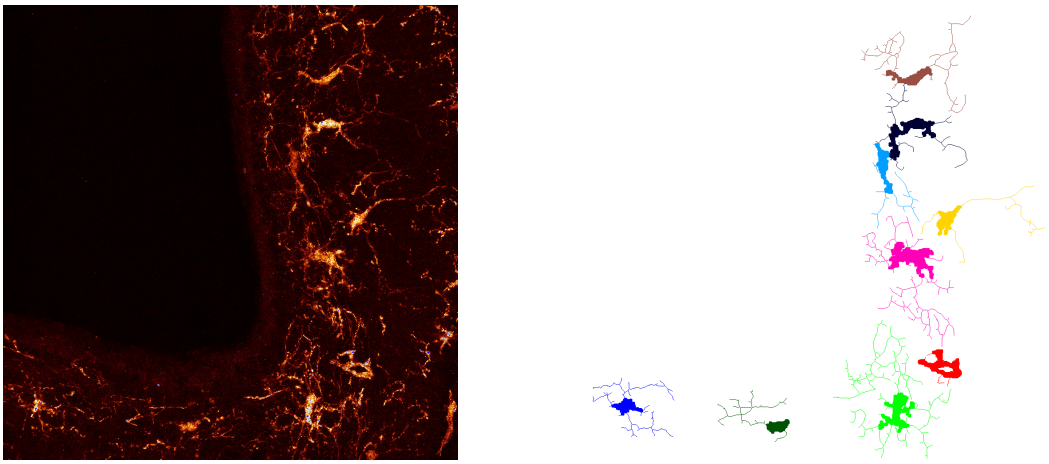


Figure 6. Maximum intensity projection (MIP) of the original image of microglias (left) and MIP of the network detection result (right).

Astrocytes are perceived as networks. Our goal is to find out if there is a difference in the organization of these networks between the two areas of interest and between the two mouse models. Regarding inflammatory cells (microglias), we first segment each cell body and their extensions using a Frangi filter bank to enhance filamentous structures. This produces network pieces that must be joined to build one network per microglia. Thus, we connect filaments to soma and filaments to filaments using minimal paths (using an image-based, anisotropic metric) computed by dynamic programming. Finally, we extract geometrical and topological parameters such as the length and width of the extensions, the number of branches ... These parameters will be used for clustering microglia networks in order to identity the different populations.

5.8. Automatic recognition of fungi phenotype by extraction and classification of morphometric parameters

Participants: Sarah Laroui, Eric Debreuve, Xavier Descombes.

This work is made in collaboration with Aurelia Vernay (Bayer) as part of a contract with Bayer.

Botrytis cinerea is a reference model of filamentous phytopathogen fungi. Some chemical treatments can lead to characteristic morphological changes, or phenotypic signatures, observable with transmitted light microscopy (see Fig. 7), which could be associated with the molecule Mode of Action.

In this context, we developed a robust image analysis and classification method relying on morphometric characteristics to automatically detect fungi observed using transmitted light microscopy, and classify them into predefined phenotypes. The detection task has been implemented in a classical way using a combination of mathematical morphology operations and active contours. The classification task has been solved in a supervised learning context.

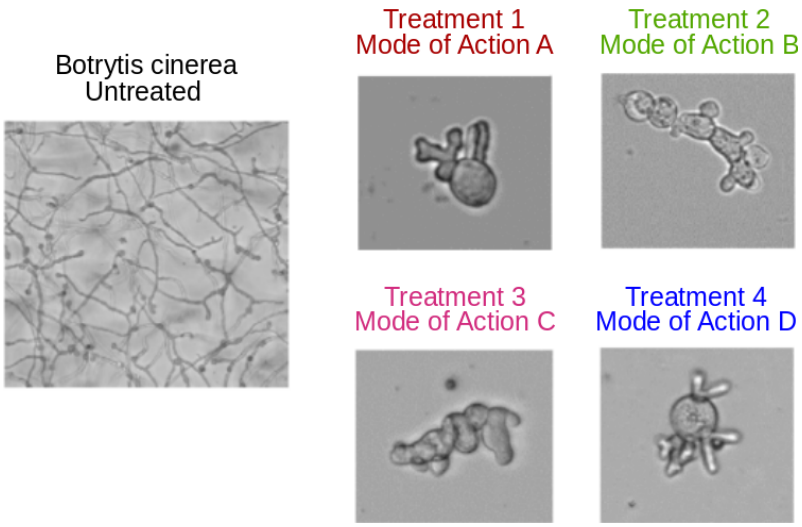


Figure 7. Characteristic phenotypic signatures for different chemical treatments (transmitted light microscopy, ImageXpress microscope, 10x lens).

Since a fungus can be described as tubular extensions connected to a spore (a roundish "root" cell), we proposed to describe such an object by its skeleton together with the distances from the skeleton to the fungus boundary. The skeleton was then converted into a valued graph. We selected a dozen topological and morphological features such as the number of nodes, the length of the longest branch, or the average and variance of the per-branch average skeleton-to-boundary distances.

These features were used in a supervised machine learning framework. Specifically, a cascade of two classifiers was proposed, the first one based on a decision tree to reject non relevant phenotypes (spores and mycelium), the second one to actually determine the phenotypes of the fungi. This second classifier was a Random Forest learned on the provided learning set composed of sample fungi from two phenotypes. Note that the classification accuracy can be computed either in a per-fungus way, or in a per-image way. Indeed, a given image corresponds to a unique chemical treatment so that all the fungi it contains exhibit the same phenotype (up to the natural biological variations), which can therefore be associated to the image itself. This per-image phenotype can be obtained by a majority vote among the individual fungus phenotypes. It represents the answer the biologists need. For the 2-phenotype problem we worked on, we obtained an image classification accuracy of around 90%, which is more than encouraging. In order to allow for a future, deeper analysis of the features characterizing each phenotype, we also computed the influence of each feature on the classification accuracy.

5.9. Density and repartition of cytoplasmic RNP (RiboNucleoprotein Particles) granules containing the Imp protein

Participants: Eric Debreuve, Xavier Descombes.

As part of the ANR project RNAGRIMP [0] (section 7.2.1), two series of images have been acquired using fluorescence microscopy: one where the cell cytoplasm has been stained with GFP (Green Fluorescent Protein), the second where the nuclei have been stained with DAPI (4',6-diamidino-2-phenylindole). The first steps are detecting the nuclei on the DAPI images and learning a classification procedure into living cell

[0] Imp = IGF-II mRNA-binding protein; IGF = Insulin-like Growth Factor; mRNA = Messenger Ribonucleic Acid.

or dead cell based on morphological and radiometric nuclei properties (average intensity, area, granularity, circularity...) (see Fig. 8).

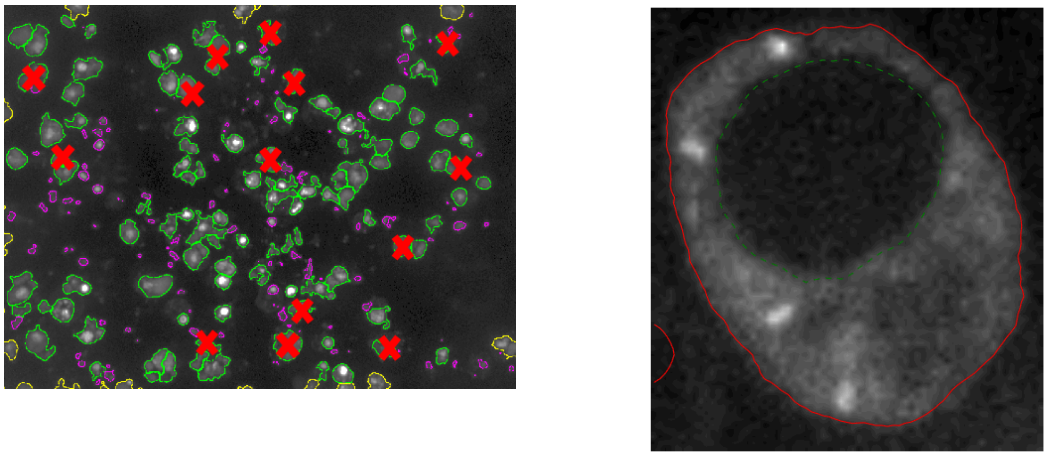


Figure 8. (left) Automatic classification of the detected nuclei into living (encircled in green) or dead (with a red cross). Objects encircled in yellow are cropped by the field of view, and objects encircled in purple are too small ; they are all discarded. (right) Active contour segmentation of the cytoplasm of a cell (previously classified as a living cell). Red contour: cytoplasm external boundary. Green, dashed contour: nucleus boundary (also cytoplasm internal boundary).

A specific CellProfiler [0] pipeline has been developed for this, and CellProfiler Analyst [0] has been used to learn a decision tree for automatic nuclei (hence, cell) classification. The next step is to segment (i.e., extract automatically the region of) the cell cytoplasms on the GFP images. Indeed, the target RNP-IMP granules appear in that compartment of the cell and are visible through their GFP response. We developed an active contour-based segmentation method relying on local image contrast with an initialization provided by a marked point process detection of ellipses [18] (see Fig. 8). Then, the detection of the particles can be performed inside the segmented cytoplasms (using a method called SPADE previously developed by the team).

5.10. Renal cell carcinoma classification from histopathological images

Participants: Mohammed Lamine Benomar, Nilgoon Zarei, Eric Debreuve, Xavier Descombes.

This work is made in collaboration with Damien Ambrosetti (MD, Pasteur Hospital, Nice).

The renal cell carcinoma is the most frequent type of kidney cancer (between 90% and 95% of all cases). Twelve classes of carcinoma can be distinguished, among which the clear cell carcinoma (CCRCC) and the papillary carcinoma (PRCC) are the two most common ones (75% and 10% of the cases, respectively). After the carcinoma has been diagnosed, the tumor is ablated and prepared for histological examination (fixation, staining, slicing, observation with a microscope) (see Fig. 9).

Along with genetic tests and protein reactions, the histological study allows to classify and grade the tumor in order to make a prognosis and monitor the patient treatment. Clinically speaking, digital histology is a recent domain (routinely, histological slices are studied by MDs directly on the microscope). The classical works on digital histology deal with the automatic analysis of cells (size, density ...). However, one crucial factor

[0] http://cellprofiler.org
[0] http://cellprofiler.org/cp-analyst

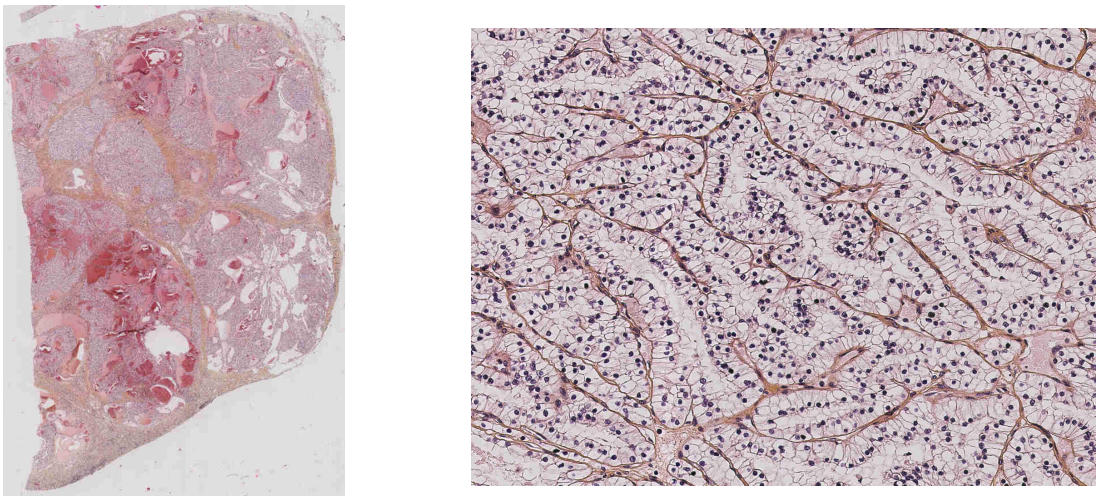


Figure 9. A histological slice through a kidney tumor: the whole slice (left) and a close-up (right) (the vascular network has a brownish color; the cell nuclei have a dark violet color).

for carcinoma classification is the structure of the vascular network. Coarsely, CCRCC is characterized by a "fishnet" structure while the PRCC has a tree-like structure.

In this context, we proposed to extract the vascular network from a given histological slice, compute features of the underlying graph structure, and classify the tumor into CCRCC or PRCC based on these features [24]. Then, we started to focus on performing a higher-level analysis of the vascular graphs. It can be noted that cells that are close to the vascular network naturally tend to align with it. Thus there might be specific "cell-vascular network" arrangements for each type of carcinoma. Our plan is to look for repeated subgraph patterns using pattern matching methods on labeled graphs, where a pattern would be a combination of (i) topological features from the graph, (ii) nearby cell features, and (iii) measures characterizing the coherence between nearby cells and the network (cell-to-network distances, cell density along the network, degree of alignment with the network...). There are chances that each carcinoma type exhibits a set of patterns that appear with a high frequency, therefore being characteristic of the given type. Such patterns would then represent discriminant features for carcinoma classification.

5.11. Comprehensive comparison of multi-labeled images

Participants: Gaël Michelin, Grégoire Malandain.

The data used for this work are courtesy of Yassin Refahi (Sainsbury Laboratory, Cambridge university) and Ulla-Maj Fiuza (CRBM, CNRS, Montpellier 1 & 2 university).

In the context of developmental biology, 3D+t microscopy imaging allows to quantitatively study the morphogenesis at the cellular level, but requires automated segmentation methods to handle the huge quantities of data. To minimize the necessary and tedious user interaction to correct unavoidable errors (3D images may have up to thousands of cells), it is desirable to improve such segmentation methods. This, in turn, motivates the need for a comprehensive evaluation methodology that will allow to automatically compare the outputs of two segmentation methods, not only in terms of cell border accuracy, but also in terms of cell detection.

The aim of the present work is to propose such an original comprehensive segmentation comparison method that provides an objective way for multi-object segmentation comparison. This method enables to determine automatically a region-to-region correspondence map and provides asymmetric shape similarity indexes

between two segmented images, with a robustness to potential region border variations. We illustrate the applicability of the proposed method with two examples in figure 10 .

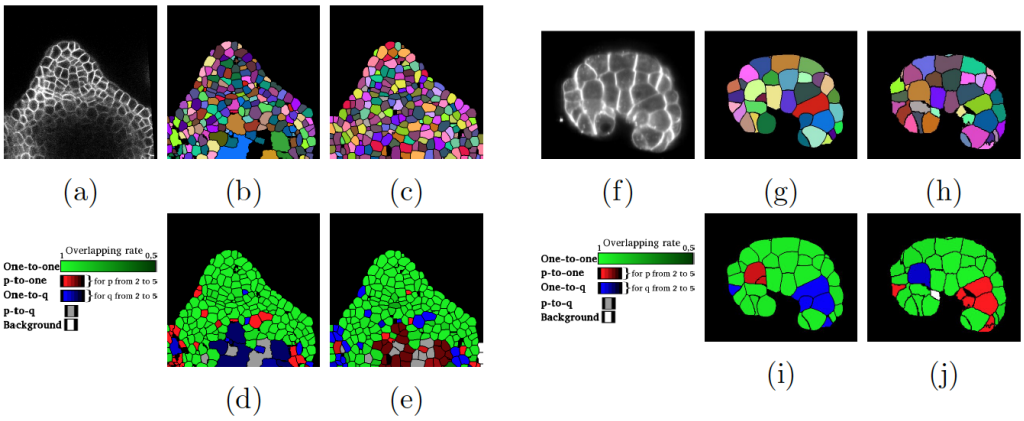


Figure 10. Cut-views of original 3D intensity images (a,f), the associated pairs of corresponding segmentations (b-c,g-h) and the results of regions association for each segmentation determined by the proposed method with the proposed method. (a-e) Floral meristem image. (f-j) Ascidian image.

5.12. Grouped Local Automated Cell Extractor (GLACE)

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Julien Laussu, Patrick Lemaire (CRBM, CNRS, Montpellier 1 & 2 university), Emmanuel Faure (IRIT, CNRS, Toulouse) and Christophe Godin (Inria Virtual Plants team, Montpellier).

In developmental biology, the embryogenesis study relies in particular on image-based studies. Today, fluorescent confocal microscopy is a means for *in vivo* imaging of developing organisms at cell level with a high spatio-temporal resolution. To handle such 3D+t image sequences, adapted computer-assisted methods are highly desirable in order to extract essential information from these data.

More specifically, for developing ascidian embryos, an existing framework called `ASTEC` [19] is used by biologists in order to extract the cell segmentation and lineage from some 3D+t sequences. However, remaining issues about segmentation accuracy motivated us to propose a new framework as an alternative to `ASTEC` for cell segmentation and tracking. The originality of the proposed Grouped Local Automated Cell Extractor (`GLACE`) framework is to segment the i–th image of a sequence by applying *locally* the original 3D cell segmentation framework of [21] for all the regions of interest defined by the segmented cells of the $i-1$–th image of the sequence. The union of all the local reconstructions provides the segmentation of the i–th image of the sequence (figure 11). The `GLACE` framework does not replace the `ASTEC` framework, however they provide complementary results for embryo image sequence reconstructions.

5.13. Ascidian embryo cell lineage registration in 3D+t image sequences

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Julien Laussu, Patrick Lemaire (CRBM, CNRS, Montpellier 1 & 2 university) and Christophe Godin (Inria Virtual Plants team, Montpellier).

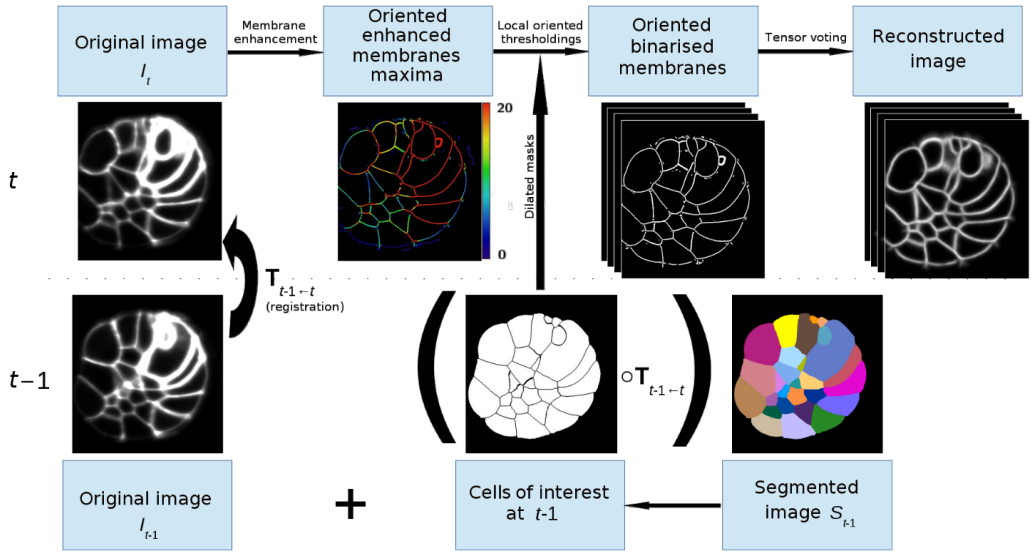


Figure 11. Pipeline for GLACE.

Until gastrulation, ascidian embryos have a very stereotyped and invariant development, so that it is possible to establish a cell-to-cell mapping between two developing embryos at a same developing stage. We proposed in a previous work a method for geometric registration that determines a linear (affine) transformation superimposing a test embryo into a reference one and that draws a cell-to-cell mapping up [20].

In the current work, we extend this framework for the determination of cell lineage mapping between two developing ascidian embryos by propagating an initial cell-to-cell mapping to the cell descendants since the cell correspondences are inherited for the ascidian embryo (figure 12 (top)). To do so, we use the information provided by the 3D+t sequences segmentation and lineage such as cell volume, life-span and relative position in the embryos. We experimented on real data the proposed cell lineage registration framework (figure 12 (bottom)).

5.14. Towards construction of digital atlases of plant tissues

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Yassin Refahi (Sainsbury Laboratory, Cambridge university), Jonathan Legrand, Jan Traas (RDP, ENS Lyon, INRA, CNRS, Lyon) and Christophe Godin (Inria Virtual Plants team, Montpellier).

In developmental biology, the study of model organisms aims for the understanding of genetic mechanisms responsible of morphogenesis. Today, fluorescent confocal microscopy is a means for in vivo imaging of developing plants at cell level with a high spatio-temporal resolution.

We propose in this work some dedicated computational tools for the study of such 3D+t sequences. These methods offer the means to compare temporal sequences of flower development and to build 4D digital atlases of developing arabidposis floral meristems on which every individual can be projected (figure 13), opening the avenue to the statical analysis of populations.

5.15. 3D Coronary vessel tracking in x-ray projections

Participants: Emmanuelle Poulain, Grégoire Malandain.

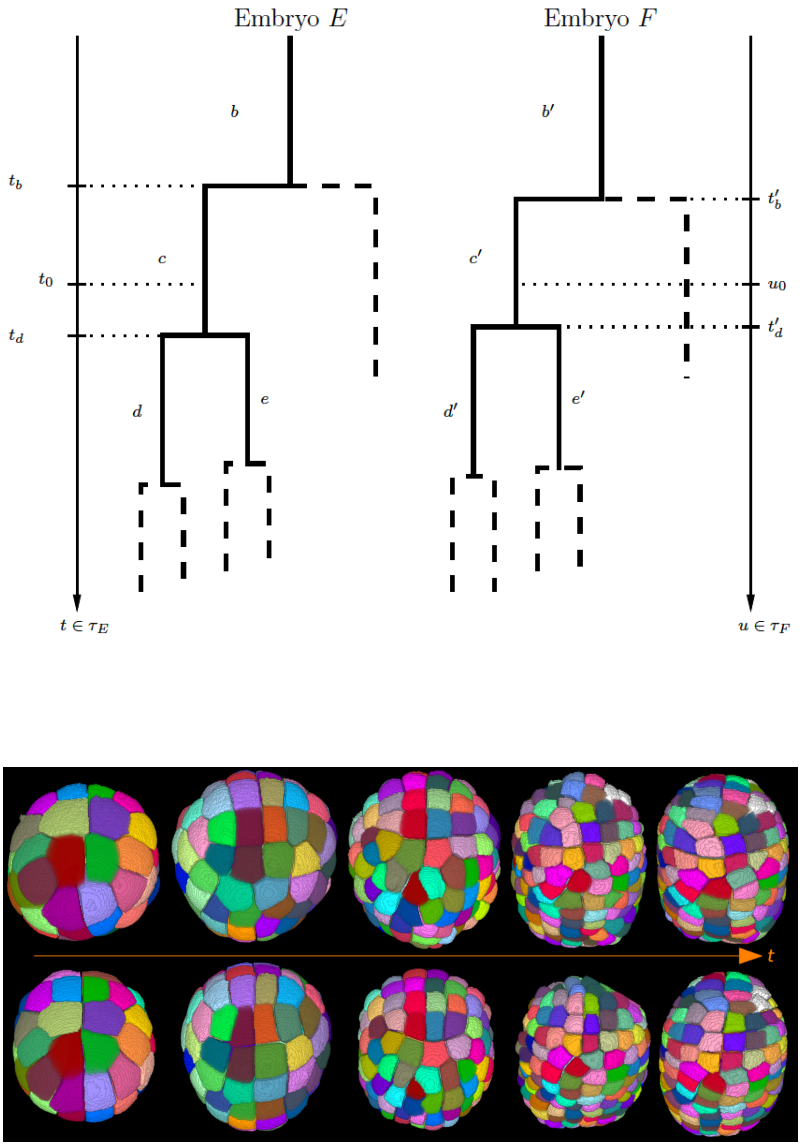


Figure 12. Ascidian embryo cell lineage registration. Top: sub-lineages from embryos E and F showing labels c and c' in correspondence with their birth (t_b) and death (t_d) (respectively t'_b and t'_d) time-points, mother cells (b and b') and daughters ((d, e) and (d', e')) along embryo lifespans τ_E and τ_F. Bottom: result of lineages registration between two developing embryos. Mapped cells appear with the same color. Cells in white are those for whom no corresponding cell was found in the other embryo. First column: cell-to-cell initial mapping.

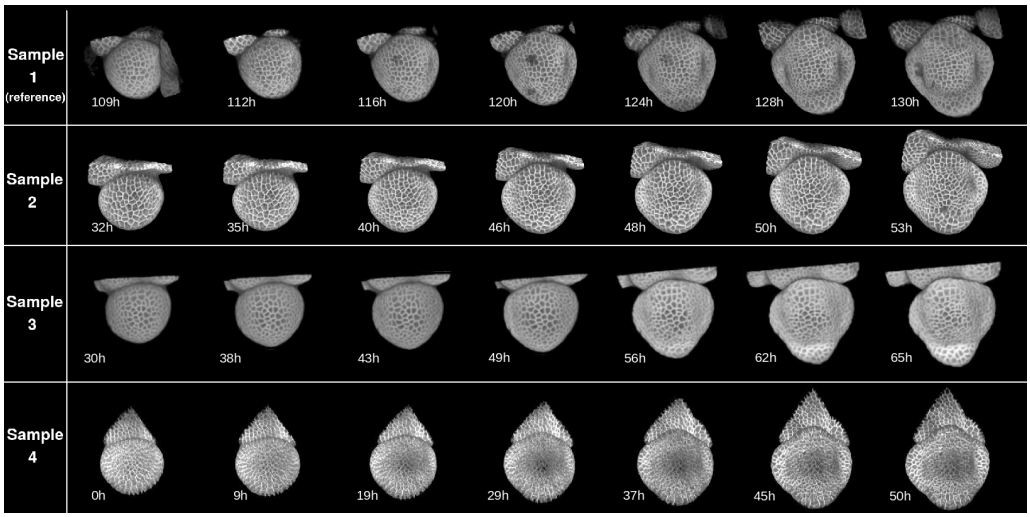


Figure 13. Visualization of the valid spatio-temporal sample alignments following the proposed registration method at different floral meristem developmental phases. One can observe the reliability of the registration method to identify developmental phases equivalences between the different samples.

This work is made in collaboration with Régis Vaillant (GE-Healthcare, Buc, France) and Nicholas Ayache (Inria Asclepios team).

Percutaneous Coronary Intervention (PCI) is a minimally procedure which is used to treat coronary artery narrowing. During the guidewire navigation, the lesion is crossed and in some cases, the physician could benefit from a visual assessment of the coronary wall. The x-ray imaging interventional system used for per-operative guidance is not able to display this information mostly by lack of density resolution. On the contrary, Computed Tomography Angiography (CTA) is a modality which has the capability of capturing the characteristics of the vessel wall.

Fusing pre-operative CT angiography with per-operative angiographic and fluoroscopic images is thus considered by physicians as a potentially useful tool for improved guidance. To be adopted, this tool has required the development of tracking methods adapted to the deformations of the arteries caused by the cardiac motion. We have proposed a 3D/2D temporal tracking of one coronary vessel, based on a spline deformation, using pairings with a controlled 2D stretching or contraction along the paired curves and a preservation of the length of the 3D curve which corresponds to the anatomic propriety [8], [9]. Experiments were conducted on a database of 10 vessels from 5 distinct patients, with dedicated metrics assessing both the global registration and the local coherency of the position along the vessel. The proposed results demonstrate the efficiency of the proposed method, with an average standard deviation of 2 mm for the localization of landmarks (see Fig. 14).

5.16. Modelling axon growth from in vivo data

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

Axons develop embedded in mechanically constrained environments. Thus, to fully understand this dynamical process, one must take into account collective mechanisms and mechanical interactions within the axonal populations. However, techniques to directly measure this from living brains are today lacking or heavy to implement. This interdisciplinary work intends to close the gap between classic in vitro experimental assumptions and real in vivo situations, where the final neuronal morphology is acquired through a dynamical and environmental-dependent process. We use as biological model Drosophila γ axon remodeling and analyze,

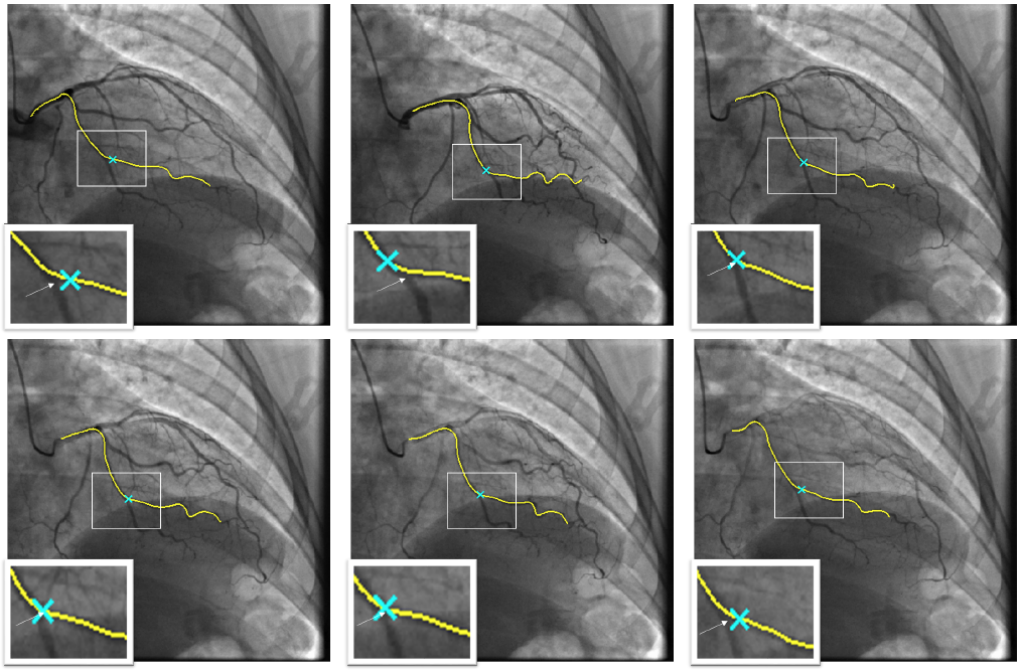


Figure 14. Visualization of the valid spatio-temporal sample alignments following the proposed registration method at different floral meristem developmental phases. One can observe the reliability of the registration method to identify developmental phases equivalences between the different samples.

for the first time to our knowledge, the mechanical situation of a whole population of γ neurons (650 individuals) growing together in a constraint space (i.e. medial lobe of the Mushroom Body).

We have designed a mathematical model of single axon growth based on Gaussian Markov Chains with two parameters, accounting for axon rigidity and attraction to the target field. We used this model to simulate the growing axons embedded in space constraint populations to test our hypothesis. We explored new branch formation mechanisms to mimic the growth of wild type γ axons population , as well as predict different mutant phenotypes. This approach allowed also to analyze dynamical aspects of the γ neuron collective growth process such as speed and density in function of space and time, which help to explain several characteristics of the γ neuron morphology and behavior during development. Among the obtained results, the proposed model is able to reproduce the intra-population morphological variability. Interestingly, applying the ESA distance between trees previously developed in the team [22] showed that real axons present shapes that showcase a compromise between collective elongation and morphological variability, essential for axonal connectivity (Figure 15). Finally, we explored other branch occurrence strategies –from uniformly random to occurrence upon mechanical interactions- to contrast and validate with previously developed hypothesis on the importance of branching for axonal elongation in vivo.

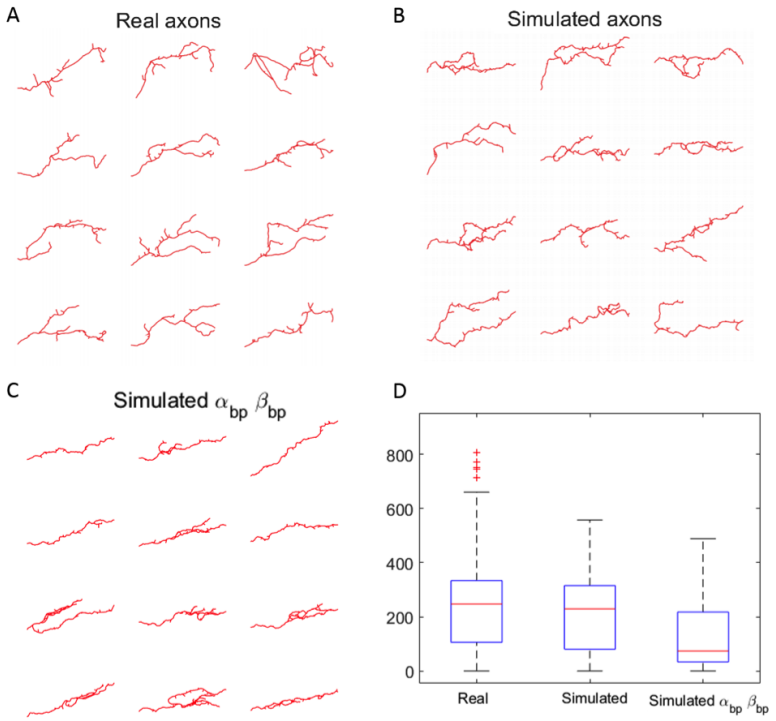


Figure 15. Impact of the parameter value on axonal morphologies. (A) Real wild type γ axons. (B) Axons simulated with parameters estimated from data. (C) Axons simulated with optimal parameters regarding collective elongation. (D) Intra-group variability measured with the ESA distance between all the axons in each group (A-C).

5.17. Jump point detection and parameter estimation from piecewise homogeneous Markov chains

Participants: Agustina Razetti, Xavier Descombes.

Piecewise homogeneous Markov chain processes can be applied to diverse phenomena of various nature, such as genetics, physics. Recent bibliography has focused on these systems, proposing different alternatives to detect the jump points and be able to separate between different phases of the signals. The Markov chain is usually defined by its transition matrix and the change points are modeled by a hidden Markov process. In this work, we focus on the Gaussian case with a Bernoulli distribution governing the change points. We have developped two different theoretical frameworks: one Bayesian with a Bernoulli prior, and the other one statistic-oriented, proposing a test of hypothesis based on ratio of likelihoods. For both cases we provide with robust algorithms to detect the jump points and reduce the error in the estimations of the parameters of the main model. We compare both methods and investigate their limits and advantages. We finally provided practical examples to showcase the power of the proposed approach (see Figure 16).

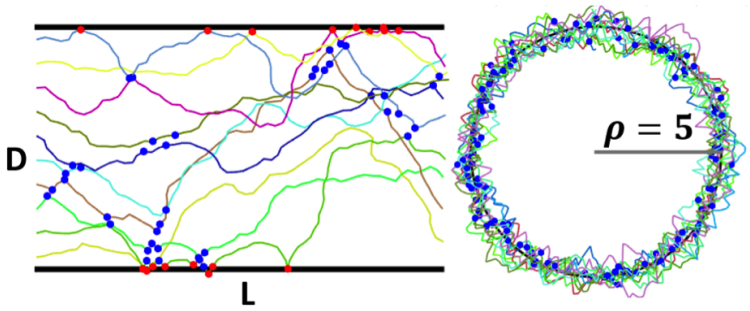


Figure 16. Two application examples. 10 particles of equal mass moving are shown at each case. When they collide another particle or the external limits, they follow elastic punctual collisions (shown by stars and circles). Left: particles inside a tube of diameter D and length L; right: particles moving around a circle of radius ρ.

PLEIADE Team

6. New Results

6.1. Alcyone system for repeatable e-science

One of PLEIADE's goals is to assist scientific users in deploying analysis software in their desktop environments. Increasingly, this is not a question of installing software packages locally, but of building bespoke environments that comprise many cooperating software tools. A typical example is a local Galaxy instance, communicating with a project-specific database that is shared with visualization and analysis tools, and cooperating with an electronic notebook such as Jupyter. In order to foster repeatable science, the configuration of each such environment should be reliably recorded, in a way that allows it to be redeployed in the future or shared with a colleague.

PLEIADE's **Alcyone** system provides a mechanism for specifying and deploying software environments for scientific users in bioinformatics and biodiversity. Alcyone offers three facilities:

1. A *specification* using configuration-by-convention style, combining specification files in YAML format and raw data files.
2. A *collection of Dockerized services* that can be chosen in the specification.
3. A *deployment system* that compiles the specification into a master container image, which orchestrates the deployment and management of the service containers.

The user's environment is fully specified in files that can be archived and shared, allowing future reuse. The use of Docker containers guarantees that future deployments run exactly as before, since the precise versions of the service containers are recorded.

Furthermore, Alcyone specifications are files, that can be managed by the Git source code control system. Different versions of the environment, including different analysis pipelines and intermediate results, are stored in the Git history and any version can be resurrected and deployed. Git branches can also be used to share configurations between users in the same lab.

Alcyone is being tested internally by PLEIADE and is undergoing intense development. Existing service containers are PLEIADE's Magus knowledge base, Magecal gene prediction pipeline, and Mimoza metabolic network explorer ; as well as third-party tools Galaxy, Gbrowse, and Jbrowse.

6.2. Clavispora lusitaniae

Clavispora lusitaniae, an environmental saprophytic yeast belonging to the CTG clade of *Candida* and a teleomorph of *Candida lusitaniae*, is an environmentally ubiquitous ascomycetous yeast with no known specific ecological niche. It can be isolated from different substrates, such as soils, waters, plants, and gastrointestinal tracts of many animals including birds, mammals, and humans. In immunocompromised hosts, *C. lusitaniae* can be pathogenic and is responsible for about 1% of invasive candidiasis, particularly in pediatric and hematology-oncology patients.

The Laboratoire de Microbiologie Fondamentale et Pathogénicité UMR-CNRS 5234 and PLEIADE sequenced and annotated the genome of *C. lusitaniae* type strain CBS 6936, and analyzed it in comparison with the strains ATCC 42720, isolated from the blood of a patient with myeloid leukemia, and MTCC 1001, a self-fertile strain isolated from citrus. In spite of a conserved genome structure, the genomes have undergone significant divergence. In particular the SNP density of 1 SNP per 90 bp is twice the level observed between strains SC5314 and WO-1 of *Candida albicans*, which are members of different subgroups within the species and qualified as having diverged relatively recently.

This work contributes to PLEIADE's long-term goal of developing understanding how diversity measured at the genome level can be made to correspond with observed functional diversity.

6.3. Introgressions as a source of diversity

Several prominent mechanisms of genomic evolution have been described for the yeasts, among them interspecific hybridization, reticulated evolution, aneuploidization, recent or ancient poly-ploidization events, large chromosomal duplication or more limited gene duplication, and horizontal transfer. These mechanisms are usually so closely intertwined that it is difficult to determine which ones are causes or consequences. Regardless of mechanisms the result has been a drastic reshaping of yeasts genome along evolution. Understanding these mechanisms is important, not only for strain construction in biotechnology, but also more fundamentally for insight into the causes and effects of genome reshaping on much shorter time scales.

Introgression, the transfer of large or more limited genetic information from one species to another, is an evolutionary mechanism of particular interest in industrial applications such as wine making where large vat cultures are used. Introgression results in mosaic genomes, and can be the result of interspecific hybridization fol- lowed by the extensive loss of one parental genome, either through repeated backcross with one parental species or through missegregation of the hybrid at meiosis.

In collaboration with the Institut des Science de la Vigne et du Vin and Bordeaux Sciences Agro, PLEIADE developed tools to rapidly assess the presence of introgressed regions in a large population of *Saccharomyces uvarum* isolates (104 strains), focusing on Holarctic isolates from natural, cider and wine environments since introgressed regions are absent in Southern hemisphere isolates. The overall number of introgressed regions is significantly higher in cider-associated strains compared to wild strains, and is higher in wine isolates. However, only a subset of the introgressed regions were found to be overrepresented in anthropic activities and their number and quality varied between cider- and wine-making processes.

Paradoxically, the low Holarctic genetic diversity observed in [1] contrasts with the relative high phenotypic diversity found for technological traits. This contradiction suggests that interspecific introgressions found among Holarctic *S. uvarum* strains could be the most important source of genetic, and by extention of phenotypic, diversity.

6.4. New results Biodiversity

The activity of PLEIADE in computational biodiversity has consisted mainly in reinforcing a cooperation with actors in High Performance Computing, namely Inria team Hiepacs, for method developments in metabarcoding. Metabarcoding is a supervised or unsupervised statistical learning method, to build taxonomic inventories from so called environmental samples, i.e. sets of short reads of a same marker for a whole community or guild. Most of tools used therefore still rely on some classical ones shaped in Multivariate Data Analysis. Those tools are indeed well known, but still are often behind the scene in current developments in Machine Learning (like kernel PCA, Support Vector Machines, etc. ...). Most of them, if not all, are based on Singular Value Decomposition of a matrix. If p features are observed on n items, the size of the matrix is $n \times p$. The complexity of such algorithms is in $O(p^3)$. The recent development of NGS data has had as a consequence to multiply by a factor $10^2/10^3$ the size of data sets. This leads to a factor $10^6/10^9$ of required computation time. Reaching such a goal is beyond resources currently offered by parallelization. Hence, a new approach has been selected, by using other methods. Indeed, it has been known for some years now that concentration of measure phenomena (a sort of extension of law of large numbers) leads to a blessing of dimensionality, i.e. some randomized methods are available as heuristics to make some matrix computations efficiently and accurately. This is the case for running SVD. Therefore, a cooperation has been set up between HiePacs and PLEIADE through Pierre Blanchard (a former Hiepacs PhD student who has held a post-doc position during 7 months in PLEIADE) to implement those methods in the framework of metabarcoding. Former work in PLEIADE had led (with a DARI project 2014-2016) to the production of many high-dimensional pairwise distance matrices of DNA environmental samples (amplicon based metabarcoding). Classical Multidimensional Scaling of some of those matrices has been programmed in C++, with dedicated libraries in domain of so called random projection, or column selection (`fmr` library). This has permitted to build a point cloud of an environmental sample of 1.2×10^5 reads, and see its "shape", with eyes, from projections on first axis, and build a low dimensional approximation of it. The outcome is twofolds: (i) build a

point cloud attached to an environmental sample, for further ecological studies and (ii) delivery of a scientific library in High Performance Computing for randomized matrix computations. These research lines will be carried on in 2018, and the cooperation extended to mésocentre GRICAD in Grenoble for HPC and C++ code development.

PLEIADE has carried on statistical learning methods, both supervised and unsupervised in metabarcoding. A cooperation with IMBE at Marseille has permitted to associate MDS as developed above with graph based methods (building connected components of a graph built from pairwise distance matrices after thresholding), and test these methods for unsupervised statistical learning (OTU building) of data sets from an ongoing PhD in Marseille Bay. Cooperation with Institut Pasteur at Cayenne has lead to a joint publication [12] for a proof of concept of an inventory by metagenomics of viromes of bats in French Guiana, with two objectives: (i) detect as soon as possible some strains which could potentially be transmitted to man and (ii) develop a viral ecology by studying further how environmental factors and nature of the host drive the virome composition.

Meanwhile, PLEIADE has carried on cooperation with SLU Universty at Uppsala especially on metabarcoding of diatom communities in rivers and lakes in Sweden (co-direction of a PhD student located at Uppsala in SLU) , and first steps in biogeography of diatoms in Fennoscandia (cooperation with a PostDoc in SLU).

SERPICO Project-Team

7. New Results

7.1. Statistical methods for image denoising and reconstruction

Participants: Emmanuel Moebel, Charles Kervrann.

In the line of the Non-Local (NL) means [39] and ND-SAFIR [11], [12], [6] denoising algorithms, we have proposed a novel adaptive estimator based on the weighted average of observations taken in a neighborhood with weights depending on image data. The idea is to compute adaptive weights that best minimize an upper bound of the pointwise L_2 risk. In the framework of adaptive estimation, we show that the "oracle" weights depend on the unknown image and are optimal if we consider triangular kernels instead of the commonly-used Gaussian kernel. Furthermore, we propose a way to automatically choose the spatially varying smoothing parameter for adaptive denoising. Under conventional minimal regularity conditions, the obtained estimator converges at the usual optimal rate. The implementation of the proposed algorithm is also straightforward. The simulations show first that our algorithm improves significantly the classical NL-means. Second, the simulations demonstrate that it is competitive when compared to state-of-the-art denoisers both in terms of PSNR values and visual quality.

Meanwhile, we investigated statistical aggregation methods which optimally combine several estimators to produce a boosted solution [13]. This approach has been especially investigated to restore spectral information in the missing wedge (MW) in cryo-electron tomography (CET). The MW is known to be responsible for several types of imaging artifacts, and arises because of limited angle tomography: it is observable in the Fourier domain and is depicted by a region where Fourier coefficient values are unknown (see Fig. 3). The proposed stochastic method tackles the restoration problem by filling up the MW by iterating following steps: adding noise into the MW (step 1) and applying a denoising algorithm (step 2). The role of the first step is to propose candidates for the missing Fourier coefficients and the second step acts as a regularizer. A constraint is added in the spectral domain by imposing the known Fourier coefficients to be unchanged through iterations. Different denoising algorithms (BM3D, NL-Bayes, NL-means...) have been compared. Furthermore, different transforms have been tested in order to apply the constraint (Fourier transform, Cosine transform, pseudo-polar Fourier transform). Finally, we showed that this strategy can be embedded into a Monte-Carlo simulation framework and amounts to computing an aggregated estimator [13]. Convincing results have been achieved (see Fig. 3) using the Fourier Shell Correlation (FSC) as an evaluation metric.

References: [18]

Collaborators: Qiyu Jin (School of Mathematical Science, Inner Mongolia University, China),
Ion Grama and Quansheng Liu (University of Bretagne-Sud, Vannes),
Damien Larivière (Fondation Fourmentin-Guilbert),
Julio Ortiz (Max-Planck Institute, Martinsried, Germany).

7.2. Algorithms for dejittering and deconvolving fluorescence Tissue MicroArray (TMA) images

Participant: Charles Kervrann.

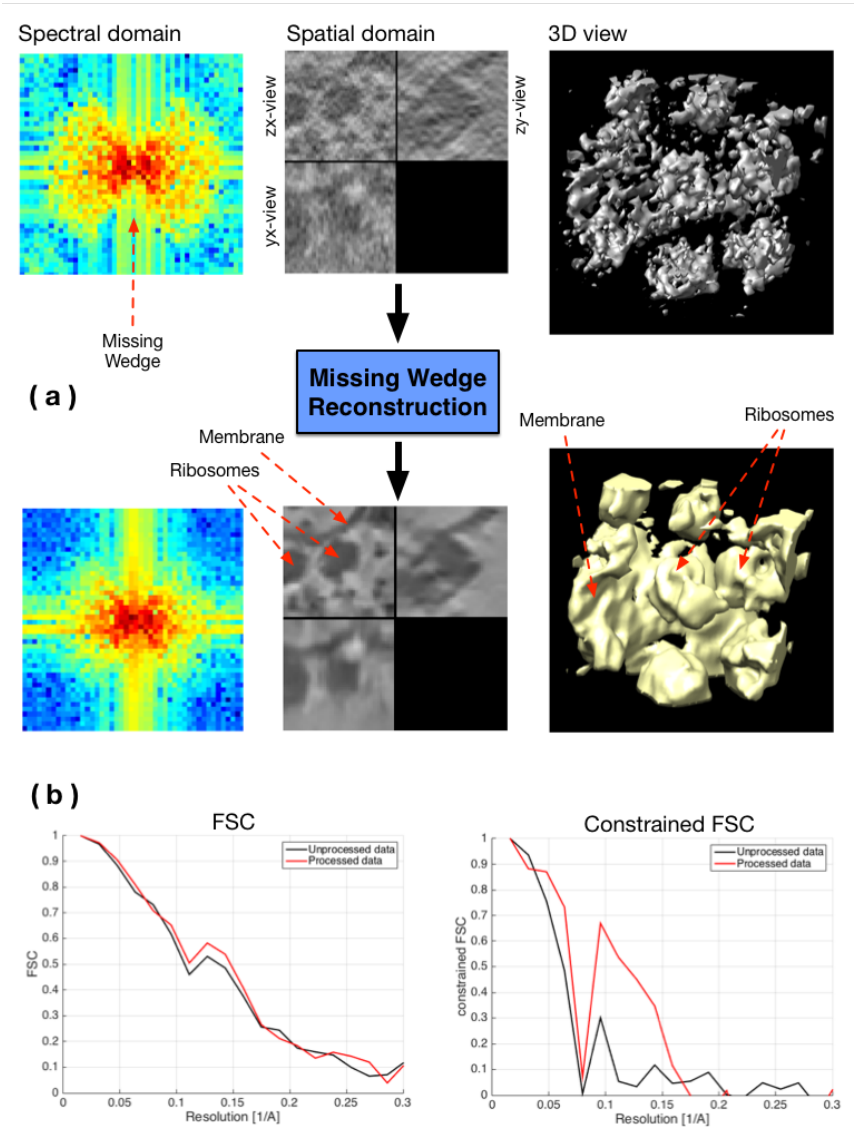


Figure 3. Experimental sub-tomogram containing ribosomes attached to a membrane. (a) Top row: original data in the spectral (left) and spatial (middle) domains and 3D view of the thresholded data (right). Bottom row: denoised data shown as above. (b) FSC and constrained FSC measures of the method input (in black) and output (in red). All measures are wrt the same reference.

In the thesis of H.-N. Nguyen, we developed dedicated image processing methods to improve quality of Tissue Microarray (TMA) images acquired by fluorescence scanners. Images are first acquired pixel by pixel along each line, with a change of scan direction between two subsequent lines. Such scanning system often suffers from pixel mis-positioning (jitter) due to imperfect synchronization of mechanical and electronic components. To correct these scanning artifacts, we proposed a variational method based on the estimation of pixel displacements on subsequent lines. This method, inspired from optical flow methods, consists in estimating a dense displacement field by minimizing an energy function composed of a non-convex data fidelity term and a convex regularization term. We used half-quadratic splitting technique to decouple the original problem into two small sub-problems: one is convex and can be solved by standard optimization algorithm, the other is non-convex but can be solved by a complete search. We showed that our method is able to remove efficiently the rolling effect due to jitter, even in the case of huge images and large non-integer displacements.

Second, to improve the resolution of acquired fluorescence images, we introduced a method of image deconvolution by considering a family of convex regularizers. The considered regularizers are generalized from the concept of Sparse Variation which combines the L1 norm and Total Variation (TV) to favors the colocalization of high-intensity pixels and high-magnitude gradient. The experiments showed that the proposed regularization approach produces competitive deconvolution results on fluorescence images, compared to those obtained with other approaches such as TV or the Schatten norm of Hessian matrix. The final deconvolution algorithm has been dedicated to large 2D 20000×60000 images acquired with ISO scan imager (see Fig.4). The method is able to process a 512×512 image in 250 ms (Matlab) with a non optimized implementation.

References: [32], [34]

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys company),
Hoai-Nam Nguyen.

7.3. Correlation-based method for membrane diffusion estimation during exocytosis in TIRFM

Participants: Ancageorgiana Caranfil, Charles Kervrann.

The dynamics of the plasma membrane of the cell is not fully understood yet; one of the crucial aspects to clarify is the diffusion process during exocytosis. Several image acquisition modalities exist, including TIRFM (Total Internal Reflection Fluorescence Microscopy), that have successfully been used to determine the successive steps of exocytosis. However, computing characteristic values for plasma membrane dynamics is problematic, as the experimental conditions have a strong influence on the obtained data, and a general model of molecular interaction dynamics cannot be determined.

In the PhD thesis of A. Caranfil, we have developed a computational approach to adapt the popular temporal image correlation spectroscopy (TICS) method to the analysis of a single fusing vesicle. The biophysical diffusion model parameters (for TfR protein) are estimated by an Approximate Bayesian Computing procedure which supplies the conditional expectation and maximum a posteriori estimators from temporal correlation data. Unlike TICS, our approach is robust to noise, estimation window size, spot location and non-uniform background. It can serve in biological studies investigating diffusion processes involved in exocytosis mechanisms.

Collaborators: Francois Waharte (UMR 144 CNRS-Institut Curie, PICT-IBiSA).

7.4. Classification of diffusion dynamics from particle trajectories

Participants: Vincent Briane, Charles Kervrann.

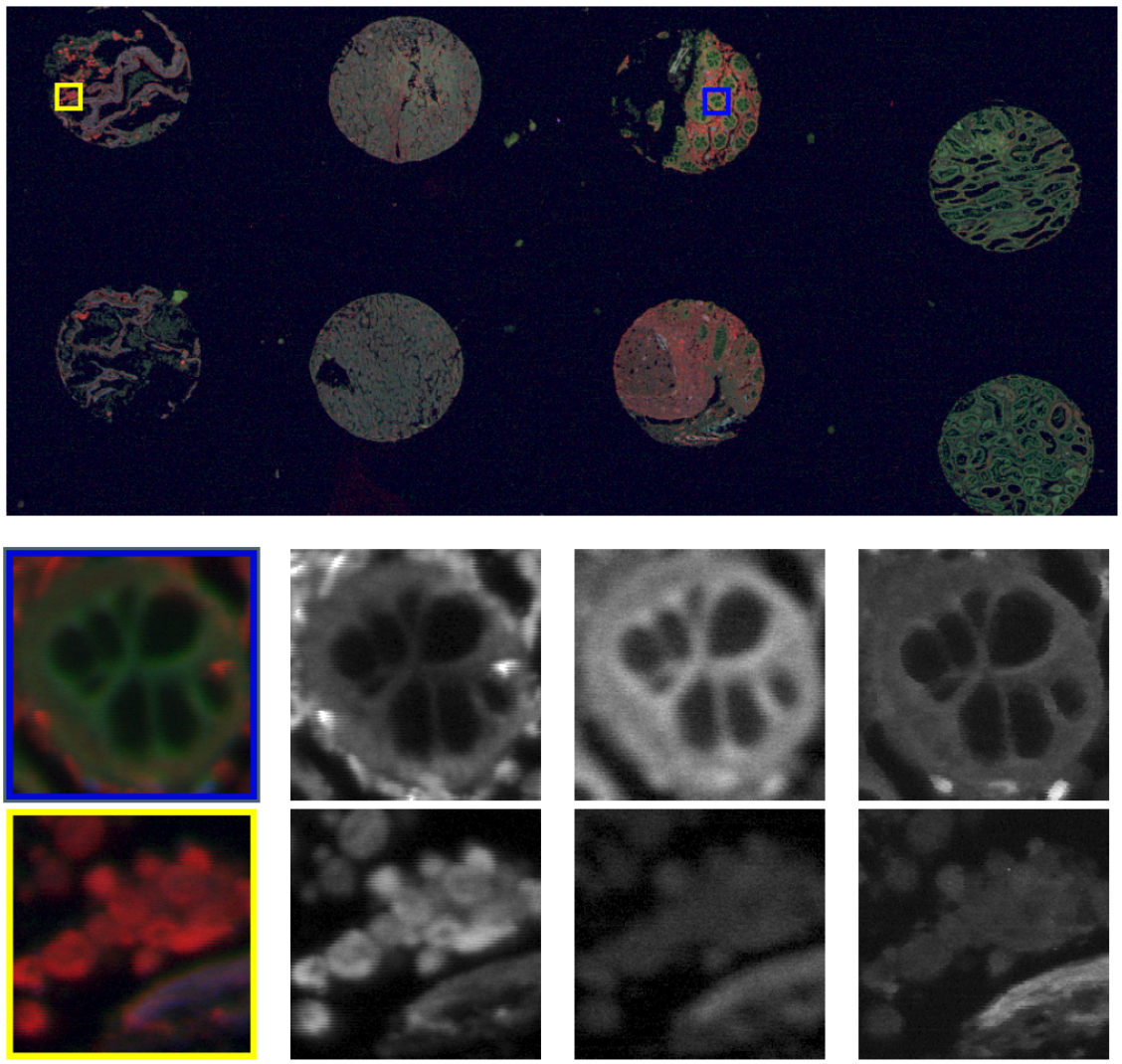


Figure 4. Three-color fluorescence image of eight tissue microarray cores. A region of interest of $4.7 \times 2.8mm^2$ *was scanned using the fluorescence scanner named InnoScan 1100AL equipped with three excitation wavelengths* ($488nm, 532nm$ *and* $635nm$) *at the spatial resolution* $0.5\mu m^2$*/ pixel, corresponding to an image of* 9544×4704 *pixels. Two areas which are bordered by two blue and yellow boxes are selected for visual comparison. First row: full size image. Second and third rows: zoom-in views of two selected areas; from left to right: 3 synchronized colors (red* ($488nm$)*, green* ($532nm$) *and blue* ($635nm$) *channels) displayed separately (courtesy of Innopsys).*

In this study, we are currently interested in describing the dynamics of particles inside live cell. Inference on the modes of mobility of molecules is central in cell biology since it reflects the interactions between the structures of the cell. In this work, we assume that the motions of particles follow a certain class of random process: the diffusion processes. Diffusions are stochastic processes with continuous paths and can model a large range of intracellular movements. Biophysicists distinguish three main types of diffusions, namely Brownian motion, superdiffusion and subdiffusion. These different diffusion processes correspond to distinct biological scenarios. A particle evolving freely inside the cytosol or along the plasma membrane is modelled by Brownian motion; the particle does not travel along any particular direction and can take a very long time to go to a precise area in the cell. Active intracellular transport can overcome this difficulty so that motion is faster in a given direction. In this case, particles are carried by molecular motors along microtubular filament networks and their motion is modelled with superdiffusion. Subdiffusion can be observed in two cases i/ when the particle is confined in a microdomain, ii/ when the particle is hindered by molecular crowding and encounters dynamic or fixed obstacles.

To address several issues in dynamics classification, we have developed a statistical test for classifying the observed trajectories into the three groups of diffusion of interest, namely Brownian motion, super-diffusion and subdiffusion. We have also designed an algorithm to detect the changes of dynamics along a single trajectory (see Fig. 5). We define the change points as the instants at which the particle switches from one diffusion type (Brownian motion, superdiffusion or subdiffusion) to another one. Finally, we have combined a clustering algorithm with our test procedure to identify micro domains, that is, zones where the particles are confined. Molecular interactions of great importance for the functioning of the cell take place in such areas.

Collaborators: Myriam Vimond (ENSAI Rennes),
Jean Salamero (UMR 144 CNRS-Institut Curie).

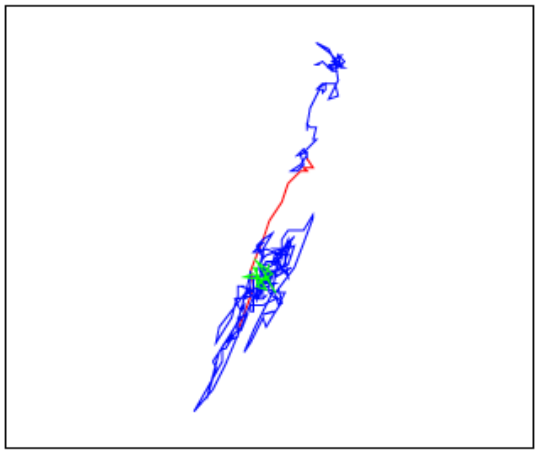


Figure 5. Change point detection on trajectories depicting neuronal mRNPs. The blue parts correspond to Brownian portions of the trajectory, red part to superdiffusive portions, green part to the subdiffusive portion.

7.5. Spatial statistics, point patterns, and colocalization in fluorescence imaging

Participants: Frédéric Lavancier, Charles Kervrann.

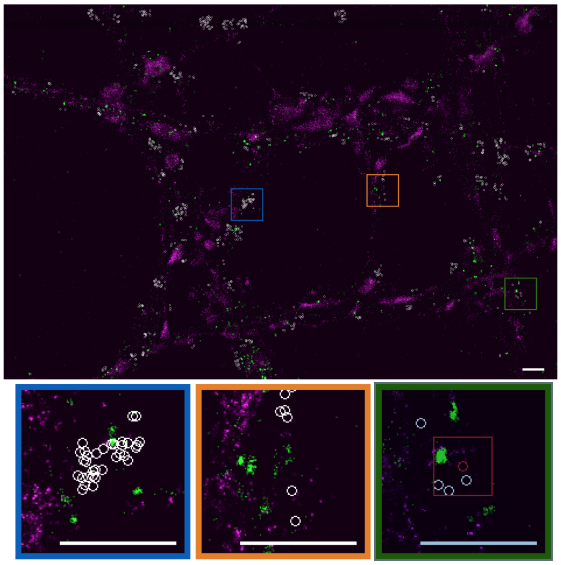


Figure 6. DSTORM acquisition of cells from hippocampi of mice expressing BDNF proteins (green channel) and vGlut (purple channel), with three zoomed-in regions (bottom). The colocalization regions identified by GcoPS are represented as white circles. The red rectangle represents the window used to find the colocalization hit shown as a red circle. The scale bars correspond to 1μm

In the context of bioimaging, colocalization refers to the detection of emissions from two or more fluorescent molecules within the same pixel of the image. This approach enables to quantify the protein-protein interactions inside the cell, just at the resolution limit of the microscope. It refers to the detection of emissions from two or more fluorescent molecules within the same pixel of the image. Colocalization is an open problem for which no satisfying solution has been found up to now. Accordingly, we proposed an objective, robust-to-noise colocalization method (GcoPS – Geo-coPositioning System)) which only requires the adjustment of a p-value that guarantees more reproducibility and more objective interpretation. It is based on the statistical analysis of the intersection (area/volume) between the two 2D or 3D binary segmented images. GcoPS handles 2D and 3D images, variable signal-to-noise ratios and any fluorescence image pair acquired with conventional or super-resolution microscopy (see Fig. 6). To our knowledge, no existing method offers the same robustness and precision level with such an easy control of the algorithm. In a recent study (internships 2017), we started to adapt this framework to analyze the spatiotemporal molecular interactions from set of 3D computed trajectories or motion vector fields (e.g., co-alignment), and then to fully quantify specific molecular machineries.

More generally, analysis of molecule and protein localization, of interactions and spatial distributions in living cells is helpful to understand functions in the cell and to compare spatialized phenotypes. This is also true with the emergence of single-molecule localization microscopy techniques (e.g., PALM), relying on the cumulative spatial localization of fluorescently tagged markers, and whose outputs are sets of spatial coordinates of single molecules. Accordingly, we were interested in the spatial distribution of single molecules that exhibit some randomness, regularity and spatial clustering (or aggregation) at large scales, while having a minimal distance between them. In that context, we theoretically studied several point processes able to represent the spatial organization of points. We focused on determinantal point processes (DDP), since they are able to describe spatial point patterns where nearby points repel or repulse each other. We also partly solved a 30 years old conjecture by proving the consistency of the likelihood procedure for a large class of Gibbs models (e.g., Strauss model, area-interaction model) which are commonly used models in practice. We extended the pseudo-

likelihood procedure to infinite range Gibbs interactions, and we proved its consistency and its asymptotic normality. All these models are now well understood and will be used in future works to analyse point patterns in cell imaging, generally described by Poisson point processes.

References: [30], [31], [35]

Collaborators: Jean Salamero and Liu Zengzhen (UMR 144 CNRS-Institut Curie),
David Dereudre (Laboratoire Paul Painlevé (UMR 8524), University of Lille 1),
Jean-François Coeurjolly (Laboratoire Jean Kutzmann, University of Grenoble).

7.6. Data assimilation and modeling of cell division mechanism

Participants: Ancageorgiana Caranfil, Charles Kervrann.

Nowadays, medical challenges demand a profound understanding of cellular mechanisms. Research in biology, biophysics and medical domain unravelled a significant part of the general processes occurring at the cellular level. It has enabled the understanding of much smaller scale processes, but our knowledge on these mechanisms is still limited as new, more complex issues need to be solved. In this context, we aim at understanding the role and interaction of the molecular key players at different scales, and their individual and collective impact on the global mechanism at the cell level. To this purpose, we have focused on the dynamics of the spindle during cell division mechanism. Our approach consists in creating a biophysical model for this mechanism, and uses data assimilation to adjust the model and optimally integrate the information from the observations. The overall spindle behaviour is led by the spindle poles behaviour. This year, we have proposed a new biophysical model for the posterior spindle pole functioning during metaphase and anaphase, that explains the oscillatory behaviour with a minimum number of parameters. Estimating the model parameters is ongoing, and will provide insights on molecular players role as well as guidance for future experiments to further investigate the dynamics of the spindle during cell division. First, we have focused on the temporal aspect. Spatial information on microtubules and molecular motors will be included in the model in the next part of this work.

Collaborators: Yann Le Cunff and Jacques Pécréaux (IGDR Institute of Genetics & Development of Rennes).

7.7. Quantifying the spatial distribution of intracellular events

Participant: Charles Kervrann.

Automated processing of fluorescence microscopy data allows to quantify cell phenotypes in an objective and reproducible way. However, most computational methods are based on the complex combination of heterogeneous features expressing geometrical, morphological and frequency properties, which makes difficult to draw definitive biological conclusions. Additionally, most experimental designs pool together data coming from replicated experiments of a given condition, neglecting the biological variability between individual cells. Hence, we developed a generic and nonparametric density framework (QuantEv) to discriminate spatiotemporal distributions (using circular Earth mover's distance) of moving proteins detected by any appropriate algorithm. The main advantage of QuantEv is to robustly process 2D and 3D data, and accurately analyse homogeneous and heterogeneous populations. As proof-of-principle, we first quantitatively characterized protein trafficking of Rab6 positive membranes between the Golgi apparatus and the plasma membrane. Next, we demonstrated that Rab11 positive membranes uniformly distribute around the Endosomal Recycling Compartment (ERC), regardless of the cell shape. Finally, we showed that actin organization is cell shape dependent, and evaluated its influence on the distribution of exocytosis/recycling vesicles. QuantEv is a flexible method which enables to quantify any intracellular trafficking in 3D flat or rounded, constrained or non-constrained, adherent or non-adherent cells.

References: [36]

Collaborators: Thierry Pécot (Hollings Cancer Center, Medical Univ. South Carolina, Charleston, USA),
Jean Salamero, Jérôme Boulanger and Liu Zengzhen (UMR 144 CNRS-Institut Curie).

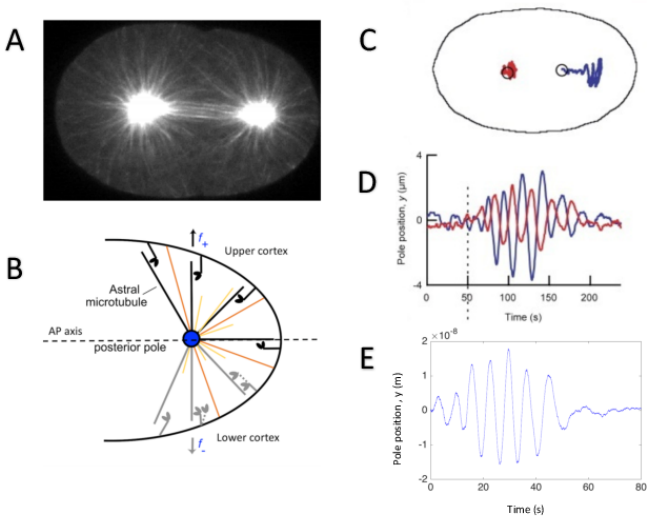


Figure 7. Illustration of the cell division mechanism observed in fluorescence microscopy (A). Sketch of one centrosome and connected to microtubules in the cell (B), experiments and tracking of the two centrosomes (C and D), and simulation of centrosome oscillations (E).

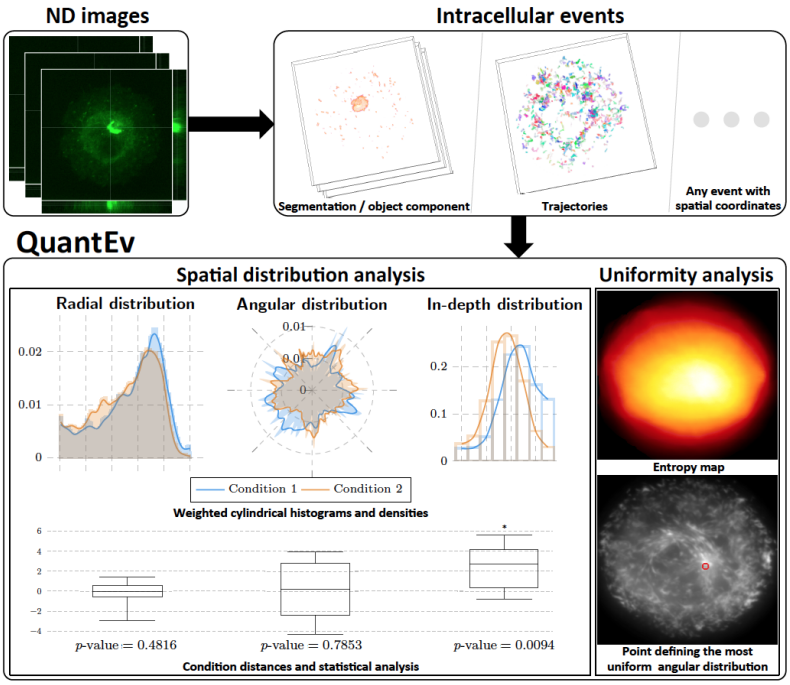


Figure 8. Overview of QuantEv approach.

7.8. 3D registration for correlative light-electron microscopy

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy.

In recent years, correlative light and electron microscopy (CLEM) has become an attractive tool in the bio-imaging field. Biologists can collect complementary information from light microscopy (LM) and electron microscopy (EM), respectively on the dynamics and on the structure of the cell. An overlay of the LM and EM images is needed to combine information from the LM and EM sources. We are developing a 3D automated CLEM method to register EM and LM image stacks. Given the significant gap between the field of view, position and orientation of the EM and LM stacks, it is not possible to estimate directly the 3D registration. We first compute a 2D maximum intensity projection (MPI) of the LM stack along the Z-axis, and we match 2D EM regions of interest (ROI), extracted from different EM slices, into the 2D LM-MPI image. From the resulting location candidates, we estimate with a robust criterion the 2D XY-shift to pre-align the LM and EM stacks. Afterwards, a 3D affine transformation between 3D-LM-ROI and 3D-EM–ROI can be estimated using mutual information. We successfully tested this framework on two first 3D correlative microscopy datasets.

Collaborators: Xavier Heiligenstein (UMR 144 CNRS-Institut Curie),
Grégoire Malandain (Inria, Morpheme EPC, Sophia-Antipolis).

7.9. Fast optical flow methods for 3D fluorescence microscopy

Participants: Sandeep Manandhar, Patrick Bouthemy, Charles Kervrann.

Estimating motion of cells and of subcellular particles is crucial for deciphering cell mechanisms and understanding cell behaviors. Modern 3D light microscopy (LM) for cell biology enables to observe cell dynamics at a good resolution, in both space and time, motivating the development of 3D optical flow methods. However, the acquired 3D LM image sequences exhibit several specificities making 3D motion computation a difficult problem. We have defined an original and efficient two-stage estimation method for light microscopy image volumes. The method, developed in the frame of S. Manandhar PhD thesis, takes a pair of LM image volumes as input, segments the 2D slices of the source volume in super-pixels, and first estimates the 3D displacement vectors of the super-pixel centers. To this end, we have extended the well-known PatchMatch method to 3D volumes, where correspondences act between voxels. Both the propagation and the random search steps were adapted to 3D volumes. Then, a weighted interpolation has been designed to recover the dense 3D flow field for all the voxels, from the sparse 3D displacement field. The super-pixel segmentation is exploited to define the neighborhood for interpolation, and the interpolation weights take into account intensity edges and local motion differences to preserve flow discontinuities. The experimental results show good gain in execution speed, and accuracy evaluated in computer-generated 3D data with ground-truth. The results were promising on two real 3D LM image sequences supplied by USTW. The sequences depict blebbing in a MV3 cell (see Fig. 9). The cell membrane protrudes increasing the surface area of the cell. These protrusions, referred to as blebs, appear and disappear in interval of minutes, the bleb appearance corresponding to the stretching of a local region of the cell membrane. The total computation time was for the first sequence 163 seconds (resp. 101s for the second sequence), with 19 (resp. 49), 120 (resp. 44) and 24 (resp. 8) seconds for super-pixel generation, 3D patch matching, and interpolation respectively, on a computer with 2.8 GHz Intel i7 processor and 16 GB of RAM.

Collaborators: Philippe Roudot and Gaudenz Danuser (UTSW, Dallas, USA).

7.10. 3D Convolutional Neural Networks for macromolecule localization in cryo-electron tomograms of intact cells

Participants: Emmanuel Moebel, Charles Kervrann.

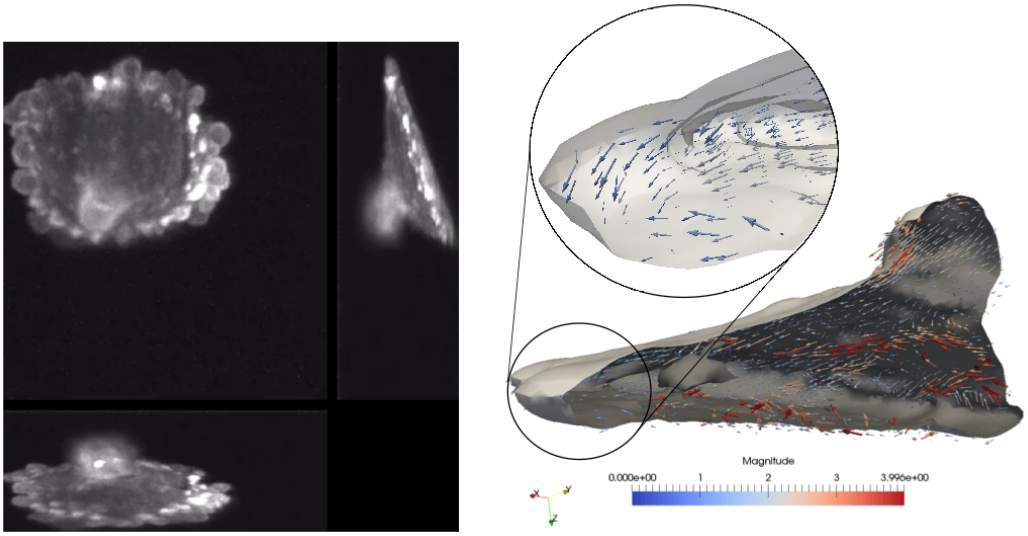


Figure 9. Illustration of 3D optical flow computation to analyze bleb deformation during cell migration in Bessel beam light sheet microscopy (input images by courtesy of Danuser lab, UTSW Dallas, USA).

In this study, we focus on macromolecule localization and classification in cryo-electron tomography (CET) images. Biologists are in need for efficient methods to localize macro-molecules (e.g. ribosomes) in frozen cell samples. The high amount of noise and imaging artifacts are the reasons why very few computational methods exist for this task. In fact, the most used method today is template matching (TM) whose resulting score map comprises a high amount of false positives. Therefore, it is necessary to apply post-processing techniques (ROI selection, classification) in order to refine the localization results. We propose an alternative localization method to TM, based on a convolutional neural network (CNN). The idea is to propose a robust and more straight-forward approach, allowing to bypass the conventional processing chain. By using python toolboxes optimized for GPU computing (elektronn, keras), we are able to reach computation time much lower than the current approach. Results on synthetic data demonstrate the superiority of our approach compared to TM. In addition, we applied our method on experimental data in order to localize sub-classes of ribosomes (membrane-bound and cytoplasmic ribosomes), a task difficult to achieve with TM alone. We are currently in the process of publishing these results. Future perspectives include localizing smaller macro-molecules, like proteasomes.

Collaborators: Damien Larivière (Fondation Fourmentin-Guilbert),
Julio Ortiz, Antonio Martinez (Max-Planck Institute, Martinsried, Germany).

7.11. Estimation of parametric motion models with deep neural networks

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

We have proposed an end-to-end learning architecture for estimating a parametric motion model for a moving scene. We handle motion outliers by using the supervised training trick that is used by stacked denoising auto-encoders. Here, we define motion outliers as regions of the image whose motion does not correspond with the estimated parametric motion model. In other words, we seek to find a parametrized dominant motion of the dynamic scene. We leverage stacked hourglass networks with a final hard-coded block corresponding to

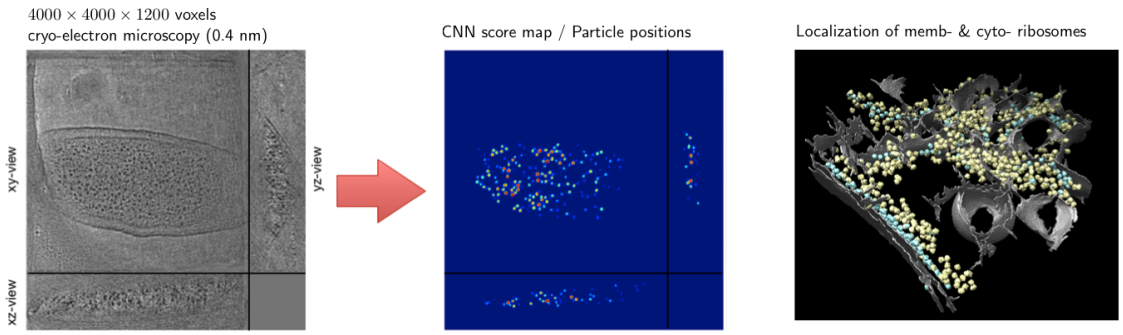


Figure 10. Illustration of 3D CNN to localize ribosomes isolated in the cytoplasm and close to the cell membrane in cryo-electron tomography (courtesy of Max-Planck Institute, Martinsried, Germany).

the global parametric motion model estimator. This block replaces the decoder part of a convolutional auto-encoder network, and it is end-to-end trainable since it involves linear operations only. Moreover, the hard-wired decoder allows the network to output the values of the parametric motion model given an input moving scene, even when the supervision acts on optical flow maps and not the motion model values. This means that our network is able to provide, as a by-product, a concise code that can be used as motion descriptor.

Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

7.12. Motion saliency in video sequences

Participants: Léo Maczyta, Patrick Bouthemy.

Dynamic (or motion) saliency is a means to detect unexpected or rare dynamic behaviors in video sequences acquired by a stationary or a mobile imaging device. Finding salient dynamic information in each image of a sequence is indeed crucial in many situations. We aim to extract saliency only from motion information, and to exhibit salient motion in contrast to its space-time context with no prior on the nature of both. So far, we have investigated a simpler problem than saliency map estimation. We deal with the classification of each image of a sequence as dynamically salient or not, that is, containing salient motion or not. We have explored convolutional neural network (CNN). We have designed two different networks. The first one relies on two intensity images, the first input image and the second image warped with the parametric dominant motion estimated between the two input images. The second one takes as input the difference between the computed optical flow and parametric dominant flow.

Collaborators: Olivier Lemeur (EPC Sirocco, Inria Rennes - Bretagne Atlantique).

TAPDANCE Team (section vide)

VIRTUAL PLANTS Project-Team

6. New Results

6.1. Analysis of structures resulting from meristem activity

6.1.1. Acquisition and design of plant geometry

Participants: Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier.

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 (light, 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 [67], [58]. 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 laser scanner and pictures.

- *Reconstruction of tree structures from 3D laser scanner data.* (Bilan Yonis-Omar, Emma Carrié, Frédéric Boudon, Christophe Godin, Benoit Pallas [AFEF, AGAP], Evelyne Coste [AFEF, AGAP])

  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 the branching systems, we previously proposed a reconstruction method to reconstruct plausible branching structures from laser scanner data based on the concept of space colonization [66]. Additionally, a number of automatic methods were proposed in the literature. A graphical editor has been developed and makes it possible to test these different methods and correct manually the reconstruction on laser scans. An additional validation pipeline makes it possible to compares automatic reconstruction with ground truth data using two indices of geometrical and structural similarities [53].

  An application for the reconstruction of an apple tree core collection (1000 trees) has been conducted during the internships of B. Yonis-Omar and E. Carrié in a collaboration with the AFEF Team of UMR AGAP. A protocol that minimize the number of movement of the scanner has been setup. Some first method to characterize and reconstruct architectural traits from the scan has been defined.

- *Characterizing wheat canopy characteristics from LiDAR measurements.* (Shouyang Liu [Emmah,Inra], Fred Baret [Emmah,Inra], Frédéric Boudon, Christian Fournier)

  Green area index (GAI) has been difficult to estimate accurately at large scales due to the cost prohibitive nature of classical in-situ methods. We propose to use LiDAR to overcome this problem. Through this work, we proposed a self-learning method to estimate GAI using LiDAR-derived metrics over a wheat field.

  Specifically, we developed a LiDAR simulator to carry out scanning on digital 3D objects, mimicking the measuring principle and setups of actual LiDAR sensors. The footprint and the geometrical configuration of the LiDAR are explicitly accounted for. Comparison with measurements of actual LiDAR demonstrates that the simulator generates a 3D point cloud having the same statistical properties as those derived from the actual LiDAR measurements.

We then used a machine learning algorithm to correlate LiDAR-derived metrics and GAI over synthetic datasets. 3D wheat canopy scenes were generated with AdelWheat model for two contrasting development stages across a wide range of combination of the model parameters. The scenes were transformed into 3D point clouds using the LiDAR simulator. Results demonstrate that emerging properties, such as leaf area index (GLAI), could be retrieved with a good accuracy.

- *Reconstruction of annual plants from multi-view images.* (Simon Artzet, Jerome Chopard, Christian Fournier, Christophe Pradal, Christophe Godin, Xavier Sirault [CSIRO-HRPPC, Canberra], Tsu-Wei Chen[Inra, LEPSE])

  Image-based phenotyping platforms in semi-controlled conditions offer large possibilities to perform genetic analyses of plant growth, architecture, light interception, and biomass accumulation over large time series for thousands of plants. However, methods for image analysis currently available are still very crude and need improvement and robustness to process huge amount of data. We are developing a python software framework dedicated to the analysis of high throughput phenotyping data and models named Phenomenal. This software framework currently consists of 2D and 3D image analysis workflow which ranges from 2D organs segmentation, 3D multi-view reconstruction, image-base meshing transformation, 2D/3D morphological thinning/skeletonization, 3D segmentation and tracking of plant organs maize (under development). We have processed images from phenoarch platform of the last four years and have built for each plant (maize, cotton, etc.) a voxel point cloud and image-base meshing representation and also for 725 maize plants a voxel point cloud automatically segmented (currently stem and mature leaf). Each process is run on distant server (private or virtual machines on FranceGrille cloud) and results can be viewed via a jupyter notebook server. Furthermore, 3D FSPM model for maize architectural development (named Cereals), is used to help segmenting plant images and to automate the mapping between segmented 3D objects and plant organs defined in the model. The 3D reconstructed model is combined with meteorological data to feed a light distribution model and estimate light use efficiency or establish response curve of morphogenetic processes to light environment [37]. This software framework was presented to “BMVA technical meeting: Plants in Computer Vision”.

  This research theme is supported by the PIA Phenome.

- *Tracking the growth of maize ear and silks in a high-throughput phenotyping platform using a robot-assisted imaging pipeline* (Simon Artzet, Jerome Chopard, Christian Fournier, Christophe Pradal, Nicolas Brichet [LEPSE, INRA], Llorenç Cabrera-Bosquet [LEPSE, INRA])

  In maize, silks are hundreds of filaments that simultaneously emerge from the ear for collecting pollen over a period of 1–7 days, which largely determines grain number especially under water deficit. Silk growth is a major trait for drought tolerance in maize, but its phenotyping is difficult at throughputs needed for genetic analyses.

  We have developed a reproducible pipeline [31] that follows ear and silk growths every day for hundreds of plants, based on an ear detection algorithm that drives a robotized camera for obtaining detailed images of ears and silks. We first select, among 12 whole plant side views, those best suited for detecting ear position. Images are seg mented, the stem pixels are labelled and the ear position is identified based on changes in width along the stem. A mobile camera is then automatically positioned in real time at 30 cm from the ear, for a detailed picture in which silks are identified based on texture and colour. This allows analysis of the time course of ear and silk growths of thousands of plants. The pipeline was tested on a panel of 60 maize hybrids in the PHENOARCH phenotyping platform. Over 360 plants, ear position was correctly estimated in 86% of cases, before it could be visually assessed. Silk growth rate, estimated on all plants, decreased with time consistent with literature. The pipeline allowed clear identification of the effects of genotypes and water deficit on the rate and duration of silk growth.

  The pipeline presented here, which combines computer vision, machine learning and robotics, provides a powerful tool for large scale genetic analyses of the control of reproductive growth to

changes in environ mental conditions in a non invasive and automatized way. It is available as Open Source software in the OpenAlea platform.

- *Review on morphological plant modelling.* (Christophe Pradal, Mathilde Balduzzi, Alexander Bucksch [Georgia Univ., USA], Daniel H. Chitwood [Donald Danforth Plant Science Center, USA], Erin E Sparks [Univ. of Delaware, USA])

  Plant morphology is inherently mathematical. The geometries of leaves and flowers and intricate topologies of the root have fascinated plant biologists and mathematicians alike. Beyond providing aesthetic inspiration, understanding plant morphology has become pressing in an era of climate change and a growing population. Gaining an understanding of how to modify plant architecture through molecular biology and breeding is critical to improving agriculture, and the monitoring of ecosystems and global vegetation is vital to modeling a future with fewer natural resources. In this review [8], we begin by summarizing the rich history and state of the art in quantifying the form of plants, mathematical models of patterning in plants, and how plant morphology manifests dynamically across disparate scales of biological organization. We then explore the fundamental challenges that remain unanswered concerning plant morphology, from the barriers preventing the prediction of phenotype from genotype to modeling the fluttering of leaves in a light breeze. We end with a discussion concerning the education of plant morphology synthesizing biological and mathematical approaches and ways to facilitate research advances through outreach, cross-disciplinary training, and open science. Never has the need to model plant morphology been more imperative. Unleashing the potential of geometric and topological approaches in the plant sciences promises to transform our understanding of both plants and mathematics.

- *Qualitative and Quantitative Descriptors for Plant Morphology.* (Christophe Pradal, Mathilde Balduzzi, Alexander Bucksch [Georgia Univ., USA], Erin E Sparks [Univ. of Delaware, USA])

  An emerging challenge in plant biology is to develop qualitative and quantitative measures to describe the appearance of plants through the integration of mathematics and biology. A major hurdle in developing these metrics is finding common terminology across fields. In this review, we define approaches for analyzing plant geometry, topology, and shape, and provide examples for how these terms have been and can be applied to plants. In leaf morphological quantifications both geometry and shape have been used to gain insight into leaf function and evolution. For the analysis of cell growth and expansion, we highlight the utility of geometric descriptors for understanding sepal and hypocotyl development. For branched structures, we describe how topology has been applied to quantify root system architecture to lend insight into root function. Lastly, we discuss the importance of using morphological descriptors in ecology to assess how communities interact, function, and respond within different environments. This review [30] aims to provide a basic description of the mathematical principles underlying morphological quantifications.

6.1.2. Modeling the plant ontogenic program

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

This research theme is supported by one PhD program.

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. [47], [59], [60], [57], 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 program" [65], "age state" [56] or "physiological age" [50]. 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 [56], [50]. 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 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. leaf growth rate, leaf surface, internode length, 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 [21] that enable to tackle hierarchically-structured categorical response variables. Typically, we need to distinguish different timing of branching events (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 [20].

- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Alix Allard [AGAP, AFEF team], Evelyne Costes [AGAP, AFEF team])

  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 shoots, which correspond to portions of stems that were grown during the same year). One or several replications of each genotype were available.

  Three families of indices were proposed for early detection of alternation during the juvenile phase. The first family was based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. The second family was based on a 2nd-order Markov chain with fixed and random effect in transition probabilities. The third family was based on entropy indices, in which flowering probabilities were corrected from fixed effects using Generalized Linear Models.

  This allowed early quantification of alternation from the yearly numbers of inflorescences at tree scale. Some quantitative trait loci (QTL) were found in relation with these indices [43], [35].

  New data sets where collected in other F1 progenies. Ancestral relationships between parents of different progenies were taken into account to enhance the power of QTL detection using Bayesian methods. Other QTLs are expected to be found using these new indices and genetic material. However, the amount of replicate per genotype and of data per replicate is quite reduced compared to those of our previous work. This is why we will investigate the loss of power in QTL detection due to a degraded amount of data, by simulating data deletion in our reference results.

- *Characterizing tree patchiness using a tree segmentation/clustering approach.* (Pierre Fernique, Anaëlle Dambreville, Jean-Baptiste Durand, Christophe Pradal, Yann Guédon, Frédéric Normand [CIRAD, HortSys, Réunion Island], Pierre-Eric Lauri [INRA, System]).

  Many tropical trees are affected by strong phenological asynchronisms entailing patchiness. Patchiness is characterized by clumps of homogeneous botanical entities (e.g. a clump of flowering growth units) within tree canopy. It is therefore assumed that there are subtrees within which the characteristics of the botanical entities follow the same or nearly the same distribution, and between which these

characteristics have different distributions. The detection of such subtrees can thus be stated as tree-indexed data segmentation. We therefore transposed multiple change-point models to tree-indexed data. The output of the segmentation procedure is a partition of trees such that two non-adjacent subtrees can be very similar in terms of botanical entity characteristics. We thus incorporated a second stage of clustering of subtrees based on a mixture model in order to group non-adjacent similar subtrees. Finally, directed acyclic graphs we built for summarizing the succession of patches over time within the canopy. This statistical modeling framework was applied to young mango trees [11].

- *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 [INRA, System])

  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. With such model, we showed the importance of architectural and temporal factors in the development of the units of the trees, see 1 . The model also simulates the phenology of shoots and inflorescences. For this, the sizes of the different organs is modelled by statistical laws estimated from measurements that depends on their locations in the architecture. The growth speed of organs is modulated by the temperature. The model has been then coupled with an ecophysiological model of fruit growth [62], [63]. The global aim is to have a crop simulation model to predict fruit yield and quality on mango tree.

Figure 1. Simulation of the development of a mango tree over two cycles [52]. The first and last image corresponds to the end of the vegetative period of the 3rd and 5th growing cycle (June), respectively while the second and third images correspond to the flowering phase (August) of the 3rd and 4th cycles, respectively. The different colours of the inflorescences of the 3rd image show different developmental stages and the flowering asynchronism over the tree.

In the context of the PhD of S. Persello, we aim at extending this model with the effect of agricultural practices. For this, a number of experiment has been conducted this year with some mango trees being pruned with different intensity (global mass removed) and severity (depth of the removed elements). Analysis and characterization of the effect of pruning on the subsequent vegetative development of the tree in currently under investigation.

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

  Our aim was to characterize the phenology of various 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 [19]. A multivariate generalization of the synchronous segmentation approach was developed in the context of Marc Labadie's PhD [14], the idea being to characterize not only the flowering pattern as in our first study but more generally the developmental pattern combining vegetative development, branching and flowering.

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

  In a previous work [13], 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 class of self-nested trees presents remarkable compression properties because of the systematic repetition of subtrees in their structure. In a collaboration with two other Inria project-teams (MISTIS and BIGS), studied methods to approximate a tree with a tree in the class of self-nested trees. We first provided a better combinatorial characterization of this specific family of trees. We then showed that self-nested trees may be considered as an approximation class of unordered trees. We finally compared our approximation algorithms with a competitive approach of the literature on a simulated dataset [4].

6.1.3. Analyzing the influence of the environment on the plant ontogenic program

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 6.1.2), 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.

- *Investigating how architectural development interfere with epidemics and epidemic control.* (Christian Fournier, Corinne Robert [Ecosys, INRA], Guillaume Garin [ITK, Montpellier], David Claessens [ENS ULM, Paris], 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 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 work is to provide a framework to both extend and simplify the modelling of pathosystems using FSPMs. Complex models are disassembled into separate *knowledge sources* originating from different specialist areas of expertise and these can be shared and reassembled into multidisciplinary models. Following the PhD of Guillaume Garin, we finalised a sensitivity analysis of the response of the severity of septoria to architectural traits, and an analysis of the influence of the wheat architecture on the competition between septoria and brown rust. These studies allowed to populate the framework with consistent example of application, and lead to the

development of operational modules that allows the fitting and validation of pathosystem models with experimental data.

- *Investigating how hydraulic structure interfere with gas-exchange dynamics of complex plants canopies under water deficit* (Christophe Pradal, Christian Fournier, Rami Albasha [LEPSE, Inra], THierry Simmoneau [LEPSE, Inra] and Eric Lebon [LEPSE, Inra])

  Individual leaves positioning within a plant canopy is a major determinant of the spatial distribution pattern of gas-exchange rates and energy budget within that canopy. Under water deficit, this distribution may be altered since soil drying affects stem hydraulic conductivity and, consequently, leaves stomatal conductance, suggesting that the hydraulic structure of the shoot may shape the intra-canopy variability of gas-exchange rates under water deficit. In this project, we design HydroShoot [1], a functional-structural plant model which allows simulating the hydraulic structure, energy budget and gas-exchange fluxes of complex plant canopies under water deficit. Model parameters are calibrated and validated using sapflow and entire plant gas exchange data collected in 2009 and 2012 from grapevine (Vitis vinifera L. cv. Syrah) experiments under three training systems (Lyre, GDC and VSP) having contrasted canopy structures. The model is then used to evaluate the role of the hydraulic structure in predicting the intra-canopy variability of temperature and intrinsic water use efficiency of trained grapevines. The resulting HydroShoot model allows to capture the effect of the different training systems on the spatial distribution of temperature and foliar photosynthesis within the canopy. We show that the intra-canopy variability of gas-exchange dynamics were mainly explained by the variability of local climate conditions, while the role of the hydraulic structure appeared only as secondary. Finally, the proposed HydroShoot model has been implemented for grapevine in the OpenAlea platform and will be extended to other plant architectural systems.

- *Eucalyptus development in response to different water stress and fertilization levels* (Yann Guédon, Charlène Arnaud (CIRAD AMAP and BioWooEB), Yves Caraglio, Sylvie Sabatier (CIRAD AMAP)

  Eucalyptus grandis has been grown successfully in plantations in many tropical regions including southern Brazil. The objective of the PhD of Charlène Arnaud (CIRAD AMAP and BioWooEB) is to study the modulation of the development of Eucalyptus main stems in response to water stresses and different levels of potassium or sodium fertilization. Eucalyptus main stem is characterized by a three-scale growth pattern with (i) at coarse scale, roughly stationary growth phases with phase changes often corresponding to cold seasons (ii) at intermediate scale, some growth fluctuations corresponding to the influence of the climatic factors (mainly temperature and cumulated rainfall) and (ii) at fine scale, more ore less systematic alternation of short and long internodes as a consequence of the phylotactic pattern. We thus developed a pipeline of statistical models that incorporates specific multiple change-point models (piecewise 1st-order stationary autoregressive models) for characterizing this three-scale growth pattern.

6.2. 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.

6.2.1. 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, UM], Jan Traas [RDP, ENS Lyon], Patrick Lemaire [CRBM, CNRS], Yassin Refahi [RDP, ENS-Lyon / Sainsbury Lab, Cambridge, UK].

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 [10] 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.

We finalized the development of a a new segmentation and tracking pipeline, ASTEC (Adaptive Segmentation and Tracking of Embryonic Cells). ASTEC is a one-pass algorithm (in contrast to MARS-ALT, that perform first the segmentation and then the tracking in two-passes) that is best suited for movies with numerous close time-points acquired at high spatio-temporal resolution. This pipeline takes advantage of information redundancy across the movies and biological knowledge on the segmented organism to constrain and improve the segmentation and the tracking. We used this one-pass algorithm to segment and track all cell shapes of a developing embryo of the marine invertebrate *Phallusia mammillata*. As a result we obtained the full track of the shapes of all the cells from the 64 cell stage up to the early tailbud stage (1030 cells undergoing 640 division events followed across 180 time-points through 6 hours of development imaged every 2 minutes, Figure 2). To our knowledge, it is the first time that such high-resolution 4D digital tissues have been generated taking into account the cell shapes.

Based on this quantitative digital representation, we systematically identified cell fate specification events up to the late gastrula stage. Computational simulations revealed that remarkably simple rules integrating measured cell-cell contact areas with spatio-temporal expression data for extracellular signalling molecules are sufficient to explain most early cell inductions. This work suggests that in embryos developing with stereotyped cell shapes and positions (like *Phallusia mammillata* embryos), the genomic constraints for precise gene expression levels are relaxed, thereby allowing rapid genome evolution. A paper describing the whole approach has been submitted in December 2017.

- *Creating mesh representation of cellular structures.*
  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 tools to reconstruct a 3D cell complex representing the tissue, based on the dual simplicial complex of cell adjacencies. This set of tetrahedra is optimized from a reasonable initial guess to match the adjacencies in the tissue, which proved to produce a very faithful reconstruction [55]. We also developed a set of methods to triangulate such reconstructions, and enhance the quality of triangular mesh representations of plant tissue, simultaneously along several criteria [54].

  These tools are integrated in the DRACO-STEM computational pipeline released as an independent package to enable biomechanical simulations on real-world data.

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

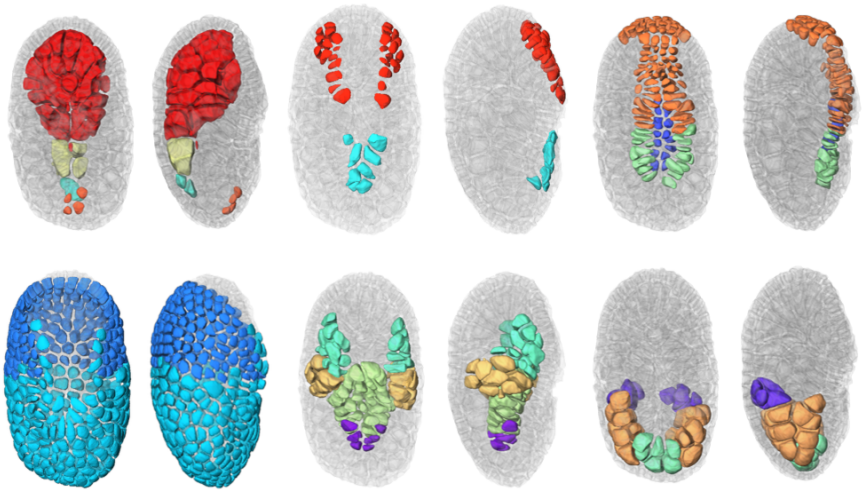


Figure 2. 3D projection of the segmented embryo at the early tailbud stage. The cells are colored by tissue type. The cells are slightly roded to allow their distinction. The other cells of the embryo are in transparent grey. The doral and lateral sides are shown.

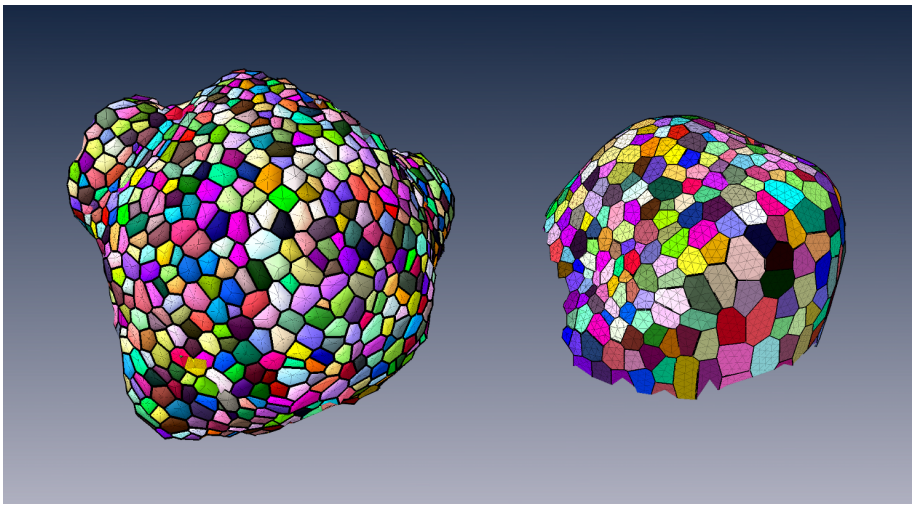


Figure 3. Triangular mesh representations of shoot apical meristem and flower meristem tissues obtained from MARS segmentations

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 [61]. 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 being investigated (a manuscript is in preparation).

6.2.2. Shape analysis of meristems

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

The MARS-ALT pipeline provides rich spatio-temporal data sets for analyzing the development of meristems, since it allows to performs 3D cell-segmentation and to compute cell-lineage. This enable the extraction and study of spatio-temporal properties of a tissue at cellular scale. To facilitate the analysis and to structure the obtained data have implemented a dedicated temporal graph structure. In this graph, vertex are cells and edges are spatial or temporal relationships, thus proposing a natural representation of the growing tissue. 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 space, cellular patterns can be identified, by clustering methods for instance.

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 may be 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.

6.2.3. Mechanical models of plant tissues

Participants: Olivier Ali, Hadrien Oliveri, Christophe Godin, Jan Traas [ENS-Lyon].

This research theme was supported, between 2012 and 2017, by the Inria Project Lab Morphogenetics and Jan Traas's ERC.

During the previous years, we set up a multi-scale mechanical model of a growing *shoot apical meristem* (the specific tissue at the very tip of plants where stem cells are active and produce new organs such as branches, leafs and flowers) with sub-cellular resolution, a detailed description of the core elements of this modelling approach has been developped in our previous reports. The aim of this project is to provide a computational framework for simulating growth of multicellular plant tissue. Several papers (and a reivew) have been published over the past few years on this work, in close collaboration with biologists: [51], [3], [26], [2].

Last year, our simulations pointed out that cell wall remodelling and growth initiation have to be co-regulated in order to initiate young organs formation. Biologists unraveled a biochemical signaling pathway that could explain this synergy. This joint work is currently under submission in a high impact factor journal.

Two years ago, we started to work on the integration of a feedback loop between mechanical stresses and growth (PhD work of Hadrien Oliveri started in Oct. 2015). A close study of this feedback mechanism made us refine several aspects of our modelling approach (tensor formalism to quantify cell polarity, ...). Ever since, Hadrien Oliveri has been studying the influence of this feedback loop on the morphogenesis of an epithelium. FEM-based simulations have been carried out on simple structures as proof of concept. This first step of the work is currently being submitted. This year, we also started to study the influence of such a mechanical-based feedback mechanism on the morphogenesis of real tissues. The specific question we want to investigate concerns flatness: How can plants produce flat organs such as leaves of sepals ? We investigate this question in the context of the sepal formation, always in close collaboration with biologists doing experiments on the very same topic.

This year, we also started in investigate to a quantitative manner the mechanical influence of inner tissues in the morphogenesis process. Indeed, up to now, our modelling approach was focused on the mechanics of the epidermis, known to be the main load-bearing layer. However, new experimental evidence suggest that inner tissues may influence and/or trigger morphogenesis processess. In order to investigate such mechanisms, our strategy relies on the use of high quality digitized tissues in 3D. Such structures, composed of triangular meshes, are produced through a workflow based on the updated version of the MARS pipeline and the DRACO-STEM module, developped within the team. Currently, numerical simulations are being carried out to analyse the mechanical equilibrium of the structures loaded with pressurized forces, see 4 . Preliminary results tend to confirm the leading mechanical role of the epidermis. Interestingly, sharp differences in the mechanical characteristics emerge between epidermal cells and inner ones.

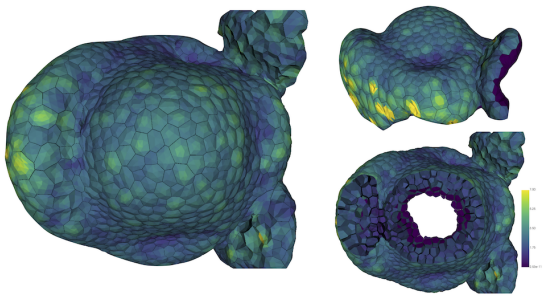


Figure 4. Heatmap of the stress tensor amplitude in a digitalized flowerbud.

6.2.4. Mechanical modelling of embryo morphogenesis.

Participants: Bruno Leggio, Emmanuel Faure, Patrick Lemaire [CRBM, CNRS], Christophe Godin.

A work on data analysis and modelling of morphogenesis and development in embryos of ascidians has been started this year. It comprises two main branches: starting from segmented data at cellular resolution, global and local symmetries of embryo development were analyzed. An analysis in terms of entropy of conserved embryonic properties was developed in order to characterise different stages of development as well as different tissues. In parallel, a mechanical and topological analysis of cell-cell interactions was carried out. This lead us to develop a new and original physical model of cleavage-plane determination in different tissues, with the goal of understanding the role of purely mechanical interactions in shaping ascidian embryos.

6.2.5. 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 meristem repeatedly fails in making a complete transition to the flower. The network was validated by multiple analyses, including sensibility analyses, stable state analysis, mutant analysis, among others. Once the network was validated, it was coupled with an architectural model of plant development using L-systems. The coupled model was used to study how does changes in gene dynamics and expression could change the architectural properties of plants and produce cauliflowers instead of flowers. Finally, the architectural model without the network was used to study how changes in certain parameters could generate different curd morphologies, including the normal cauliflower and the romanesco one.

We have three main types of predictions. (1) How does gene expression is modified from WT to cauliflower organisms. (2) How gene regulate plants shape in order to produce curds instead of flowers. (3) The main parameter regulating curds shapes. The predictions made using the model are currently being experimentally tested. All our results could provide a comprehensive understanding of how does genes and plant architecture are linked in a dynamical way.

6.2.6. Model integration

Participants: Frédéric Boudon, Christophe Godin, Guillaume Cerutti, Jean-Louis Dinh, Eugenio Azpeitia, Jan Traas.

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

One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development, TissueLab. This platform, based on *OpenAlea*, will be used to carry out integration of the different models developed in this research axis. In the past year, progress has been made in defining a generic tissue data structure that would be visualized, manipulated and updated through 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 being investigated. A 2D version is being developed in the context of Jean-Louis's Dinh PhD thesis, and will be described in a forthcoming book chapter.

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). An introduction to the modeling of some main components of such integrated system was published as a book chapter in the series of Ecole de Physique des Houches [12]. 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 sub-model 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 [64] and the CPIB group in Nottingham, UK [49].

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

6.3.1. Modeling water transport in roots

Participants: Mikaël Lucas [IRD], Christophe Pradal, Christophe Godin, Yann Boursiac [BPMP], 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 developed 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 UMR 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 challenge 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.

As the simulations may be time consuming and results sometimes difficult to interpret on complex branching systems, we started to investigate new methods to compute efficiently hydraulic conductivities and corresponding flows on complex root systems using architecture compression technics developed in the 1st axis of the project. First results show that very efficient computations of complex hydraulic architectures can be derived from the use of these compression techniques on idealized root architectures. These encouraging results provide a new abstraction that will be used in combination with the detailed modeling approach described above to break down the complexity of the analysis these huge branching systems.

6.3.2. Functional-Structural Root System Models in the Context of Drought Tolerance Breeding

Participants: Mikaël Lucas [IRD], Christophe Pradal.

This research theme is supported by a PhD program at IRD.

In this work, we study the impact of hydraulic architecture on water fluxes, and we review the conception and use of functional-structural root models in the broader context of research on root-driven drought tolerance, on the basis of root system architecture (RSA) phenotyping [40]. Such models result from the integration of architectural, physiological and environmental data. Here, we consider the different phenotyping techniques allowing for root architectural and physiological study and their limits. We discuss how QTL and breeding studies support the manipulation of RSA as a way to improve drought resistance. We then go over the integration of the generated data within architectural models, how those architectural models can be coupled with functional hydraulic models, and how functional parameters can be measured to feed those models. We then consider the assessment and validation of those hydraulic models through confrontation of simulations to experimentations. Finally, we discuss the up and coming challenges facing root systems functional-structural modeling approaches in the context of breeding.

6.3.3. 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.

Fruits and plants in general are large scale hydraulic systems in which growth is closely linked to water fluxes: thanks to osmotic pressure difference, the cells are able to absorb water from their environment and therefore increase their volume; as the cells are bounded by rigid walls, this results in both hydrostatic pressure (the so-called turgor pressure) in the cell and tension in the cell walls; above a threshold, synthesis of new cell wall material occurs and relaxes the tension. This process allows cells to grow, and along with cell division, is responsible for plant growth. In fruits, phloem and xylem vascular networks provide the water fluxes necessary for growth, while the osmotic pressure is mainly regulated by sugar intake from the phloem. The goal of this project is to combine a description of water and sugar fluxes at the fruit scale with a modelling of growth at cell level, as described above.

As a first step in this direction, we have developed a bidimensional multicellular model that couples, on the one hand, water fluxes between cells (symplastic pathway) and between cells and intercellular space (apoplastic pathway), and on the other hand, mechanical properties of the cell walls and mechanical equilibrium of this complex system. Existing multicellular models for plant growth overlook this coupling. From a mathematical point of view, it corresponds to a coupling between (1) the ordinary differential equations that describe fluxes and cell walls properties and (2) the highly non linear system of equations that describes the mechanical equilibrium of the cell walls.

We have developed a numerical method for this coupled system, that allows to simulate in a reasonable amount of time a hundred of connected cells. Numerical simulations exhibit a highly non linear behaviour with respect to the governing parameters. Thanks to the detailed analysis of a simplified setup, we have identified two clearly distinct growth regimes: one regime that allows large growth heterogeneities by amplifying the effect of differences between cells, and conversely another regime that smoothes differences out and yields a homogeneous growth. On the biological level, the first regime is well adapted to morphogenesis, whereas the second one is well adapted to homothetic growth after the differentiated tissues have been created. A publication of these completely new results is in preparation.

We have developped a collaboration with biophysicists in RDP laboratory in Lyon (Arezki Boudaoud and Yuchen Long) in order to compare the results of this model to experiments at the microscopic scale of the meristem. A publication is in preparation.

In the longer term, we plan extend this model to the larger scale of tissues and organs in order to model fruit growth.

6.3.4. 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 was supported by two PhD programmes.

New 2D and 3D root phenotyping plateforms are emerging with associated image analysis toolbox (e.g. SmartRoot, RhizoScan) and the identification of developmental patterns within these complex phenotyping data requires new approaches. Here, we aim at developing a pipeline of methods for analyzing root systems at three scales:

1. tissular scale to identify and characterize the division, elongation and mature zones along a root apex using piecewise heteroscedastic linear models for segmenting epidermal cell length profiles [17].
2. individual root scale to analyze the dynamics of lateral root elongation. We in particular applied semi-Markov switching linear models for classifying roots on the basis of the identification of phases within growth rate profiles,
3. root system scale to analyze the primary root branching structure using variable-order Markov chains.

This pipeline of analysis methods was applied to different species (maize, Pearl millet [18]) with contrasting biological objectives (study of genetic diversity for Pearl millet and of metabolic and hormonal controls of morphogenesis for maize).

6.3.5. 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. In the last year, it has also been demonstrated that sugar effect on bud outgrowth was preceded by a modification of the hormonal levels involved in bud outgrowth, which suggests that auxin and sugar pathways do interact in a non- trivial way. However, 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 regulate bud entrance into sustained growth. For this, a series of experiments has been carried out 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 of the regulatory networks controlling stem-bud molecular interaction has been developed.

6.4. Generic methodological results

6.4.1. OpenAlea scientific workflows and grid computing

Participants: Christophe Pradal, Sarah Cohen-Boulakia, Christian Fournier, Didier Parigot [Inria, Zenith], Patrick Valduriez [Inria, Zenith].

This research theme has been supported by IBC.

Plant phenotyping consists in the observation of physical and biochemical traits of plant genotypes in response to environmental conditions. Challenges, in particular in context of climate change and food security, are numerous. High-throughput platforms have been introduced to observe the dynamic growth of a large number of plants in different environmental conditions. Instead of considering a few genotypes at a time (as it is the case when phenomic traits are measured manually), such platforms make it possible to use completely new kinds of approaches. However, the data sets produced by such widely instrumented platforms are huge, constantly augmenting and produced by increasingly complex experiments, reaching a point where distributed computation is mandatory to extract knowledge from data. We design the infrastructure InfraPhenoGrid [42] to efficiently manage data sets produced by the PhenoArch plant phenomics platform in the context of the French Phenome Project. Our solution consists in deploying *OpenAlea* scientific workflows on a Grid using a middleware, SciFloware, to pilot workflow executions. Our approach is user-friendly in the sense that despite the intrinsic complexity of the infrastructure, running scientific workflows and understanding results obtained (using provenance information) is kept as simple as possible for end-users.

6.4.2. Reproductibility in the Life Science with Scientific workflows

Participants: Christophe Pradal, Sarah Cohen-Boulakia, Jerome Chopard.

This research theme has been supported by IBC and the GDR MADICS/ReproVirtuFlow.

With the development of new experimental technologies, biologists are faced with an avalanche of data to be computationally analyzed for scientific advancements and discoveries to emerge. Faced with the complexity of analysis pipelines, the large number of computational tools, and the enormous amount of data to manage, there is compelling evidence that many if not most scientific discoveries will not stand the test of time: increasing the reproducibility of computed results is of paramount importance. In the context of the project 7.2.5.3 , we study how scientific workflows can help to improve the reproducibility of computational experiment in the domain of life science [34]. We characterize and define the criteria that need to be catered for by *reproducibility-friendly* scientific workflow systems, and use such criteria to place several representative and widely used workflow systems and companion tools within such a framework.

6.4.3. Statistical modeling

Participants: Yann Guédon, Jean Peyhardi, Pierre Fernique, Jean-Baptiste Durand Peyhardi, Catherine Trottier [IMAG, Montpellier].

We develop statistical models and methods for identifying and characterizing developmental patterns in plant phenotyping data. Phenotyping data are very diverse ranging from the tissular to the whole plant scale but are often highly structured in space, time and scale. Problems of interest deal with the definition of new family of statistical models specifically adapted to plant phenotyping data and the design of new methods of inference concerning both model structure, model parameters and latent structure. This is illustrated this year by [41].

6.4.4. Lossy compression of tree structures

Participants: Christophe Godin, Romain Azaïs, Jean-Baptiste Durand, Alain Jean-Marie.

In in [13], we defined the degree of self-nestedness of a tree 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 introduced a lossy compression method that consists in computing in polynomial time for trees with bounded outdegree the reduction of a self-nested tree that closely approximates the initial tree. This approximation relies on an indel edit distance that allows (recursive) insertion and deletion of leaf vertices only. We showed in a conference paper presented at DCC'2016 [48] with a simulated dataset that the error rate of this lossy compression method is always better than the loss based on the nearest embedded self-nestedness tree [13] while the compression rates are equivalent. This procedure is also a keystone in our new topological clustering algorithm for trees. In addition, we obtained new theoretical results on the combinatorics of self-nested structures and their ability to approximate complex trees in a costless manner [4].

6.4.5. Version climber

Participants: Christophe Padal, Dennis Shasha, Sarah Cohen-Boulakia, Patrick Valduriez.

This research has been supported by the Inria International Chair of Dennis Shasha.

Imagine you are a data scientist (as many of us are/have become). Systems you build typically require many data sources and many packages (machine learning/data mining, data management, and visualization) to run. Your working configuration will consist of a set of packages each at a particular version. You want to update some packages (software or data) to their most recent possible version, but you want your system to run after the upgrades, thus perhaps entailing changes to the versions of other packages.

One approach is to hope the latest versions of all packages work. If that fails, the fallback is manual trial and error, but that quickly ends in frustration.

We advocate a provenance-style approach in which tools like *ptrace* and *reprozip*, combine to enable us to identify version combinations of different packages. Then other tools like *pip* and *VirtualEnv* enable us to fetch particular versions of packages and try them in a sandbox-like environment.

Because the space of versions to explore grows exponentially with the number of packages, we have developed a memorizing algorithm that avoids exponential search while still finding an optimum version combination.

Experimental results have been tested (with full reproducibility) on well known packages used in data science to illustrate the effectiveness of our approach as well as life science computational experiment.

6.4.6. Automatic generation of python bindings for C++ libraries

Participants: Pierre Fernique, Christophe Padal.

This research has been supported by the Inria ADT SCOOP

Most of Python and R scientific packages incorporate compiled scientific libraries to speed up the code and reuse legacy libraries. While several semi-automatic solutions exist to wrap these compiled libraries, the process of wrapping a large library is cumbersome and time consuming. In this paper, we introduce AutoWIG, a Python package that wraps automatically compiled libraries into high-level languages using LLVM/Clang technologies and the Mako templating engine. Our approach [46] is automatic, extensible, and applies to complex C++ libraries, composed of thousands of classes or incorporating modern meta-programming constructs.

ARAMIS Project-Team

7. New Results

7.1. Fiberprint: A subject fingerprint based on sparse code pooling for white matter fiber analysis

Participants: Kuldeep Kumar [Correspondant], Christian Desrosiers, Kaleem Siddiqi, Olivier Colliot, Matthew Toews.

White matter characterization studies use the information provided by diffusion magnetic resonance imaging (dMRI) to draw cross-population inferences. However, the structure, function, and white matter geometry vary across individuals. Here, we propose a subject fingerprint, called Fiberprint, to quantify the individual uniqueness in white matter geometry using fiber trajectories. We learn a sparse coding representation for fiber trajectories by mapping them to a common space defined by a dictionary. A subject fingerprint is then generated by applying a pooling function for each bundle, thus providing a vector of bundle-wise features describing a particular subject's white matter geometry. These features encode unique properties of fiber trajectories, such as their density along prominent bundles. An analysis of data from 861 Human Connectome Project subjects reveals that a fingerprint based on approximately 3000 fiber trajectories can uniquely identify exemplars from the same individual. We also use fingerprints for twin/sibling identification, our observations consistent with the twin data studies of white matter integrity. Our results demonstrate that the proposed Fiberprint can effectively capture the variability in white matter fiber geometry across individuals, using a compact feature vector (dimension of 50), making this framework particularly attractive for handling large datasets.

More details in [21].

7.2. Individual analysis of molecular brain imaging data through automatic identification of abnormality patterns

Participants: Ninon Burgos [Correspondant], Jorge Samper-González, Anne Bertrand, Marie-Odile Habert, Sébastien Ourselin, Stanley Durrleman, M. Jorge Cardoso, Olivier Colliot.

We introduce a pipeline for the individual analysis of positron emission tomography (PET) data on large cohorts of patients. This pipeline consists for each individual of generating a subject-specific model of healthy PET appearance and comparing the individual's PET image to the model via a novel regularised Z-score. The resulting voxel-wise Z-score map can be interpreted as a subject-specific abnormality map that summarises the pathology's topographical distribution in the brain. We then propose a strategy to validate the abnormality maps on several PET tracers and automatically detect the underlying pathology by using the abnormality maps as features to feed a linear support vector machine (SVM)-based classifier. We applied the pipeline to a large dataset comprising 298 subjects selected from the ADNI2 database (103 cognitively normal, 105 late MCI and 90 Alzheimer's disease subjects). The high classification accuracy obtained when using the abnormality maps as features demonstrates that the proposed pipeline is able to extract for each individual the signal characteristic of dementia from both FDG and Florbetapir PET data.

More details in [27].

7.3. Multilevel Modeling with Structured Penalties for Classification from Imaging Genetics data

Participants: Pascal Lu [Correspondant], Olivier Colliot.

In this paper, we propose a framework for automatic classification of patients from multimodal genetic and brain imaging data by optimally combining them. Additive models with unadapted penalties (such as the classical group lasso penalty or L_1-multiple kernel learning) treat all modalities in the same manner and can result in undesirable elimination of specific modalities when their contributions are unbalanced. To overcome this limitation, we introduce a multilevel model that combines imaging and genetics and that considers joint effects between these two modalities for diagnosis prediction. Furthermore, we propose a framework allowing to combine several penalties taking into account the structure of the different types of data, such as a group lasso penalty over the genetic modality and a L_2-penalty on imaging modalities. Finally , we propose a fast optimization algorithm, based on a proximal gradient method. The model has been evaluated on genetic (single nucleotide polymorphisms-SNP) and imaging (anatomical MRI measures) data from the ADNI database, and compared to additive models. It exhibits good performances in AD diagnosis; and at the same time, reveals relationships between genes, brain regions and the disease status.

More details in [33].

7.4. Towards Fully-reproducible Research on Classification of Alzheimer's Disease

Participants: Jorge Samper-González [Correspondant], Ninon Burgos, Sabrina Fontanella, Hugo Bertin, Marie-Odile Habert, Stanley Durrleman, Theodoros Evgeniou, Olivier Colliot.

In recent years, the number of papers on Alzheimer's disease classification has increased dramatically, generating interesting methodological ideas on the use machine learning and feature extraction methods. However, practical impact is much more limited and, eventually, one could not tell which of these approaches are the most efficient. While over 90% of these works make use of ADNI an objective comparison between approaches is impossible due to variations in the subjects included, image pre-processing, performance metrics and cross-validation procedures. In this paper, we propose a framework for reproducible classification experiments using multimodal MRI and PET data from ADNI. The core components are: 1) code to automatically convert the full ADNI database into BIDS format; 2) a modular architecture based on Nipype in order to easily plug-in different classification and feature extraction tools; 3) feature extraction pipelines for MRI and PET data; 4) baseline classification approaches for unimodal and multimodal features. This provides a flexible framework for benchmarking different feature extraction and classification tools in a reproducible manner. Data management tools for obtaining the lists of subjects in AD, MCI converter, MCI non-converters, CN classes are also provided. We demonstrate its use on all (1519) baseline T1 MR images and all (1102) baseline FDG PET images from ADNI 1, GO and 2 with SPM-based feature extraction pipelines and three different classification techniques (linear SVM, anatomically regularized SVM and multiple kernel learning SVM). The highest accuracies achieved were: 91% for AD vs CN, 83% for MCIc vs CN, 75% for MCIc vs MCInc, 94% for AD-ABeta+ vs CN-ABeta- and 72% for MCIc-ABeta+ vs MCInc-ABeta+. The code is publicly available at https://gitlab.icm-institute.org/aramislab/AD-ML.

More details in [34].

7.5. Early Cognitive, Structural, and Microstructural Changes in Presymptomatic C9orf72 Carriers Younger Than 40 Years

Participants: Anne Bertrand [Correspondant], Junhao Wen, Sabrina Fontanella, Alexandre Routier, Stanley Durrleman, Olivier Colliot.

Presymptomatic carriers of chromosome 9 open reading frame 72 (C9orf72) mutation, the most frequent genetic cause of frontotemporal lobar degeneration and amyotrophic lateral sclerosis, represent the optimal target population for the development of disease-modifying drugs. Preclinical biomarkers are needed to monitor the effect of therapeutic interventions in this population. The aim of our study was to assess the occurrence of cognitive, structural, and microstructural changes in presymptomatic C9orf72 carriers. The PREV-DEMALS study is a prospective, multicenter, observational study of first-degree relatives of individuals

carrying the C9orf72 mutation. Eighty-four participants entered the study between October 2015 and April 2017; 80 (95%) were included in cross-sectional analyses of baseline data. All participants underwent neuropsychological testing and magnetic resonance imaging; 63 (79%) underwent diffusion tensor magnetic resonance imaging. Gray matter volumes and diffusion tensor imaging metrics were calculated within regions of interest. Anatomical and microstructural differences between individuals who carried the C9orf72 mutation (C9+) and those who did not carry the C9orf72 mutation (C9-) were assessed using linear mixed-effects models. Data were analyzed from October 2015 to April 2017.Of the 80 included participants, there were 41 C9+ individuals (24 [59%] female; mean [SD] age, 39.8 [11.1] years) and 39 C9- individuals (24 [62%] female; mean [SD] age,45.2 [13.9] years). Compared with C9- individuals, C9+ individuals had lower mean (SD)praxis scores (163.4 [6.1] vs 165.3 [5.9]; P = .01) and intransitive gesture scores (34.9 [1.6] vs 35.7 [1.5]; P = .004), atrophy in 8 cortical regions of interest and in the right thalamus, and white matter alterations in 8 tracts. When restricting the analyses to participants younger than 40 years, compared with C9- individuals, C9+ individuals had lower praxis scores and intransitive gesture scores, atrophy in 4 cortical regions of interest and in the right thalamus, and white matter alterations in 2 tracts.Our work demontrates that cognitive, structural and microstructural alterations are detectable in young C9+ individuals. Early and subtle praxis alterations, underpinned by focal atrophy of the left supramarginal gyrus, may represent an early and nonevolving phenotype related to neurodevelopmental effects of C9orf72 mutation. White matter alterations reflect the future phenotype of frontotemporal lobar degeneration/amyotrophic lateral sclerosis, while atrophy appears more diffuse. Our results contribute to a better understanding of the preclinical phase of C9orf72

More details in [5].

7.6. Loss of brain inter-frequency hubs in Alzheimer's disease

Participants: Jeremy Guillon, Yohan Attal, Olivier Colliot, Valentina La Corte, Bruno Dubois, Denis Schwartz, Mario Chavez, Fabrizio de Vico Fallani [Correspondant].

Alzheimer's disease (AD) causes alterations of brain network structure and function. The latter consists of connectivity changes between oscillatory processes at different frequency channels. We proposed a multi-layer network approach to analyze multiple-frequency brain networks inferred from magnetoencephalographic recordings during resting-states in AD subjects and age-matched controls. Main results showed that brain networks tend to facilitate information propagation across different frequencies, as measured by the multi-participation coefficient (MPC). However, regional connectivity in AD subjects was abnormally distributed across frequency bands as compared to controls, causing significant decreases of MPC. This effect was mainly localized in association areas and in the cingulate cortex, which acted, in the healthy group, as a true inter-frequency hub. MPC values significantly correlated with memory impairment of AD subjects, as measured by the total recall score. Most predictive regions belonged to components of the default-mode network that are typically affected by atrophy, metabolism disruption and amyloid-β deposition. We evaluated the diagnostic power of the MPC and we showed that it led to increased classification accuracy (78.39%) and sensitivity (91.11%). These findings shed new light on the brain functional alterations underlying AD and provide analytical tools for identifying multi-frequency neural mechanisms of brain diseases.

More details in [17].

7.7. A statistical model for brain networks inferred from large-scale electrophysiological signals

Participants: Catalina Obando, Fabrizio de Vico Fallani [Correspondant].

Network science has been extensively developed to characterize the structural properties of complex systems, including brain networks inferred from neuroimaging data. As a result of the inference process, networks estimated from experimentally obtained biological data represent one instance of a larger number of realizations with similar intrinsic topology. A modelling approach is therefore needed to support statistical inference on the bottom-up local connectivity mechanisms influencing the formation of the estimated brain networks. Here, we adopted a statistical model based on exponential random graph models (ERGMs) to reproduce brain

networks, or connectomes, estimated by spectral coherence between high-density electroencephalographic (EEG) signals. ERGMs are made up by different local graph metrics, whereas the parameters weight the respective contribution in explaining the observed network. We validated this approach in a dataset of N ¼ 108 healthy subjects during eyes-open (EO) and eyes closed (EC) resting-state conditions. Results showed that the tendency to form triangles and stars, reflecting clustering and node centrality, better explained the global properties of the EEG connectomes than other combinations of graph metrics. In particular, the synthetic networks generated by this model configuration replicated the characteristic differences found in real brain networks, with EO eliciting significantly higher segregation in the alpha frequency band (8–13 Hz) than EC. Furthermore, the fitted ERGM parameter values provided complementary information showing that clustering connections are significantly more represented from EC to EO in the alpha range, but also in the beta band (14–29 Hz), which is known to play a crucial role in cortical processing of visual input and externally oriented attention. Taken together, these findings support the current view of the functional segregation and integration of the brain in terms of modules and hubs, and provide a statistical approach to extract new information on the (re)organizational mechanisms in healthy and diseased brains. More details in [23].

7.8. A Topological Criterion for Filtering Information in Complex Brain Networks

Participants: Fabrizio de Vico Fallani [Correspondant], Vito Latora, Mario Chavez.

In many biological systems, the network of interactions between the elements can only be inferred from experimental measurements. In neuroscience, non-invasive imaging tools are extensively used to derive either structural or functional brain networks in-vivo. As a result of the inference process, we obtain a matrix of values corresponding to a fully connected and weighted network. To turn this into a useful sparse network, thresholding is typically adopted to cancel a percentage of the weakest connections. The structural properties of the resulting network depend on how much of the inferred connectivity is eventually retained. However, how to objectively fix this threshold is still an open issue. We introduce a criterion, the efficiency cost optimization (ECO), to select a threshold based on the optimization of the trade-off between the efficiency of a network and its wiring cost. We prove analytically and we confirm through numerical simulations that the connection density maximizing this trade-off emphasizes the intrinsic properties of a given network, while preserving its sparsity. Moreover, this density threshold can be determined a-priori, since the number of connections to filter only depends on the network size according to a power-law. We validate this result on several brain networks, from micro- to macro-scales, obtained with different imaging modalities. Finally, we test the potential of ECO in discriminating brain states with respect to alternative filtering methods. ECO advances our ability to analyze and compare biological networks, inferred from experimental data, in a fast and principled way.

More details in [11].

7.9. Preclinical Alzheimer's disease: a systematic review of the cohorts underlying the concept

Participants: Stéphane Epelbaum [Correspondant], Remy Genthon, Enrica Cavedo, Marie Odile Habert, Foudil Lamari, Geoffroy Gagliardi, Simone Lista, Marc Teichmann, Hovagim Bakardjian, Harald Hampel, Bruno Dubois.

Preclinical Alzheimer's disease (AD) is a relatively recent concept describing an entity characterized by the presence of a pathophysiological biomarker signature characteristic for AD in the absence of specific clinical symptoms. There is rising interest in the scientific community to define such an early target population mainly due to failures of all recent clinical trials despite evidence of biological effects on brain amyloidosis for some compounds. A conceptual framework has recently been proposed for this preclinical phase of AD. However, few data exist on this silent stage of AD. We performed a systematic review in order to investigate how the concept is defined across studies. The review highlights the substantial heterogeneity concerning the three main determinants of preclinical AD: "normal cognition", "cognitive decline" and "AD pathophysiological signature". We emphasize the need for a harmonized nomenclature of the preclinical AD concept and standardized population-based and case-control studies using unified operationalized criteria.

More details in [12].

7.10. Free and Cued Selective Reminding Test - accuracy for the differential diagnosis of Alzheimer's and neurodegenerative diseases: A large-scale biomarker-characterized monocenter cohort study (ClinAD)

Participants: Marc Teichmann [Correspondant], Stéphane Epelbaum, Dalila Samri, Marcel Levy Nogueira, Agnes Michon, Harald Hampel, Foudil Lamari, Bruno Dubois.

The International Working Group recommended the Free and Cued Selective Reminding Test (FCSRT) as a sensitive detector of the amnesic syndrome of the hippocampal type in typical Alzheimer's disease (AD). But does it differentiate AD from other neurodegenerative diseases? We assessed the FCSRT and cerebrospinal fluid (CSF) AD biomarkers in 992 cases. Experts, blinded to biomarker data, attributed in 650 cases a diagnosis of typical AD, frontotemporal dementia, posterior cortical atrophy, Lewy body disease, progressive supranuclear palsy, corticobasal syndrome, primary progressive aphasias, "subjective cognitive decline," or depression. The FCSRT distinguished typical AD from all other conditions with a sensitivity of 100% and a specificity of 75%. Non-AD neurodegenerative diseases with positive AD CSF biomarkers ("atypical AD") did not have lower FCSRT scores than those with negative biomarkers. The FCSRT is a reliable tool for diagnosing typical AD among various neurodegenerative diseases. At an individual level, however, its specificity is not absolute. Our findings also widen the spectrum of atypical AD to multiple neurodegenerative conditions.

More details in [13].

7.11. Parallel transport in shape analysis : a scalable numerical scheme

Participants: Maxime Louis, Alexandre Bône, Benjamin Charlier, Stanley Durrleman.

The analysis of manifold-valued data requires efficient tools from Riemannian geometry to cope with the computational complexity at stake. This complexity arises from the always-increasing dimension of the data, and the absence of closed-form expressions to basic operations such as the Riemannian logarithm. In this work, we adapted a generic numerical scheme recently introduced for computing parallel transport along geodesics in a Riemannian manifold to finite-dimensional manifolds of diffeomorphisms. We provided a qualitative and quantitative analysis of its behavior on high-dimensional manifolds, and investigated an application with the prediction of brain structures progression.

More details in [32].

7.12. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks

Participants: Igor Koval, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière, Stanley Durrleman.

We introduced a mixed-effects model to learn spatiotemporal patterns on a network by considering longitudinal measures distributed on a fixed graph. The data come from repeated observations of subjects at different time points which take the form of measurement maps distributed on a graph such as an image or a mesh. The model learns a typical group-average trajectory characterizing the propagation of measurement changes across the graph nodes. The subject-specific trajectories are defined via spatial and temporal transformations of the group-average scenario, thus estimating the variability of spatiotemporal patterns within the group. To estimate population and individual model parameters, we adapted a stochastic version of the Expectation-Maximization algorithm, the MCMC-SAEM. The model was used to describe the propagation of cortical atrophy during the course of Alzheimer's Disease. Model parameters show the variability of this average pattern of atrophy in terms of trajectories across brain regions, age at disease onset and pace of propagation. We showed that the personalization of this model yields accurate prediction of maps of cortical thickness in patients.

More details in [29]

7.13. Prediction of the progression of subcortical brain structures in Alzheimer's disease from baseline

Participants: Alexandre Bône, Maxime Louis, Alexandre Routier, Jorge Samper, Michael Bacci, Benjamin Charlier, Olivier Colliot, Stanley Durrleman.

We proposed a method to predict the subject-specific longitudinal progression of brain structures extracted from baseline MRI, and evaluated its performance on Alzheimer's disease data. The disease progression is modeled as a trajectory on a group of diffeomorphisms in the context of large deformation diffeomorphic metric mapping (LDDMM). We first exhibited the limited predictive abilities of geodesic regression extrapolation on this group. Building on the recent concept of parallel curves in shape manifolds, we then introduced a second predictive protocol which personalizes previously learned trajectories to new subjects, and investigate the relative performances of two parallel shifting paradigms. This design only requires the baseline imaging data. Finally, coefficients encoding the disease dynamics are obtained from longitudinal cognitive measurements for each subject, and exploited to refine our methodology which was demonstrated to successfully predict the follow-up visits.

More details in [28]

7.14. Prediction of amyloidosis from neuropsychological and MRI data for cost effective inclusion of pre-symptomatic subjects in clinical trials

Participants: Manon Ansart, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, Stanley Durrleman.

We proposed a method for selecting pre-symptomatic subjects likely to have amyloid plaques in the brain, based on the automatic analysis of neuropsychological and MRI data and using a cross-validated binary classifier. By avoiding systematic PET scan for selecting subjects, it reduces the cost of forming cohorts of subjects with amyloid plaques for clinical trials, by scanning fewer subjects but increasing the number of recruitments. We validated our method on three cohorts of subjects at different disease stages, and compared the performance of six classifiers, showing that the random forest yields good results more consistently, and that the method generalizes well when tested on an unseen data set.

More details in [25]

7.15. Geodesic shape regression with multiple geometries and sparse parameters

Participants: James Fishbaugh, Stanley Durrleman, Marcel Prastawa, Guido Gerig.

Many problems in medicine are inherently dynamic processes which include the aspect of change over time, such as childhood development, aging, and disease progression. From medical images, numerous geometric structures can be extracted with various representations, such as landmarks, point clouds, curves, and surfaces. Different sources of geometry may characterize different aspects of the anatomy, such as fiber tracts from DTI and subcortical shapes from structural MRI, and therefore require a modeling scheme which can include various shape representations in any combination. In this paper, we present a geodesic regression model in the large deformation (LDDMM) framework applicable to multi-object complexes in a variety of shape representations. Our model decouples the deformation parameters from the specific shape representations, allowing the complexity of the model to reflect the nature of the shape changes, rather than the sampling of the data. As a consequence, the sparse representation of diffeomorphic flow allows for the straightforward embedding of a variety of geometry in different combinations, which all contribute towards the estimation of a single deformation of the ambient space. Additionally, the sparse representation along with the geodesic constraint results in a compact statistical model of shape change by a small number of parameters defined by the user. Experimental validation on multi-object complexes demonstrate robust model estimation across a variety of parameter settings. We further demonstrate the utility of our method to support the analysis of derived shape features, such as volume, and explore shape model extrapolation. Our method is freely available in the software package deformetrica which can be downloaded at www.deformetrica.org.

More details in [14]

7.16. A sub-Riemannian modular framework for diffeomorphism based analysis of shape ensembles

Participants: Barara Gris, Stanley Durrleman, Alain Trouvé.

Deformations, and diffeormophisms in particular, have played a tremendous role in the field of statistical shape analysis, as a proxy to measure and interpret differences between similar objects but with different shapes. Diffeomorphisms usually result from the integration of a flow of regular velocity fields, whose parameters have not enabled so far a full control of the local behaviour of the deformation. In this work, we propose a new mathematical and computational framework, in which diffeomorphisms are built on the combination of local deformation modules with few degrees of freedom. Deformation modules contribute to a global velocity field, and interact with it during integration so that the local modules are transported by the global diffeomorphic deformation under construction. Such modular diffeomorphisms are used to deform shapes and to provide the shape space with a sub-Riemannian metric. We then derive a method to estimate a Fréchet mean from a series of observations, and to decompose the variations in shape observed in the training samples into a set of elementary deformation modules encoding distinctive and interpretable aspects of the shape variability. We show how this approach brings new solutions to long lasting problems in the fields of computer vision and medical image analysis. For instance, the easy implementation of priors in the type of deformations offers a direct control to favor one solution over another in situations where multiple solutions may fit the observations equally well. It allows also the joint optimisation of a linear and a non-linear deformation between shapes, the linear transform simply being a particular type of modules. The proposed approach generalizes previous methods for constructing diffeomorphisms and opens up new perspectives in the field of statistical shape analysis.

More details in [16]

7.17. A Bayesian Framework for Joint Morphometry of Surface and Curve meshes in Multi-Object Complexes

Participants: Pietro Gori, Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman.

We present a Bayesian framework for atlas construction of multi-object shape complexes comprised of both surface and curve meshes. It is general and can be applied to any parametric deformation framework and to all shape models with which it is possible to define probability density functions (PDF). Here , both curve and surface meshes are modelled as Gaussian random varifolds , using a finite-dimensional approximation space on which PDFs can be defined. Using this framework , we can automatically estimate the parameters balancing data-terms and deformation regularity , which previously required user tuning. Moreover , it is also possible to estimate a well-conditioned covariance matrix of the deformation parameters. We also extend the proposed framework to data-sets with multiple group labels. Groups share the same template and their deformation parameters are modelled with different distributions. We can statistically compare the groups ' distributions since they are defined on the same space. We test our algorithm on 20 Gilles de la Tourette patients and 20 control subjects , using three sub-cortical regions and their incident white matter fiber bundles. We compare their morphological characteristics and variations using a single diffeomorphism in the ambient space. The proposed method will be integrated with the Deformetrica software package, publicly available at www.deformetrica.org.

More details in [15]

7.18. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations

Participants: Jean-Baptiste Schiratti, Stéphanie Allassonnière, Olivier Colliot, Stanley Durrleman.

We propose a generic Bayesian mixed-effects model to estimate the temporal progression of a biological phenomenon from observations obtained at multiple time points for a group of individuals. The progression is modeled by continuous trajectories in the space of measurements. Individual trajectories of progression result from spatiotemporal transformations of an average trajectory. These transformations allow to quantify the changes in direction and pace at which the trajectories are followed. The framework of Rieman-nian geometry allows the model to be used with any kind of measurements with smooth constraints. A stochastic version of the Expectation-Maximization algorithm is used to produce produce maximum a posteriori estimates of the parameters. We evaluate our method using series of neuropsychological test scores from patients with mild cognitive impairments later diagnosed with Alzheimer's disease, and simulated evolutions of symmetric positive definite matrices. The data-driven model of the impairment of cognitive functions shows the variability in the ordering and timing of the decline of these functions in the population. We show also that the estimated spatiotemporal transformations effectively put into correspondence significant events in the progression of individuals.

More details in [40]

ASCLEPIOS Project-Team

6. New Results

6.1. Medical Image Analysis

6.1.1. Segmentation and Anatomical Variability of the Cochlea from Medical Images

Participants: Thomas Demarcy [Correspondant], Hervé Delingette, Charles Raffaelli [CHU, Nice], Clair Vandersteen [IUFC, Nice], Dan Gnansia [Oticon Medical], Nicholas Ayache.

This work is supported by the National Association for Research in Technology (ANRT) through the CIFRE Grant 2013-1165 and Oticon Medical (Vallauris). This work is a collaboration with the Department of Ear Nose Throat Surgery (IUFC, Nice) and the Nice University Hospital (CHU).

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

- We introduced an automated and reproducible framework for cochlear shape analysis [13].
- We introduced a new cochlear segmentation method within a generative probabilistic Bayesian framework for CT images [6].
- We studied the shape variability with a large database of CT images (N = 987) and quantified the bilateral symmetry in cochlear anatomy.
- We provided a proof of concept for the estimation of postoperative cochlear implant electrode-array position from clinical CT (Fig. 1).

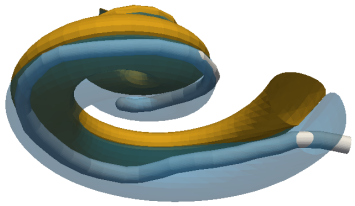


Figure 1. Cochlear implant electrode-array position (white) with respect to scala tympani (blue) and scala vestibuli (orange)

6.1.2. Prediction of Post-Ablation Outcome in Atrial Fibrillation Using Shape Parameterization and Partial Least Squares Regression

Participants: Shuman Jia [Correspondent], Claudia Camaioni, Marc Michel Rohe, Pierre Jaïs, Xavier Pennec, Hubert Cochet, Maxime Sermesant.

The authors acknowledge the partial funding by the Agence Nationale de la Recherche (ANR)/ERA CoSysMedSysAFib and ANR MIGAT projects.

We proposed an application of diffeomorphometry and partial least squares regression to address the problem of post-ablation outcome in atrial fibrillation.

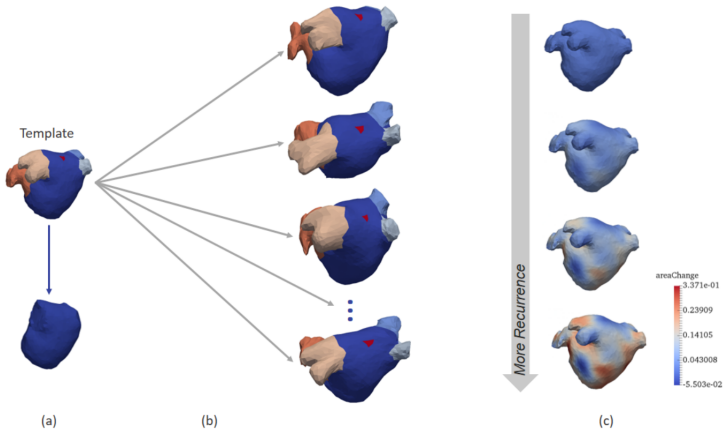


Figure 2. Extraction of remodeling information vs. recurrence. (a) average shape in control group; (b) deformation from the template to patient-specific shapes; (c) deformation mode correlated with recurrence.

As illustrated in Fig. 2 , we computed a template of left atrial shape in control group and then established point-to-point correspondence between patient-specific shapes and the template. The diffeomorphic deformations are encoded and applied in partial least squares regression to predict ablation success, which outperformed the left atrial volume index.

6.1.3. Cardiac Imaging and Machine Learning for Electrostructural Tomography

Participants: Tania Marina Bacoyannis [Correspondent], Hubert Cochet [IHU Liryc, Bordeaux], Maxime Sermesant.

This work is funded within the ERC Project ECSTATIC from the IHU Liryc, in Bordeaux.

Machine Learning, Cardiac modeling, Personalised simulation, Inverse problem of ECG, Electrical simulation, Inverse problem.

By using non-invasive electrical data (Body Surface Potential Mapping), we aim to develop a machine learning approach that can improve electrophysiological cardiac modeling in order to improve diagnosis and predict the response to therapy. This project involves measured and simulated data. For example, we processed experimental data provided by the IHU Liryc, gathered during an experiment on an healthy pig's heart (Figure 3). The simulated potentials appeared to be close to the measurements (Figure 4). The short-term goal is to reconstruct semi-automatically the simulated personalized activation maps.

6.1.4. VT-Scan: image based modelling of cardiac electrophysiology to guide catheter radiofrequency ablation of re-entrant ventricular tachycardia

Participants: Nicolas Cedilnik [Correspondent], Maxime Sermesant, Hubert Cochet, Pierre Jaïs, Frédéric Sacher.

This work was funded by IHU Liryc, Bordeaux.

cardiac electrophysiology modelling, cardiac imaging, ventricular tachycardia, catheter ablation, arrhythmia

- We used cardiac CT images to estimate infarct scar density and location using an automated thickness computation.
- A wavefront propagation speed was derived from this thickness in order to parametrize an Eikonal model of cardiac electrophysiology (see 5).

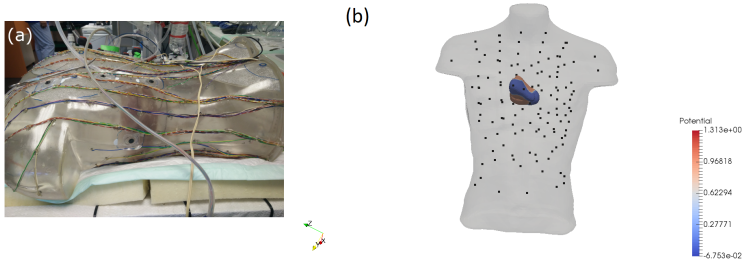


Figure 3. (a) Torso- tank experimental setup with perfused pig heart,(b) Representation of the experimental Torso for BSPM registration of the healthy pig heart, Torso's electrodes and the estimated activation map

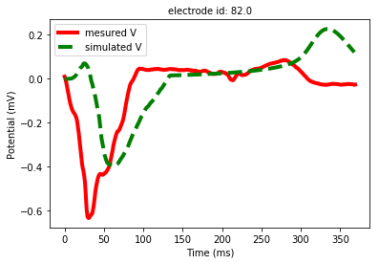


Figure 4. Example of BSPM signals on the Torso's electrode number 82

- We were able to match our simulations to recorded ventricular tachycardia patterns obtained during catheter ablation procedures, on 10 different ventricular tachycardias.
- This work was presented at the Functionnal Imaging and Modelling of the Heart conference in Toronto[34].

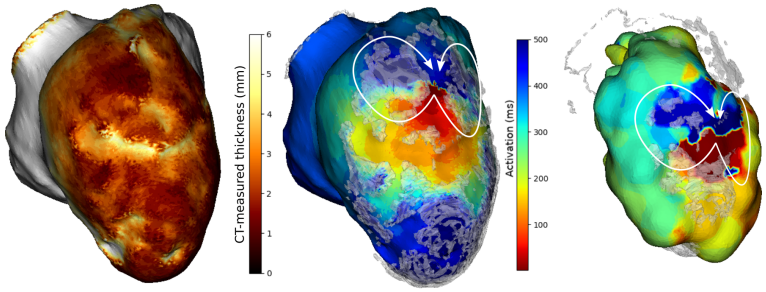


Figure 5. Example of a patient-specific simulation of a proven re-entrant wavefront propagation pattern. [Left] CT-measured myocardial wall thickness, projected on a CT-derived 3D mesh. [Middle] Myocardial activation simulation result using our framework. [Right] Activation recorded during a ventricular tachycardia catheter ablation

6.1.5. Deep Learning for Tumor Segmentation

Participants: Pawel Mlynarski [Correspondent], Nicholas Ayache, Hervé Delingette, Antonio Criminisi [MSR].

This work is funded by Inria-Microsoft joint center and is done in cooperation with Microsoft Research in Cambridge.

deep learning, semi-supervised learning, segmentation, MRI, tumors

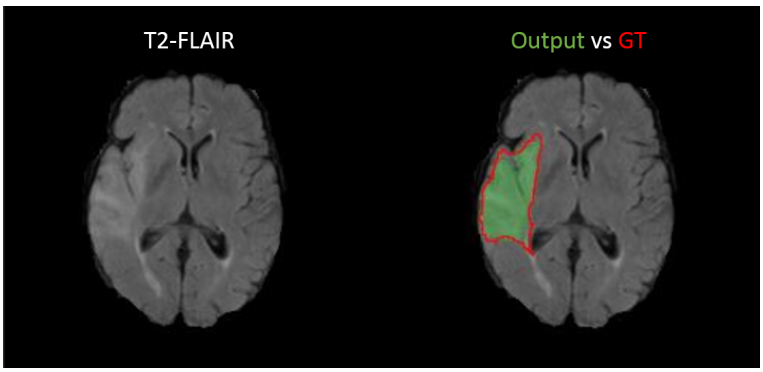


Figure 6. Left: axial slice of a brain MR image presenting a malignant tumor. Right: GT truth (red contour) vs segmentation produced by our semi-supervised method.

- We designed an algorithm for semi-supervised learning of neural nets for segmentation of tumors. The proposed system produces accurate binary segmentations (Figure 6) on unseen images with a limited number of ground truth segmentations used during the training phase.

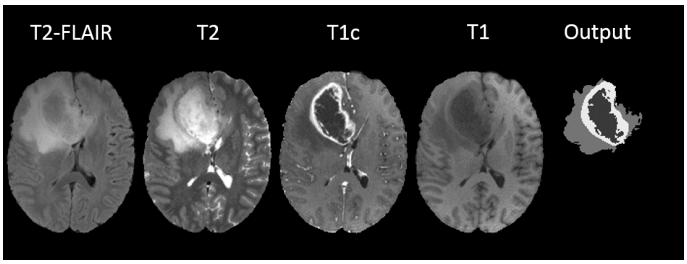


Figure 7. Multimodal MR and the multi-class tumor segmentation produced by our system.

- We proposed an efficient system based on Convolutional Neural Networks for multi-class segmentation of tumors in multimodal MR images (Figure 7). In particular, we proposed a new approach for treating different MR sequences and we introduced a new approach for ensembling 2D and 3D networks. We evaluated our method on the public benchmark of BRATS 2017 challenge and we obtained a top-3 performance among 60 participating teams.

6.1.6. Learning Brain Alterations in Multiple Sclerosis from Multimodal Neuroimaging Data

Participants: Wen Wei [Correspondent], Nicholas Ayache [Inria], Olivier Colliot [ARAMIS].

Multiple Sclerosis, MRI, PET

Multiple sclerosis (MS) is a demyelinating and inflammatory disease of the central nervous system . The goal of this topic is to develop a machine learning approach that can predict different types of PET-derived brain alterations using multiple local and regional MRI measures for MS patients. Figure 8 shows an example of multiple MRI pulse sequences for MS studies.

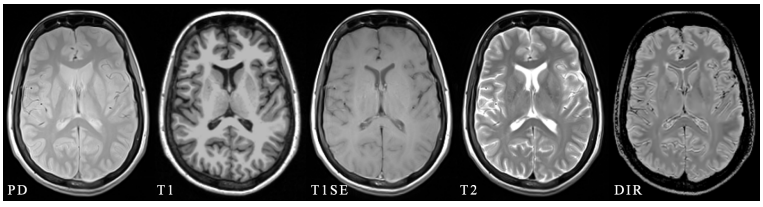


Figure 8. Proton density (PD), T1 spin-echo (T1SE), T1-w, T2-w, double inversion recovery (DIR) weighted images are used for MS studies.

6.1.7. Robust and 3D-Consistent Cardiac Segmentation by Deep Learning

Participants: Qiao Zheng [Correspondent], Hervé Delingette [Inria], Nicolas Duchateau [Université Claude Bernard Lyon 1], Nicholas Ayache [Inria].

Cardiac Segmentation, Deep Learning, MRI, Robustness, Consistency

We propose a method based on deep learning to perform cardiac segmentation on short axis MRI image stacks. An example of segmentation is presented in Figure 9 . The method is trained on a large database and then tested on other state-of-the-art cohorts. Results comparable or even better than the state-of-the-art in terms of distance measures are achieved. They prove the contribution of our method to enhance spatial consistency, and its generalization ability to unseen cases even from other databases.

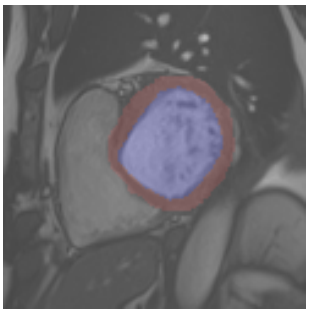


Figure 9. An example of cardiac segmentation.

6.1.8. Joint analysis of radiomic and metabolomic features to improve diagnosis and therapy in oncology

Participants: Fanny Orlhac [Correspondent], Charles Bouveyron, Hervé Delingette, Nicholas Ayache, Olivier Humbert [CAL], Jacques Darcourt [CAL], Thierry Pourcher [CEA], Fanny Vandenbos [CHU Nice].

Inria postdoctoral fellowship for 16 months

Radiomics, Metabolomics, Statistical learning

This work is done in collaboration with the Centre Antoine Lacassagne and the TIRO team (Transporter in Imagery and Radiotherapy for Oncology, CEA-UNS) located in Nice.

- The project consists to jointly analyze histogram, shape and textural features extracted from medical images (radiomics) and metabolomic data in oncology (see Figure 10).
- The goal is to better characterize tumor heterogeneity from both data sources in order to provide a personalized patient management.
- The work focuses on two pathologies: breast cancer and glioblastoma.

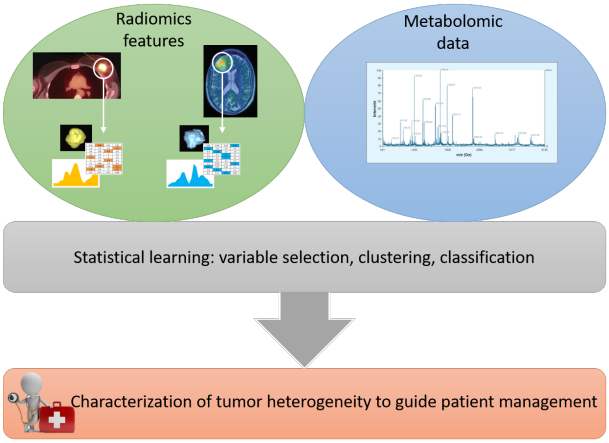


Figure 10. Joint analysis of radiomic and metabolomic features.

6.1.9. Heart & Brain: discovering the link between cardiovascular pathologies and neurodegeneration through biophysical and statistical models of cardiac and brain images.

Participants: Jaume Banús Cobo [Correspondent], Maxime Sermesant, Marco Lorenzi.

Uinversité Côte d'Azur (UCA)

Lumped models - Medical Imaging - Biophysical simulation - Machine learning

The project aims at developing a computational model of the relationship between cardiac function and brain damage from large-scale clinical databases of multi-modal and multi-organ medical images. We will use advanced statistical learning tools for discovering relevant imaging features related to cardiac dysfunction and brain damage from large datasets of medical images and clinical information; these measurements will be combined within a unified mechanistic framework to understand and validate the relationship between cardiac function, vascular pathology and brain damage. The goal is to provide an unprecedented instrument for the in-vivo assessment of latent neurodegenerative conditions in the general population, and will be validated with respect to established indices of cognitive decline and to specific sub-population for which the ground truth is known.

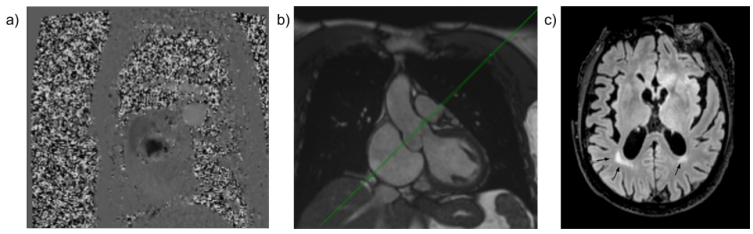


Figure 11. a) Aortic valve flow imaging, view obtained from the plane represented in the left ventricular outflow tract (LVOT) cine b); c) T2 FLAIR image in which white matter hyperintensities (WMHs) are visible.

6.1.10. Statistical learning on large databases of heterogeneous imaging, cognitive and behavioural data

Participants: Luigi Antelmi [Correspondant], Marco Lorenzi, Nicholas Ayache, Valeria Manera, Philippe Robert.

statistical learning, neuroimaging, big data, multimodal

The aim of our work is to develop scalable learning models for the joint analysis of heterogeneous biomedical data. The project will be applied to the investigation of neurological disorders from collections of brain imaging, body sensors, biological and clinical data available in current large-scale health databases. The resulting methodological framework will be tested on the UK Biobank, as well as on pathology-specific clinical data, as provided by the ADNI [0], or INSIGHT [0] initiatives.

From the methodological perspective, the project will focus on the development of computationally efficient formulations of probabilistic latent variable models. These approaches will highlight meaningful relationship among biomarkers that will be used to develop optimal strategies for disease quantification and prediction (Fig. 12).

[0] http://adni.loni.usc.edu/
[0] http://alzheimer-recherche.org/9248/etude-insight/

The research is within the MNC3 initiative (Médecine Numérique: Cerveau, Cognition, Comportement) funded by Université Côte d'Azur (UCA), and will be performed in collaboration with the Institut Claude Pompidou (CHU of Nice).

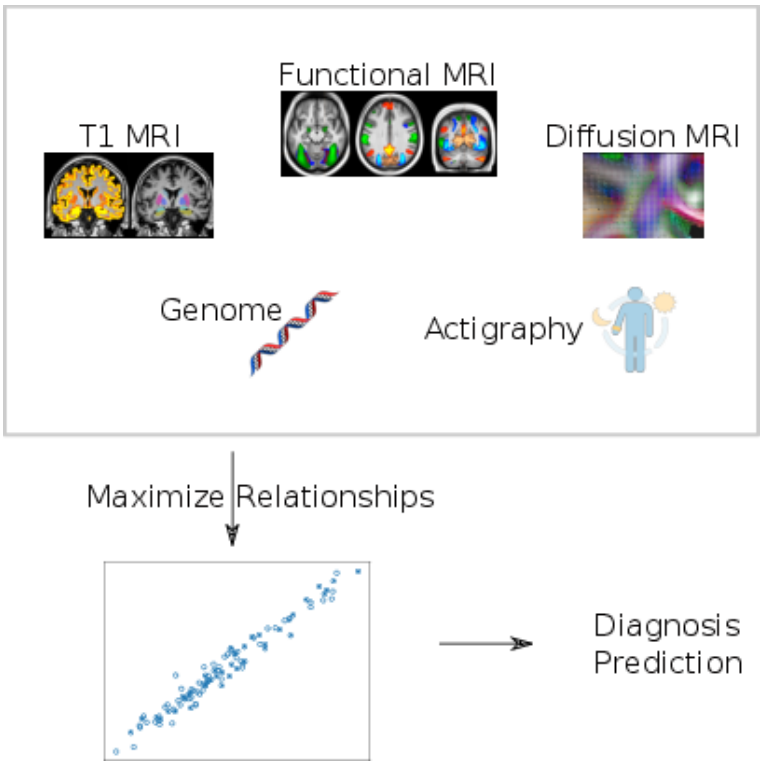


Figure 12. Adopted framework: development of computationally efficient formulations of probabilistic latent variable models to highlight meaningful relationship among biomarkers for disease quantification and prediction.

6.1.11. Robust non-rigid registration through agent-based action learning

Participants: Julian Krebs [Correspondent], Hervé Delingette, Tommaso Mansi [Siemens [Siemens Healthineers, Medical Imaging Technologies], Nicholas Ayache.

This PhD is carried out between the Asclepios research group, Inria Sophia Antipolis and Medical Imaging Technologies, Siemens Healthineers, Princeton, New Jersey, USA.

Deformable Registration, Deep Learning, Reinforcement Learning

We developed a deep learning-based approach for organ-specific deformable registration in 3-D [39] by:

- reformulating deformable registration as an agent-based learning problem (Fig. 13)
- using a low-parametric parametric statistical deformation model
- applying a novel ground truth generator which allows generating millions of synthetically deformed training samples requiring only a few real deformation estimations

Improved performance has been demonstrated with respect to traditional algorithms.

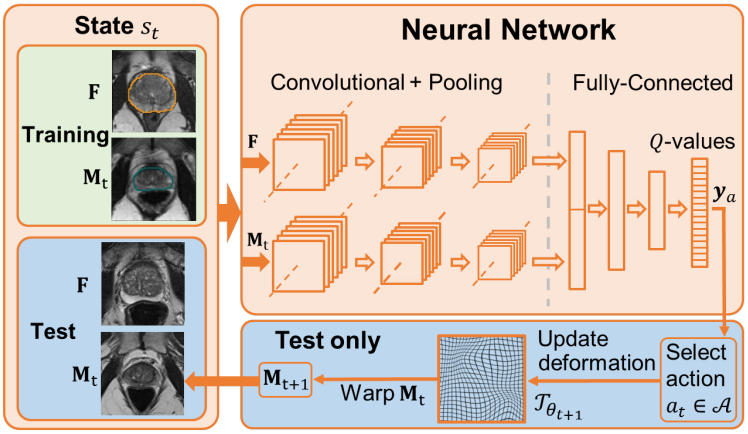


Figure 13. Dual-stream network used for the agent's action estimations including single-stage Markov Decision Process for testing (blue background).

6.2. Computational Anatomy

6.2.1. Inconsistency of Template Estimation in Quotient Spaces

Participants: Loïc Devilliers [Correspondent], Stéphanie Allassonnière [Université Paris-Descartes], Alain Trouvé [ENS Paris-Saclay], Xavier Pennec.

Inria postdoctoral fellowship for 16 months

Template estimation, Fréchet Mean, Quotient Spaces, Inconsistency, Consistency Bias

- A central issue in Computational Anatomy is to compute an unbiased template prototype of our data images (the template) in the presence of two effects: the noise in the ambient space and the unknown registration of the data. The template estimation is usually performed by minimizing the discrepancy after registration (and iterating), which corresponds geometrically to the computation of the Fréchet mean in the quotient space. So far, it was generally believed that the template estimation with this method was unbiased.
- We show in this work that inconsistency is in fact the general situation when the ambient space is an infinite dimensional linear space. In [15] we prove that this method is generally inconsistent when the action is isometric. Moreover the consistency bias has been quantified [35] thanks to a Taylor expansion in the noise level. Besides, we provide proofs of inconsistency for non isometric action [15] when the noise level is large enough.

6.2.2. Geometric statistics for Computational Anatomy

Participants: Nina Miolane [Correspondent], Xavier Pennec.

This work is conducted jointly with the Department of Statistics of Stanford, in the context of the associated team GeomStats of the program Inria@SiliconValley.

Statistics, Computational Anatomy, Differential Geometry, Template shape, asymptotic bias

The usual algorithm of brain template estimation is asymptotically biased, therefore inconsistent: even with an infinite number of brain images in the database, the template estimate may not converge to the brain anatomy it is meant to estimate. In [22]:

- we present a methodology that quantifies spatially the brain template's asymptotic bias, see Figure 14
,

- we propose a topologically constrained adaptation of the template computation, that constructs a hierarchical template with bounded bias, and we apply it to the Open Access Series of Imaging Studies (OASIS) database.

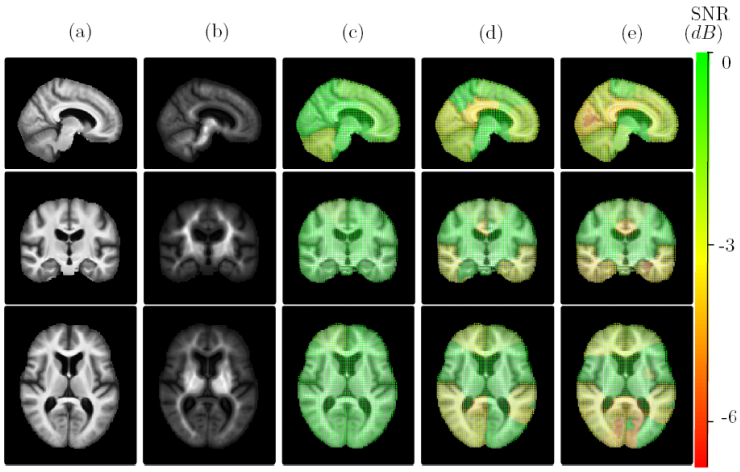


Figure 14. Here we investigate the brain template's consistency as an estimator of a unique anatomy, with respect to the signal-over-noise ratio (SNR) of different regions. The SNR is related to the ratio of the maximum difference in intensity of the region, on the intensity variability averaged on corresponding registered subjects. (a) Template, (b) Template whitened by the intersubject variability, (c) Region-wise inconsistency for a SNR threshold = 1.3, (d) for threshold = 2, (e) for threshold = 4 (dimensionless).

6.2.3. SVF-Net: Learning Deformable Registration Using Shape Matching

Participants: Marc Michel Rohe [Correspondent], Xavier Pennec, Maxime Sermesant.

The authors acknowledge the partial funding by the EU FP7-funded project MD-Paedigree (Grant Agreement 600932).

Registration, Deep Learning, Shape Matching

We propose an innovative approach for registration based on the deterministic prediction of the parameters from both images instead of the optimization of a energy criteria [44]. The method relies on a fully convolutional network (see Fig. 15). Whereas convolutional networks have seen a widespread expansion and have been already applied to many medical imaging problems such as segmentation and classification, its application to registration has so far faced the challenge of defining ground truth data on which to train the algorithm. Here, we present a novel training strategy to build reference deformations which rely on the registration of segmented regions of interest. The speed and robustness of this registration algorithm make it a strong candidate within a multi-atlas segmentation pipeline [45].

6.2.4. Reduced Representation of Segmentation and Tracking in Cardiac Images for Group-Wise Longitudinal Analysis

Participants: Marc Michel Rohe [Correspondant], Xavier Pennec, Maxime Sermesant.

The authors acknowledge the partial funding by the EU FP7-funded project MD-Paedigree (Grant Agreement 600932).

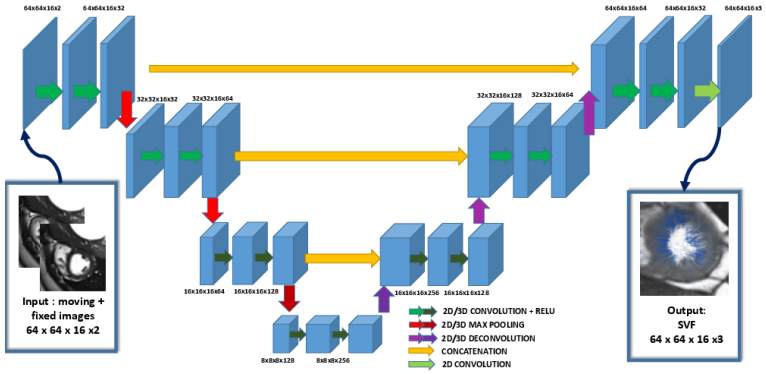


Figure 15. Fully convolutional neural networks for 3D registration: The inputs are the fixed and the moving 3D images. The output is a dense SVF symmetrically mapping the two images defined on the initial image grid.

Medical image analysis, Non-rigid registration, Deep learning, Statistical model reduction, Longitudinal analysis

We study image-based methods for the analysis of cardiac motion to enable group-wise statistics, automatic diagnosis and longitudinal study [10]. This is achieved by combining advanced medical image processing with machine learning methods and statistical modelling. The first axis of this work is to define an automatic method for the segmentation of the myocardium. The second axis of this work is focused on the improvement of cardiac motion tracking methods in order to define relevant low-dimensional representations. Finally, in the last axis, we apply the previously defined representation to the problem of diagnosis and longitudinal analysis. These three axes form an end to end framework for the study of cardiac motion starting from the acquisition of the medical images to their automatic analysis. Such a framework could be used for diagnosis and therapy planning in order to improve the clinical decision making with a more personalised computer-aided medicine.

6.2.5. *A model of brain morphological evolution*

Participants: Raphaël Sivera [Correspondent], Hervé Delingette, Marco Lorenzi, Xavier Pennec, Nicholas Ayache.

Longitudinal modeling, deformation framework, brain morphology, Alzheimer's disease, aging.

We proposed a deformation-based generative model of the brain morphological evolution that can joinly describes the effect of aging and Alzheimer's disease. It relies on longitudinal description of the aging and disease consequences and can be use to compute image-based cross-sectional progression markers (see Figure 16). This approach is able to propose a description of the disease evolution, population and subject-wise.

6.2.6. *Statistical Learning of Heterogeneous Data in Large-Scale Clinical Databases*

Participants: Clement Abi Nader [Correspondent], Nicholas Ayache, Marco Lorenzi.

The research takes place within the MNC3 initiative (Médecine Numérique: Cerveau, Cognition, Comportement) funded by Université Côte d'Azur (UCA), and is performed in collaboration with the Institut Claude Pompidou (CHU of Nice).

Longitudinal modeling, brain structure, Alzheimer's disease, aging, Gaussian processes, ICA.

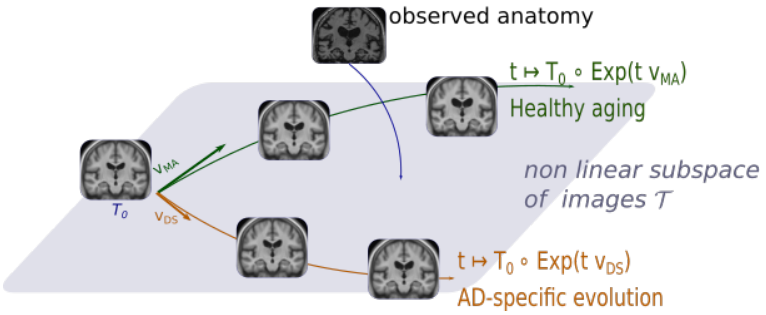


Figure 16. Left: z-values associated with group-wise differences between controls and Alzheimer-diagnosed subjects. Right: Areas of statistically significant differences.

Through this project we aim at developing novel scalable spatio-temporal analysis tools to identify clinical and biological modulators of structural and functional brain changes across time. The project relies on the extension of current un-/semi-supervised image analysis approaches (such as independent component analysis, ICA) to encode priors on spatial and temporal properties of the signal measured in brain images. The application to currently available large-scale biomedical datasets (such as the UK Biobank) will be addressed by focusing on scalable and distributed learning methods.

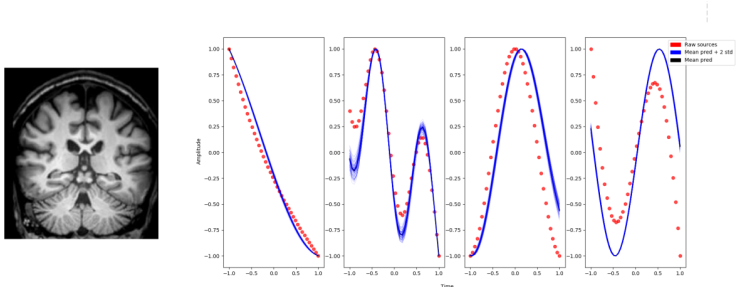


Figure 17. On the left, a coronal slice of a T1 weighted brain MRI. On the right we observe the temporal trajectories that best explain the evolution of the observed brain MRI times series from the UKBIOBANK study. Each trajectory representing the evolution of a meaningful biological and clinical sub-structure of the brain.

6.3. Computational Physiology

6.3.1. Non-invasive personalisation of a cardiac electrophysiology model from body surface potential mapping

Participants: Sophie Giffard Roisin [Correspondent], Maxime Sermesant, Nicholas Ayache, Hervé Delingette.

This work has been supported by the European Project FP7 under grant agreement VP2HF (no 611823) and the Marie Curie Actions European Industrial Doctorate CardioFunXion project (with Universitat Pompeu Fabra and Philips as partners).

Cardiac Modelling, Personalised Simulation, Inverse Problem of ECG, Electrical Simulation

Within the VP2HF project, non-invasive cardiac electrical data has been acquired at St Thomas' Hospital, London. It consists in Body Surface Potential Mapping (BSPM), which are recordings of the electrical potential on several locations on the surface of the torso. In [37], we use non-invasive data (body surface potential mapping, BSPM) to personalise complex cardiac electrical activation patterns such as multiple onset activation locations. We have used a relevance vector regression (see Figure 18) and we have evaluated our method on clinical datasets.

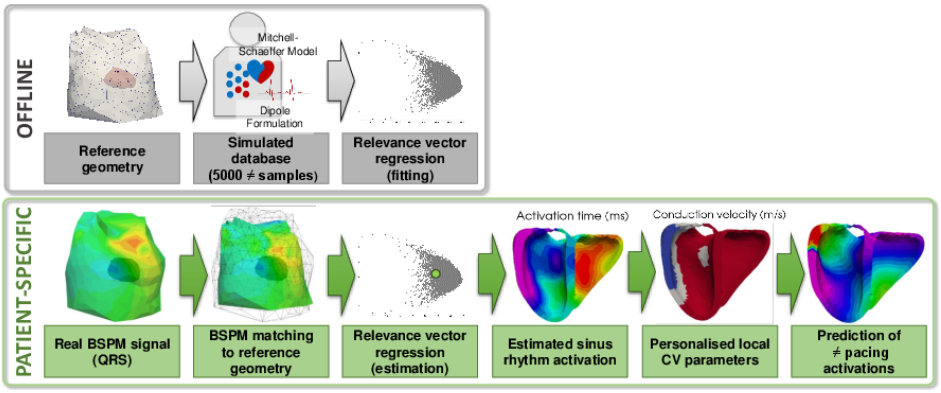


Figure 18. Pipeline of the non-invasive model personalisation

6.3.2. Mulfidelity-CMA Personalisation Algorithm and Personalised 3D Modeling for Longitudinal Analysis

Participants: Roch Philippe Molléro [Correspondent], Xavier Pennec, Hervé Delingette, Alan Garny, Nicholas Ayache, Maxime Sermesant.

This work has been partially funded by the EU FP7-funded project MD-Paedigree (Grant Agreement 600932) and contributes to the objectives of the ERC advanced grant MedYMA (2011-291080).

Cardiac Modelling, Personalised Simulation, Longitudinal Analysis, Parameter Estimation, Finite Element Mechanical Modelling

- We extended the multiscale 0D/3D personalisation approach previously published to build a fast, flexible and computationally efficient *multifidelity personalisation*. This algorithm called **Multifidelity-CMA** can be used to personalise hundreds of cases per day without specific manual supervision, fine-tuning of the algorithm or precomputation. The method was published in a scientific journal [24].
- We built more than **140 personalised 3D simulations** in the context of two longitudinal studies. We first used personalised parameters to model short-term transient effects in digestion ([33] and a poster presentation at FIMH Conference 2016), then to analyze long-term evolution of the cardiac function in cardiomyopathies ([42] and a poster presentation at MICCAI Conference 2017). In particular we showed that the use of priors reduces considerably the variance in the population of estimated parameters leading a better conditionning of parameter values whose variability in the population only reflects physiological properties of the cases. In particular we projected personalised parameters onto the axis of a classifier which discriminates between a cohort of healthy and diseased cases, and showed that the evolution of parameter values suggests an improvement of the cardiac function under therapy since the parameters of the follow-up acquisition are closer to the *healthy side* of the classifier (see Figure 19).

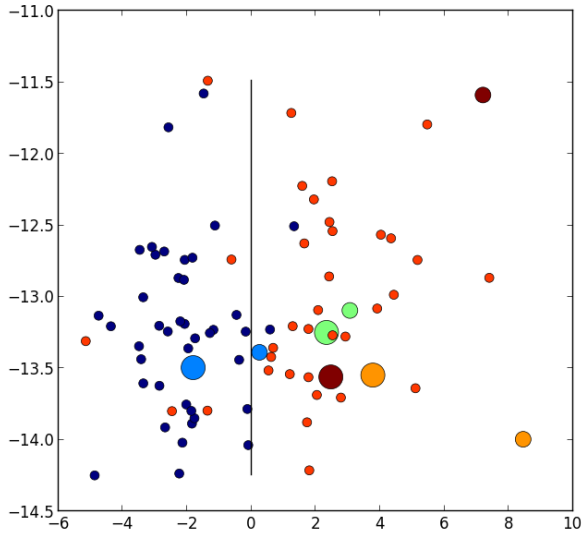


Figure 19. Projection of personalised parameters on the main direction of a LDA classifier between the healthy cases (dark blue dots) and cardiomyopathy (other dots) cases (x-axis) and an principal orthogonal direction of this vector (y-axis). The dots in light blue, brown, orange and green correspond to 4 patients for which the data was available both at baseline (small dot) and follow-up (larger dot).

ATHENA Project-Team

7. New Results

7.1. Computational Diffusion MRI

7.1.1. Spatio-Temporal dMRI Acquisition Design: Reducing the Number of qt Samples Through a Relaxed Probabilistic Model

Participants: Patryk Filipiak, Rutger Fick, Alexandra Petiet [ICM, CENIR, Paris], Mathieu Santin [ICM, CENIR, Paris], Anne-Charlotte Philippe [ICM, CENIR, Paris], Stephane Lehericy [ICM, CENIR, Paris], Demian Wassermann, Rachid Deriche.

Acquisition time is a major limitation in recovering brain microstructure with diffusion Magnetic Resonance Imaging. Finding a sampling scheme that maximizes signal quality and satisfies given time constraints is NP-hard. We alleviate that by introducing a relaxed probabilistic model of the problem, for which nearly-optimal solutions can be found effectively. Our model is defined in the qt-space, so that it captures both spacial and temporal phenomena. The experiments on in-vivo diffusion images of the C57Bl6 wild-type mice reveal superiority of our technique over random sampling and even distribution in the qt-space.

This work has been published in [33].

7.1.2. Diffusion MRI microstructure models with in vivo human brain Connectom data: results from a multi-group comparison

Participants: Uran Ferizi [CMIC, Dept. of Computer Science, UCL, UK], Rutger Fick, Rachid Deriche.

A large number of mathematical models have been proposed to describe the measured signal in diffusion-weighted (DW) magnetic resonance imaging (MRI) and infer properties about the white matter microstructure. However, a head-to-head comparison of DW-MRI models is critically missing in the field. To address this deficiency, we organized the "White Matter Modeling Challenge" during the International Symposium on Biomedical Imaging (ISBI) 2015 conference. This competition aimed at identifying the DW-MRI models that best predict unseen DW data. in vivo DW-MRI data was acquired on the Connectom scanner at the A.A.Martinos Center (Massachusetts General Hospital) using gradients strength of up to 300 mT/m and a broad set of diffusion times. We focused on assessing the DW signal prediction in two regions: the genu in the corpus callosum, where the fibres are relatively straight and parallel, and the fornix, where the configuration of fibres is more complex. The challenge participants had access to three-quarters of the whole dataset, and their models were ranked on their ability to predict the remaining unseen quarter of data. In this work, we provide both an overview and a more in-depth description of each evaluated model, report the challenge results, and infer trends about the model characteristics that were associated with high model ranking. This work provides a much needed benchmark for DW-MRI models. The acquired data and model details for signal prediction evaluation are provided online to encourage a larger scale assessment of diffusion models in the future.

This work has been published in [16].

7.1.3. Advanced dMRI signal modeling for tissue microstructure characterization

Participants: Rutger Fick, Demian Wassermann, Rachid Deriche.

Non-invasive estimation of brain white matter microstructure features using dMRI – otherwise known as Microstructure Imaging – has become an increasingly complex and difficult challenge over the last decade. Within the framework of Fick's PhD thesis [13], we contributed to the challenge to recover microstructure tissue parameters by studying the impact of using well-regularized functional basis together with multi-compartment approaches. We focused on the estimation and interpretation of microstructure-related markers, often referred to as *Microstructure Imaging* and we reviewed and compared most state-of-the-art microstructure models in PGSE-based Microstructure Imaging, emphasizing model assumptions and limitations, as well as validating them using spinal cord data with registered ground truth histology. We then presented contributions to 3D q-space imaging and microstructure recovery. We proposed closed-form Laplacian regularization for the recent MAP functional basis, allowing robust estimation of tissue-related q-space indices. We also applied this approach to Human Connectome Project data, where we used it as a preprocessing for other microstructure models. Finally, we compared tissue biomarkers in a ex-vivo study of Alzheimer rats at different ages. Last but not least, we contributed to representing the qt-space- varying over 3D q-space and diffusion time. Overall, we significantly contributed to the challenge of better understanding microstructure-related features of the brain's white matter.

This work has been published in [13].

7.1.4. White matter tractography guided by anatomical and microstructural priors

Participants: Gabriel Girard [SCIL, Sherbrooke University, CA], Maxime Descoteaux [SCIL, Sherbrooke University, CA], Demian Wassermann, Rachid Deriche.

In this work, performed within the framework of G. Girard's PhD thesis [81], we mainly focused in developing beyond the state-of-the-art and well grounded tractography solutions to recover the brain structural connectivity: We started reporting biases from tractography reconstruction and suggested to use anatomical priors, derived from a high resolution T1-weighted image to reduce these biases and to embed additional spatial information of the brain tissues in the tractography to guide tractography. We showed that optimizing tractography parameters, stopping and seeding strategies can reduce the biases in position, shape, size and length of the streamline distribution. Overall, we very nicely succeeded to show that this idea was able to significantly improve the tractography by reducing the rate of false positives produced and provides a more quantitative characterization of the WM structure. Going further, we then proposed to embed more intrinsic microstructural information in the reconstruction process and remarkably succeeded to show the great added value brought to tractography by the addition of intrinsic microstructural information such as the mean axonal diameter information estimated from the orientation of maximal diffusion probability. This is an original and important step forward in microstructure informed tractography, paving the way to a new generation of algorithms able to deal with intricate configurations of white matter fibres and providing quantitative brain connectivity analysis.

This work has been published in [13] and its part related to AxTract, the micro-informed tractography algorithm, in [19].

7.1.5. Rational invariants of ternary forms under the orthogonal group

Participants: Paul Görlach, Evelyne Hubert, Théodore Papadopoulo, Rachid Deriche.

In [79], [80], [95] we started to explore the theory of tensor invariants as a mathematical framework for computing new biomarkers for HARDI. We pursued this work and, in collaboration with the project-team GALAAD/AROMATH, we succeeded to develop a complete set of rational invariants for ternary quartics [44]. Being rational, they are very close to the polynomial invariants developed in [80] but they constitute a complete set of invariants. They are also good tools to understand better the algebraic invariants of [95] and some others based on spherical harmonics decomposition [61]. We determined a generating set of rational invariants of minimal cardinality for the action of the orthogonal group O3 on the space $R[x, y, z]_{2d}$ of ternary forms of even degree 2d. The construction relies on two key ingredients: On one hand, the Slice Lemma allows us to reduce the problem to dermining the invariants for the action on a subspace of the finite subgroup B3 of signed permutations. On the other hand, our construction relies in a fundamental way on specific bases

of harmonic polynomials. These bases provide maps with prescribed B3-equivariance properties. Our explicit construction of these bases should be relevant well beyond the scope of this work. The expression of the B3-invariants can then be given in a compact form as the composition of two equivariant maps. Instead of providing (cumbersome) explicit expressions for the O3-invariants, we provide efficient algorithms for their evaluation and rewriting. We also use the constructed B3-invariants to determine the O3-orbit locus and provide an algorithm for the inverse problem of finding an element in $R[x, y, z]_{2d}$ with prescribed values for its invariants. These are the computational issues relevant in brain imaging.

This work has been sumitted and is currently under review. A preprint is available in [44].

7.1.6. Non-parametric graphnet-regularized representation of dMRI in space and time

Participants: Rutger Fick, Alexandra Petiet [ICM, CENIR, Paris], Mathieu Santin [ICM, CENIR, Paris], Anne-Charlotte Philippe [ICM, CENIR, Paris], Stéphane Lehericy [ICM, CENIR, Paris], Demian Wassermann, Rachid Deriche.

Effective representation of the four-dimensional diffusion MRI signal – varying over three-dimensional q-space and diffusion time τ – is a sought-after and still unsolved challenge in diffusion MRI (dMRI). We propose a functional basis approach that is specifically designed to represent the dMRI signal in this qτ-space. Following recent terminology, we refer to our qτ-functional basis as "qτ-dMRI". qτ-dMRI can be seen as a time-dependent realization of q-space imaging by Paul Callaghan and colleagues. We use GraphNet regularization – imposing both signal smoothness and sparsity – to drastically reduce the number of diffusion-weighted images (DWIs) that is needed to represent the dMRI signal in the qτ-space. As the main contribution, qτ-dMRI provides the framework to – without making biophysical assumptions – represent the qτ-space signal and estimate time-dependent q-space indices (qτ-indices), providing a new means for studying diffusion in nervous tissue. We validate our method on both in-silico generated data using Monte–Carlo simulations and an in-vivo test-retest study of two C57Bl6 wild-type mice, where we found good reproducibility of estimated qτ-index values and trends. In the hopes of opening up new τ-dependent venues of studying nervous tissues, qτ-dMRI is the first of its kind in being specifically designed to provide open interpretation of the qτ-diffusion signal.

This work has been published in [17]

7.1.7. Computational diffusion & perfusion MRI in brain imaging

Participants: Marco Pizzolato, Rachid Deriche.

Diffusion and Perfusion Magnetic Resonance Imaging (dMRI & pMRI) represent two modalities that allow sensing important and different but complementary aspects of brain imaging. This work performed within the framework of M. Pizzolato's PhD thesis presents a theoretical and methodological investigation on the MRI modalities based on diffusion-weighted (DW) and dynamic susceptibility contrast (DSC) images. For both modalities, the contributions of the thesis are related to the development of new methods to improve quality, processing, and exploitation of the obtained signals. With respect to contributions in diffusion MRI, the nature of the complex DW signal is investigated to explore a new potential contrast related to tissue microstructure. In addition, the complex signal is exploited to correct a bias induced by acquisition noise of DW images, thus improving the estimation of structural scalar metrics. With respect to contributions in perfusion MRI, the DSC signal processing is revisited in order to account for the bias due to bolus dispersion. This phenomenon prevents the correct estimation of perfusion metrics but, at the same time, can give important insights about the pathological condition of the brain tissue. The contributions of the thesis are presented within a theoretical and methodological framework, validated on both synthetic and real images.

This work has been published in [15].

7.1.8. Solving the Inclination Sign Ambiguity in Three Dimensional Polarized Light Imaging with a PDE-Based Method

Participants: Abib Alimi, Marco Pizzolato, Rutger Fick, Rachid Deriche.

Three dimensional Polarized Light Imaging (3D-PLI) is a contrast-enhancing technique that measures the spatial fiber architecture in the human postmortem brain or heart at a submillimeter resolution. In a voxel, the 3D fiber orientation is defined by the direction angle and the inclination angle whose sign is unknown. To have an accurate explanation of fiber orientation, it is compulsory to clear up this sign ambiguity. A tilting process provides information about the true inclination sign, however the technique is highly sensitive to noise. In this work, a partial differential equations based method is proposed to reduce the noise: the total variation model of Rudin-Osher-Fatemi is extended to 3D orientation vector images to restore the sign. The proposed algorithm is evaluated on synthetic and human heart data and results show that the true sign of the inclination angle can be successfully extracted.

This work has been published in [27]

7.1.9. Brain correlates of apathy in Kleine Levin syndrome: a mean apparent propagator study

Participants: Anne-Charlotte Philippe [ICM, CENIR, Paris], Sophie Lavault [ICM, CENIR, Paris], Romain Valabregue [ICM, CENIR, Paris], Richard Levy [ICM, CENIR, Paris], Isabelle Arnulf [ICM, CENIR, Paris], Stéphane Lehericy [ICM, CENIR, Paris], Rutger Fick, Demian Wassermann, Rachid Deriche.

Kleine-Levin syndrome (KLS) is a rare neurological disorder characterized by episodes of severe hypersomnia, apathy, cognitive impairment, derealization and behavioral disturbances. Between episodes, patients have normal sleep, mood and behavior. Apathy is a prominent clinical feature of KLS but its pathophysiology is not known. Using new techniques to boost signal-to-noise ration and biomarker extraction in multi-shell dMRI [13], we have studied, in collaboration with the Brain and Spine Institute (ICM, Paris) the Klein-Levin syndrome (KLS) [45]. Our results highlight the presence of structural changes correlated to the apathy score i n the anterior portion of the CC during episodes, a region where fibers project onto the medial orbitofrontal cortex. As, these prefrontal regions are involved in motivation processes, this suggests that apathy in KLS could result from difficul ties to provide the affective/motivational value of a given behavioral context.

This work has been published in [45].

7.2. Unveiling brain activity using M/EEG

7.2.1. Dictionary learning for M/EEG processing

Participants: Maureen Clerc, Sebastian Hitziger, Théodore Papadopoulo.

Signals obtained from magneto- or electroencephalography (M/EEG) are very noisy and inherently multi-dimensional, i.e. provide a vector of measurements at each single time instant. To cope with noise, researchers traditionally acquire measurements over multiple repetitions (trials) and average them to classify various patterns of activity. This is not optimal because of trial-to-trial variability (waveform variations, jitters). The jitter-adaptive dictionary learning method (JADL) has been developed [82] to better handle for this variability, with a particular emphasis on jitters. It was generalized to handle variability both in jitter and in duration, in a method called Adaptive Waveform Learning [8]. These methods [83] are data-driven and learn a dictionary (prototype signals) from a set of signals, but are limited to a single channel, which restricts their capacity to work with very noisy multichannel data such as M/EEG. An extension to multidimensional signals has been developped in [96] and [41].

7.2.2. Accounting for conductivity in M/EEG leadfields

Participants: Maureen Clerc, Juliette Leblond [APICS project-team], Kostiantyn Maksymenko, Jean-Paul Marmorat [APICS project-team], Théodore Papadopoulo, Christos Papageorgakis [APICS project-team].

We aim at improving the EEG forward/inverse problem by better modelling the skull conductivity. Indeed, it has been shown that the complex conductivity profile of the skull has a major influence on the accuracy of the EEG forward/inverse problems.

- The skull conductivity is usually considered homogeneous, but the skull is actually made of several types of bone: hard (compacta) and soft (spongiosa) which may have different conductivity characteristics. By adapting a template to MR images of individual subjects, the influence of the spongiosa on source localization can be demonstrated [97]. Estimating the conductivity values of the skull compartments is an important problem, for which theoretical results on uniqueness and robustness have been obtained [64], [63], [26].
- Such studies show the need of easily obtaining EEG leadfields with various conductivity values. Recomputing a new leadfield for every different set of conductivities is expensive. We have thus developped a technique inspired by “reduced bases” which approximates the set of leadfields over a domain of conductivities using a low number of "base leadfields" [40]. The approach offers mathematical guarantees on the approximation level and provides an efficient methodological ground for attempting to compute both sources and conductivities in the EEG inverse problem.

7.2.3. Cochlear implant stimulation models

Participants: Maureen Clerc, Kai Dang, Dan Gnansia [Oticon Medical], Nicolas Guevara [CHU de Nice].

Our expertise on building forward models in bioelectromagnetism has led to a collaboration with Oticon Medical, a cochlear implant manufacturer. Through Dang’s PhD thesis [12], we developed computational models of cochlear implant stimulation, which can account for the anatomical shape of the inner ear, the shape of the implanted electrode, and the stimulation mode, for instance common ground or multi-mode grounding [67], [66]. The OpenMEEG software was extended to cope with zero-conductivity regions (e.g. the silicon electrode holder). The cochlear implant Boundary Element model was coupled with a lumped capacitor and constant phase element model, allowing time-domain simulation. Thorough validation campaigns were conducted, in vitro (notably using a 3D printer) and in situ (in human specimens).

7.3. Combining spatio-temporal CNS imaging modalities

7.3.1. Groupwise structural parcellation of the whole cortex: A logistic random effects model based approach

Participants: Guillermo Gallardo, William Wells [Harvard Medical School, Boston, MA, USA], Demian Wassermann, Rachid Deriche.

Current theories hold that brain function is highly related to long-range physical connections through axonal bundles, namely extrinsic connectivity. However, obtaining a groupwise cortical parcellation based on extrinsic connectivity remains challenging. Current parcellation methods are computationally expensive; need tuning of several parameters or rely on ad-hoc constraints. Furthermore, none of these methods present a model for the cortical extrinsic connectivity of the cortex. To tackle these problems, we propose a parsimonious model for the extrinsic connectivity and an efficient parceling technique based on clustering of tractograms. Our technique allows the creation of single subject and groupwise parcellations of the whole cortex. The parcellations obtained with our technique are in agreement with structural and functional parcellations in the literature. In particular, the motor and sensory cortex are subdivided in agreement with the human homunculus of Penfield. We illustrate this by comparing our resulting parcels with the motor strip mapping included in the Human Connectome Project data.

This work has been published in [6].

7.3.2. Spatial regularization based on dMRI to solve EEG/MEG inverse problem

Participants: Brahim Belaoucha, Théodore Papadopoulo.

In this work, we present a new approach to reconstruct dipole magnitudes of a distributed source model for magnetoencephalographic (MEG) and electroencephalographic (EEG). This approach is based on the structural homogeneity of the cortical regions which are obtained using diffusion MRI (dMRI). First, we parcellate the cortical surface into functional regions using structural information. Then, we use a weighting matrix that relates the dipoles' magnitudes of sources inside these functional regions. The weights are based on the region's structural homogeneity. Results of the simulated and real MEG measurement are presented and compared to classical source reconstruction methods.

This work has been published in [29], [11].

7.3.3. Large brain effective network from EEG/MEG data and dMR information

Participants: Brahim Belaoucha, Théodore Papadopoulo.

In this research, we aim at reconstructing the information flow in the brain for a given task. More than simple activations, we look at their relationship in time, so at networks constituted by nodes obtained from the parcellations of [55] and edges coming from tractographies obtained by dMRI. In [56] a multivariate auto-regressive model has been used to model the interactions between brain areas. Those areas are obtained using the methods depicted in paragraph 1. Then a putative network is built using connexions obtained by tractography augmented by cortico-cortical connexions (horizontal connexions between neighbor areas) which are not seen by dMRI. A two stage algorithm estimates the coefficients of the autoregressive matrices [57]. Those matrices are constrained to be sparse, so that the non-zero coefficients can be used to estimate the effective network that was activated during the task. The method was validated using simulated data and applied to real MEG and EEG datasets.

This work has been published in [28], [11].

7.3.4. Inference and Visualization of Information Flow in the Visual Pathway using dMRI and EEG

Participants: Samuel Deslauriers-Gauthier, Jean-Marc Lina [ETS - Ecole de Technologie Supérieure, Montréal, CA], Russel Buttler [SCIL, Sherbrooke University, CA], Pierre-Michel Bernier [SCIL, Sherbrooke University, CA], Kevin Whittingstall [SCIL, Sherbrooke University, CA], Maxime Descoteaux [SCIL, Sherbrooke University, CA], Rachid Deriche.

We propose a method to visualize information flow in the visual pathway following a visual stimulus. Our method estimates structural connections using diffusion magnetic resonance imaging and functional connections using electroencephalography. First, a Bayesian network which represents the cortical regions of the brain and their connections is built from the structural connections. Next, the functional information is added as evidence into the network and the posterior probability of activation is inferred using a maximum entropy on the mean approach. Finally, projecting these posterior probabilities back onto streamlines generates a visual depiction of pathways used in the network. We first show the effect of noise in a simulated phantom dataset. We then present the results obtained from left and right visual stimuli which show expected information flow traveling from eyes to the lateral geniculate nucleus and to the visual cortex. Information flow visualiza-tion along white matter pathways has potential to explore the brain dynamics in novel ways.

This work has been published in [37].

7.3.5. Information Flow in the White Matter During a Motor Task: A Structural Connectivity Driven Approach

Participants: Guillermo Gallardo, Demian Wassermann, Maxine Descoteaux [SCIL, Sherbrooke University, CA], Samuel Deslauriers-Gauthier, Rachid Deriche.

Cognitive tasks emerge from the interaction of functionally specialized cortical regions . These interactions are supported by information flow through white matter fiber bundles connecting distant cortical regions. Estimating the information flow through white matter fiber bundles would therefore provide valuable information into the necessary cortical interactions to realize a task. In this work, we build a Bayesian network representing cortical regions and their connections using a structural connectivity driven parcellation derived from diffusion MRI (dMRI). We then introduce Magnetoencephalography (MEG) measurements as evidence into this network to infer the information flow between cortical regions. We show, for the first time, results on the interaction between the precentral, postcentral and occipital regions during a hand-movement task.

This work has been published in [39].

7.4. Brain Computer Interfaces

7.4.1. Multimodal BCI

Participants: Maureen Clerc, Lorraine Perronnet [Visages project-team], Saugat Bhattacharyya [Camin team].

We are conducting research in Multimodal BCI:

- In collaboration with Camin team in Montpellier, we are investigating the use of feedback using Functional electrical stimulation (FES) of limb muscles [58] for Motor Imagery and also studying the influence of the FES on brain signals.
- A study comparing unimodal and bimodal EEG-fMRI neurofeedback for 10 healthy volunteers showed that EEG-fMRI leads to stronger activations than EEG alone [24].

7.4.2. Automatizing calibration

Participants: Maureen Clerc, Nathalie Gayraud, Alain Rakotomamonjy [Université de Rouen].

One of the drawbacks of BCI is the time required for setup and calibration before its use. Instead of fine-tuning the BCI by collecting labeled data by asking the user to perform tasks without any purpose nor feedback, we propose to fine-tune the BCI after the user has started using it. This requires an initial - suboptimal - classifier, which we propose to build through "transfer learning" by re-using labeled data acquired from other subjects and other sessions. We have investigated two main directions for this:

- **Riemannian geometry of covariance matrices**. Covariance matrices of EEG signals are interesting features for BCI. Their information geometry has led to impressive transfer learning performance, as testified by their excellent ranking in several competitions. We are studying the advantages of these features and how they can be used to build separability markers within datasets [74].
- **Optimal transport theory**. A new strand of research is to use optimal transport methods for domain adaptation. The idea is to reuse the classifiers built from existing labeled datasets by transporting the new unlabeled data onto the domain of the existing data [34].

7.4.3. Translational research

Participants: Maureen Clerc, Claude Desnuelle [CHU de Nice], Violaine Guy [CHU de Nice], Théodore Papadopoulo, Marie-Hélène Soriani [CHU de Nice].

The P300-speller is a widespread BCI paradigm for communication, studied in many laboratories. Our involvement in this paradigm was triggered by the Nice University Hospital ALS reference center. Having evaluated with them existing P300-spellers, which were found difficult to get to work properly, we decided to develop our own P300-speller based on OpenViBE in collaboration with Inserm Lyon [65], [104]. Among its distinctive features: optimal stopping of flashes, principled choice of letter groups [103] and word completion and prediction. We demonstrated the feasibility of our "Coadapt P300 speller" in collaboration with Nice University Hospital during a clinical study with 20 ALS patients who participated in 3 sessions each [99], [20].

In order to bring this type of communication BCI closer to patients, we developed a user-friendly software, bci-vizapp, with far greater portability with respect to hardware (OS, screen and amplifier).

Our work aroused the interest of patient associations, in particular “Espoir Charcot” who helped a patient hospitalized in Chambéry acquire a consumer-grade (Emotiv-EPOC) EEG in order to use the P300-speller. He eventually succeeded in using the system, with the help of a local engineer, but notably without our physical presence at any stage. This represents an important first step for us in translational research.

BIOVISION Team

7. New Results

7.1. High tech vision aid systems for low-vision patients

7.1.1. Using virtual reality to helping low-vision people read depending on their pathology

Participants: Marco Benzi [Université Côte d'Azur (France)], Stéphanie Baillif [Centre hospitalier Pasteur 2 (service d'ophtalmologie, Nice, France)], Annick Martin ["27Delvalle" (Centre d'Innovation Santé de la ville de Nice, France)], Eric Castet [Aix-Marseille Université (CNRS, Laboratoire de Psychologie Cognitive, Marseille, France)], Fabio Solari [University of Genoa (DIBRIS, Genoa, Italy)], Manuela Chessa [University of Genoa (DIBRIS, Genoa, Italy)].

By stimulating imagination, reading can be considered as the first immersive media that we are experimenting in our life. We read for leisure, to learn or to be informed. Nowadays, we read not only on printed books or newspaper but on a variety of electronic platforms (computers, tablets, phones), thus extending the possibilities to read. However, reading poses problems for almost everyone with low-vision and it is amongst the strongest need reported by patients [55], [65]. Electronic equipments such as CCTV have offered new possibilities for the patients to tune their prefered display and many studies have been done to understand the impact of most parameters in reading performance [50], [70], [42], [49], [65]. However, display is still highly limited by the small field of view offered by CCTVs, the navigation issues, and the fact that they are constrained to sit at their desk in order to read, thus providing a limited comfort to patients. Our goal is to investigate how virtual reality could be used to overcome these limitations and study new reading aid strategies depending on patients' pathologies.

This project received funding from Université Côte d'Azur (France), in the "Pré-maturation" call which finances actions that transform existing proof of concept into an operational laboratory prototype allowing either the realization of "robust" demonstrators or the complete experimental validation of concept (see Sec. 9.1.1).

7.1.2. Real-time image enhancement in virtual reality applications for low-vision people

Participants: Manuela Chessa [University of Genoa (DIBRIS, Genoa, Italy)], Alberto Patino [University of Genoa (DIBRIS, Genoa, Italy)], Horacio Rostro [University of Guanajuato (Guanajuato, Mexico)], Eric Castet [Aix-Marseille Université (CNRS, Laboratoire de Psychologie Cognitive, Marseille, France)], Fabio Solari [University of Genoa (DIBRIS, Genoa, Italy)], Pierre Kornprobst.

In the last years, virtual reality technology has experienced a boost in affordability, and an increasing number of applications have emerged proposing new immersive 360 degrees visual content. To make this content accessible for low-vision people, one should adopt the same strategies as in traditional displays, i.e., use dedicated image enhancement methods to facilitate their interpretation. This work introduces a virtual reality application for mobile devices that implements real-time content enhancement. It is implemented as a visual search task in a set of static 360 degrees environments: the immersed user can manipulate the parameters of the enhancement algorithm in a intuitive way, using an external controller. In particular, we focus on the transform proposed by Peli et al [69], which is based on an adaptive filter that controls the local contrast as a function of the local mean luminance of an image. Such a transform has been shown to improve recognition tasks in patients with moderate visual loss, central scotoma or cataracts. Our application is, to our knowledge, the first attempt to evaluate the impact of this image enhancement in an immersive virtual reality environment. In particular, our system allows the real time tuning of the transform, and provides all the quantitative data to analyse a posteriori users' behaviour and how parameters may impact their performance. Designed as a game, it is perceived as more enjoyable than traditional ophthalmologic experiments. More generally, this application could be a way for low-vision people to adjust vision enhancements to their needs in everyday virtual reality applications, also for entertainment purposes

This work was presented at the Vision conference [27].

7.1.3. ARVIP: Augmented reality for visually impaired people

Participants: Josselin Gautier, Pierre Kornprobst, Frédéric Dosière [Bosch Visiontec (Sophia Antipolis, France)], David Coupé [Bosch Visiontec (Sophia Antipolis, France)].

In Biovision, we want to develop new augmented reality systems for low-vision people, to facilitate scene interpretation by enhancing important scene characteristics. Research and investigations are conducted using automotive industry HW solutions, thanks to a partnership with Bosch Visiontec (Sophia Antipolis, France, see Sec. 8.1.1).

7.2. Human vision understanding through joint experimental and modeling studies, for normal and distrophic vision

7.2.1. Recurrent network dynamics reconciles visual motion segmentation and integration

Participants: N.v. Kartheek Medathati, James Rankin [University of Exeter (Department of Mathematics, Exeter, UK)], Andrew I. Meso [Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France)], Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France)].

In sensory systems, different computational rules are postulated to be implemented by different neuronal subpopulations characterised by their tuning function. For instance, in primate cortical area MT, different classes of direction-selective cells have been identified and related to either motion integration, segmentation or transparency. Still, how such different tuning properties are constructed is unclear. The dominant theoretical viewpoint based on linear-nonlinear feedforward cascade does not account for their complex temporal dynamics and their versatility when facing different input statistics. Here, we demonstrate that a recurrent network model of visual motion processing can reconcile these different properties. Using a ring network, we show how excitatory and inhibitory interactions can implement different computational rules such as vector averaging, winner-take-all or superposition. The model also captures ordered temporal transitions between these behaviours. In particular, depending on the inhibition regime the ring network can switch from motion integration to motion segmentation, thus being able to compute either a single pattern motion or to superpose multiple inputs as in motion transparency. We thus demonstrate that recurrent architectures can adaptively give rise to different cortical computational regimes depending upon the input statistics, thus reconciling the twin blows of sensory processing: integration and segmentation.

This work was published in [20]

7.2.2. Retinal waves

Participants: Dora Karvouniari, Lionel Gil [Institut Non Linéaire de Nice (INLN, Université Côte d'Azur (France), France)], Olivier Marre [Institut de la Vision (Paris, France)], Serge Picaud [Institut de la Vision (Paris, France)], Bruno Cessac.

Retinal waves are bursts of activity occurring spontaneously in the developing retina of vertebrate species, contributing to the shaping of the visual system organization: retina circuitry shaping, retinotopy, eye segregation [77], [56], [72], [57]. They stop a few weeks after birth. Wave activity begins in the early development, long before the retina is responsive to light. It was recently found that they can be reinitiated pharmacologically in the adult mammalian retina [54]. This could have deep consequences on therapy for several degenerative retinal diseases. The mechanism of their generation, in imature, or adult retinas, remains however incompletely understood [78].

We have proposed a model for stage II retinal waves - induced by bursting Starburst Amacrine Cells (SACs) coupled by acetylcholine - with 2 objectives: (i) being sufficiently close to biophysics to explain and propose experiments and (ii) affording a mathematical analysis. From a bifurcations analysis we have highlighted several relevant biophysical parameters controlling waves generation, mainly regulating potassium and calcium dynamics. We thus explain how SACs in different species exhibit a large variability in their bursting periods with a common mechanism. We have proposed a testable experimental prediction providing a possible link of the evolution of voltage-dependent potassium channels along development with their role on the excitability properties of SACs. We have reproduced experimental findings (statistical characteristics of waves size, duration and frequency of appearance) and analyzed how the evolution of cholinergic conductance due to the maturation of nicotinic receptors dramatically changes the retinal wave characteristics. We have also shown that the nonlinear dynamics generates heterogeneous local spatial structures inside which retinal waves propagate. This induces a wide variability in waves characteristics even though the network is perfectly homogeneous.

This work has been presented in [36], [34], [24], [25], [38], [37]

7.2.3. Pan-retinal characterisation of light responses from ganglion cells in the developing mouse retina

Participants: Gerrit Hilgen [Institute of Neuroscience (ION, Newcastle, UK)], Sahar Pirmoradian [Institute for Adaptive and Neural Computation (ANC, School of Informatics University of Edinburgh, UK)], Daniela Pamplona [ENSTA ParisTech, Autonomous Systems and Robotics (Paris, France)], Pierre Kornprobst, Bruno Cessac, Matthias H. Hennig [Institute for Adaptive and Neural Computation (ANC, School of Informatics University of Edinburgh, UK)], Evelyne Sernagor [Institute of Neuroscience (ION, Newcastle, UK)].

We have investigated the ontogeny of light-driven responses in mouse retinal ganglion cells (RGCs). Using a large-scale, high-density multielectrode array, we recorded from hundreds to thousands of RGCs simultaneously at pan-retinal level, including dorsal and ventral locations. Responses to different contrasts not only revealed a complex developmental profile for ON, OFF and ON-OFF RGC types, but also unveiled differences between dorsal and ventral RGCs. At eye-opening, dorsal RGCs of all types were more responsive to light, perhaps indicating an environmental priority to nest viewing for pre-weaning pups. The developmental profile of ON and OFF RGCs exhibited antagonistic behaviour, with the strongest ON responses shortly after eye-opening, followed by an increase in the strength of OFF responses later on. Further, we found that with maturation receptive field (RF) center sizes decrease, responses to light get stronger, and centers become more circular while seeing differences in all of them between RGC types. These findings show that retinal functionality is not spatially homogeneous, likely reflecting ecological requirements that favour the early development of dorsal retina, and reflecting different roles in vision in the mature animal.

This work has been published in [19].

7.2.4. Trajectory anticipation, from retina to V1

Participants: Selma Souihel, Bruno Cessac.

Global motion processing is a major computational task of biological visual systems. When an object moves across the visual field, the sequence of visited positions is strongly correlated in space and time, forming a trajectory. These correlated images generate a sequence of local activation of the feedforward stream. At the present stage of knowledge, it is still unclear how the early visual system processes motion trajectories. Motion integration, anticipation and prediction would be jointly achieved through the interactions between feed-forward, lateral and feedback propagations within a common spatial reference frame, the retinotopic maps. Addressing this problem is particularly challenging, as it requires to probe these sequences of events at multiple scales (from individual cells to large networks) and multiple stages (retina, primary visual cortex (V1).

In the context of the ANR Trajectory we are working on such an integrated approach. We aim at modelling the population responses at two key stages of visual motion encoding: the retina and V1 based on simultaneous micro- and mesoscopic recordings made by our partners Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France) and Institut de la Vision (Paris, France), and design a simulator of retinal output feeding V1. This study is a step toward understanding mechanisms of motion coding and anticipation with strong impact on our understanding of the visual system.

We have implemented in our retina simulator, PRANAS, gain control mechanisms allowing to reproduce motion anticipation for simple motions. We developed a simple decoding algorithm that reconstructs the stimulus using firing rates, with the goal of comparing the performance of the different models of gain control. We have also designed a biologically inspired model of connectivity, mimicking short and long range connections between ganglion cells via amacrine cells. This has allowed us to compare the pairwise correlations between ganglion cells, under the influence of a moving object both, in vivo and in silico. These results have been presented in [40], [41], [39]

7.2.5. Dimensionality reduction in spatio-temporal MaxEnt models and analysis of retinal ganglion cell spiking activity in experiments

Participants: Rubén Herzog [Centro Interdisciplinario de Neurociencia de Valparaíso (CINV, Valparaíso, Chile)], Maria-Jose Escobar [Universidad Tecnico Federico Santa María (Electronics Engineering Department, Valparaíso, Chile)], Adrian Palacios [Centro Interdisciplinario de Neurociencia de Valparaíso (CINV, Valparaíso, Chile)], Bruno Cessac.

Retinal spike response to stimuli is constrained, on one hand by short range correlations (receptive field overlap) and on the other hand by lateral connectivity (cells connectivity). This last effect is difficult to handle from statistics because it requires to consider spatio-temporal correlations with a time delay long enough to take into account the time of propagation along synapses. Although MaxEnt models are useful to fit optimal model (maximizing entropy) under the constraints of reproducing observed correlations, they do address spatio-temporal correlations in their classical form (Ising or higher order interactions but without time delay). Binning in such models somewhat integrates propagation effects, but in an implicit form, and increasing binning severely bias data. To resolve this issue we have considered spatio-temporal MaxEnt model formerly developed e.g. by Vasquez et al. [75]. The price to pay, however is a huge set of parameters that must be fitted to experimental data to explain the observed spiking patterns statistics. There is no a priori knowledge of which parameters are relevant and which ones are contributing to overfitting. We propose here a method of dimension reduction, i.e. a projection on a relevant subset of parameters, relying on the so-called Susceptibility matrix closely related to the Fisher information. In contrast to standard methods in information geometry though, this matrix handles space and time correlations. We have applied this method for retina data obtained in a diurnal rodent (Octodon degus, having 30% of cones photoreceptors) and a 252-MEA system. Three types of stimuli were used: spatio-temporal uniform light, white noise and a natural movie. We show the role played by time-delayed pairwise interactions in the neural response to stimuli both for close and distant cells. Our conclusion is that, to explain the population spiking statistics we need both short-distance interactions as well as long-distance interactions, meaning that the relevant functional correlations are mediated not only by common input (i.e. receptive field overlap, electrical coupling; spillover) but also by long range connections.

This work has been submitted to Plos Comp Bio.

7.2.6. On the mathematical consequences of binning spike trains

Participants: Bruno Cessac, Arnaud Le Ny [Laboratoire d'Analyse et de Mathématiques Appliquées (LAMA, (Université Paris-Est, France)], Eva Loecherbach [Laboratoire d'Analyse, Géométrie et Modélisation (AGM) and Département de Mathématiques (Cergy-Pontoise, France)].

We initiate a mathematical analysis of hidden effects induced by binning spike trains of neurons. Assuming that the original spike train has been generated by a discrete Markov process, we show that binning generates a stochastic process which is not Markovian any more, but is instead a Variable Length Markov Chain (VLMC) with unbounded memory. We also show that the law of the binned raster is a Gibbs measure in the DLR

(Dobrushin-Lanford-Ruelle) sense coined in mathematical statistical mechanics. This allows the derivation of several important consequences on statistical properties of binned spike trains. In particular, we introduce the DLR framework as a natural setting to mathematically formalize anticipation, i.e. to tell "how good" our nervous system is at making predictions. In a probabilistic sense, this corresponds to condition a process by its future and we discuss how binning may affect our conclusions on this ability. We finally comment what could be the consequences of binning in the detection of spurious phase transitions or in the detection of wrong evidences of criticality.

This work has been published in [17].

7.2.7. Linear response of general observables in spiking neuronal network models

Participants: Bruno Cessac, Rodrigo Cofré [Université de Genève (Switzerland) and Centro Interdisciplinario de Neurociencia de Valparaíso (CINV, Valparaíso, Chile)].

The activity of a neuronal network, characterized by action potentials (spikes), is constrained by the intrinsic properties of neurons and their interactions. When a neuronal network is submitted to external stimuli, the statistics of spikes changes, and it is difficult to disentangle the influence of the stimuli from the intrinsic dynamics. Using the formalism of Gibbs distributions, which are a generalization of Maximum Entropy distributions to non-stationary distributions, and generalization of Markov chains to infinite memory, we analyze this problem in a specific model (Conductance-based Integrate-and-Fire), where the neuronal dynamics depends on the history of spikes of the network. We derive a linear response formula allowing to quantify the influence of a weak amplitude external stimuli on the average value of arbitrary observables. This formula clearly disentangles the effect of the stimuli, intrinsic neuronal dynamics, and network connectivity. Upon some approximations, it reduces to a convolution, allowing to recover a standard formulation in computational neuroscience.

This work has been submitted to Journal of Mathematical Neurosciences [33].

7.2.8. A bio-inspired synergistic virtual retina model for tone mapping

Participants: Marco Benzi, Maria-Jose Escobar [Universidad Tecnico Federico Santa María (Electronics Engineering Department, Valparaíso, Chile)], Pierre Kornprobst.

Real-world radiance values span several orders of magnitudes which have to be processed by artificial systems in order to capture visual scenes with a high visual sensitivity. Interestingly, it has been found that similar processing happens in biological systems, starting at the retina level. So our motivation in this paper is to develop a new video tone mapping operator (TMO) based on a synergistic model of the retina. We start from the so-called Virtual Retina model [76], which has been developed in computational neuroscience. We show how to enrich this model with new features to use it as a TMO, such as color management, luminance adaptation at photoreceptor level and readout from a heterogeneous population activity. Our method works for video but can also be applied to static images (by repeating images in time). It has been carefully evaluated on standard benchmarks in the static case, giving comparable results to the state-of-the-art using default parameters, while offering user control for finer tuning. Results on HDR videos are also promising, specifically w.r.t. temporal luminance coherency. As a whole, this paper shows a promising way to address computational photography challenges by exploiting the current research in neuroscience about retina processing.

This work was published in [15].

CAMIN Team

6. New Results

6.1. Modeling and identification of the sensory-motor system

6.1.1. Inertial Sensor based Analysis of Gait for Post-stroke individuals

Participants: Christine Azevedo Coste, Benoît Sijobert, Jérôme Froger [CHU Nîmes], François Feuvrier [CHU Nîmes].

Walking impairment after stroke can be addressed through the use of drop foot stimulators (DFS). In these systems, electrical stimulation is applied to activate the common peroneal nerve and elicit ankle dorsiflexion during the swing phase of gait. DFS are generally piloted by a heel switch positioned in the shoe of the affected side with stimulation being triggered ON by heel rise of the affected foot and triggered OFF by heel strike.

Using inertial sensors for modulating FES intensity could provide a more optimized delivery of stimulation and could also enable to regulate dorsiflexion in the presence of disturbances, such as fatigue or stairs. It could also increase the number of potential users of the technology, allowing subjects walking without heel strikes to be stimulated at a correct timing. Meanwhile, pathological post-stroke gait requires the investigation of complex inertial sensors based algorithms for being able to compute different useful gait parameters for later triggering stimulation. Numerous constraints related to these clinical context, pathology and usability have to be taken into account for providing a reliable patient oriented solution. In this work, we aim to compare accuracy and feasibility of using a minimum amount of inertial sensors instead of the gold standard camera based motion capture, for assessing joint angles and gait events such as stride length or dorsiflexion speed at heel on. 29 subjects were included in this experimental protocol. Equipped with motion capture targets on which an inertial sensor is set, subjects had to perform an experimental path on a gait carpet. EMG recordings were also performed to monitor and evaluate fatigue. Algorithms were developped for computing 3D trajectory (6), dorsiflexion angles at mid-swing or before heel strike. Results shows an RMS error of 5.8° at heel on and 6.6° at mid-swing compared to motion capture data [20]. François Feuvrier has defended his medicine thesis on this topic on December 14th 2017.

6.2. Model based optimal multipolar stimulation without a priori knowledge of nerve structure: application to vagus nerve stimulation

Participants: Mélissa Dali, Olivier Rossel, David Guiraud.

Neural electrical stimulation, applied to the peripheral nervous system for motor functions restoration or neuromodulation, is a thriving technology, especially implanted stimulation using cuff electrodes positioned around a peripheral nerve. The main obstacle to the development of stimulation systems is the difficulty in obtaining the independent stimulation or inhibition of specific target functions (i.e. functional selectivity). The parameters involved in selectivity are not always intuitive and the number of degrees of freedom (choice of electrode, number of contacts, pulse shape etc.) is substantial. Thus, testing all these hypotheses in a clinical context is not conceivable. This choice of parameters can be guided using prior numerical simulations predicting the effect of electrical stimulation on the neural tissue. Numerous studies developed new strategies to achieve selectivity based on modeling results that have been validated a posteriori by experimental works. We presented a general method based on a spatiotemporal model to optimize and assess multipolar neural electrical stimulation without a priori knowledge of the nerve structure. The model consists of two independent components: a lead field matrix (LFM) and an activation model. It represents the transfer function from the applied current to the extracellular voltage present on the nodes of Ranvier along each axon. The determination of fibers activation is used to optimize the spatial layout for the selective activation of specific fibers or nerve areas. Optimization is not only based on selectivity but also on robustness and efficiency of the stimulation settings.

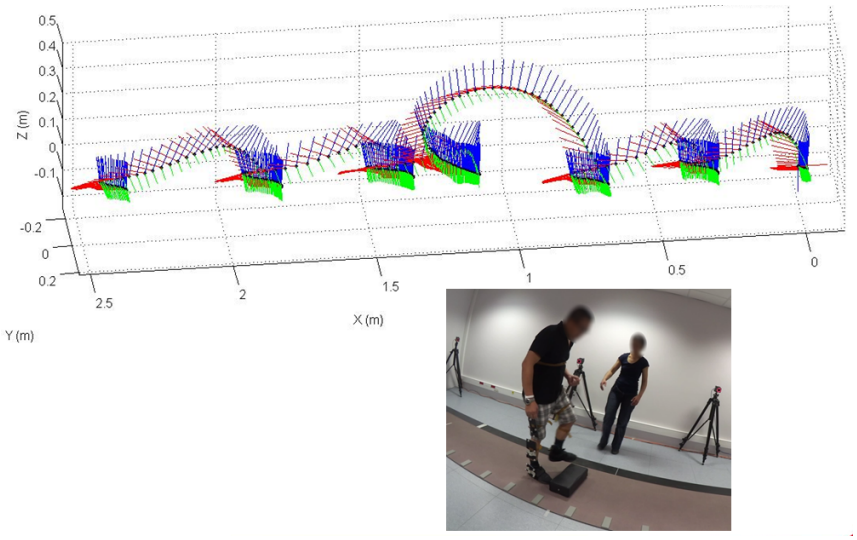


Figure 6. 3D trajectory reconstruction from inertial sensors

The results show that state-of-the- art solutions are part of the optimized solutions but new ones can emerge depending on the trade-off between the criteria and the targeted area. We successfully assessed the solutions in-vivo to selectively induce a decrease in cardiac rhythm through vagus nerve stimulation. Experiments on animal model allowed us to evaluate the effectiveness and genericness of the method. These encouraging results suggest that this approach will have broader applications that would benefit from multicontact cuff electrodes to elicit very accurate and selective responses. This work was performed as part of a larger project on vagus nerve stimulation (INTENSE project) in which one of the applications focused on the treatment of cardiac disorders. The main objective was to selectively activate a specific population of nerve fibers to improve therapy and decrease side effects. Within the framework of the INTENSE project, the second application investigated vagus nerve stimulation as a therapy for morbid obesity. Activation of target axons related to gastric functions requires a significant amount of charge injection. Several studies suggest that non-rectangular waveforms can activate axons of the peripheral nervous system with a reduced amount of charge compared to the reference rectangular pulse shape. Our last contribution focuses on the experimental study and the modeling of these complex waveforms. The modeling approach, if performed properly and while bearing in mind its limits, provides a relevant and even indispensable analysis tool for the clinical adjustment of neuroprostheses.

6.3. Alterations of EEG rhythms and dynamics during motor preparation following wide-awake brain surgery

Participants: Anthony Boyer, Sofiane Ramdani [LIRMM], Hugues Duffau [CHU Montpellier], Bénédicte Poulin-Charronnat [Université de Bourgogne], David Guiraud, François Bonnetblanc.

Awake brain surgery of tumour is used to optimize the resection of tumoral tissue. Postoperatively, patients show mild and temporary neurological deficits despite massive cerebral resections. Reasons for these impairments along with the compensation mechanisms operating within the cortex and subcortical structures are barely understood. The objective of this project is to reveal the remote effects of the tumour and its resection, to determine their nature measuring changes induced in functional Magnetic Resonance Imagery (fMRI) and electroencephalographic signals using standard and nonlinear methods. Recently, we focused on postoperative

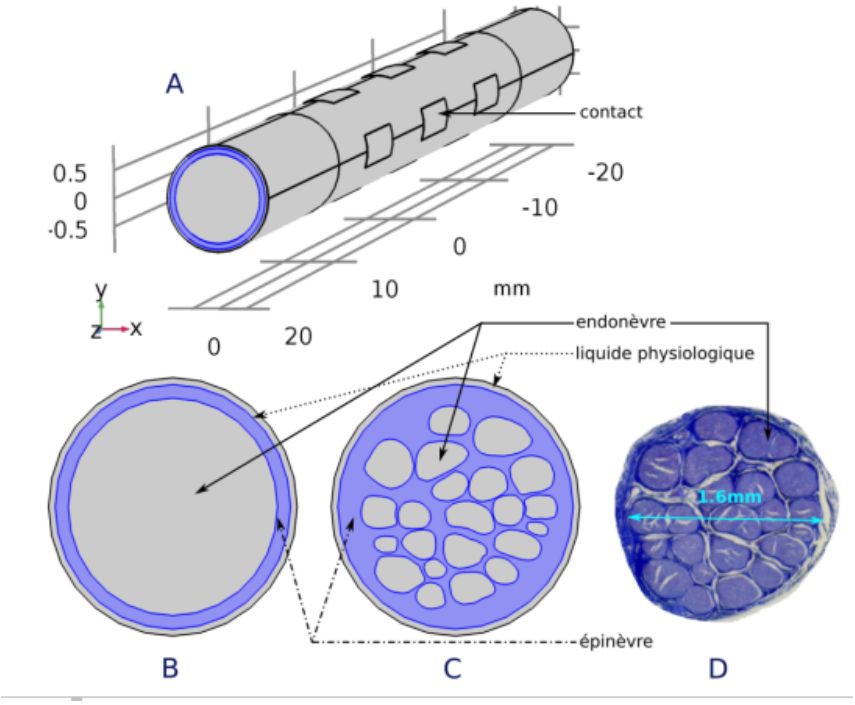


Figure 7. Generic nerve model (A,B) and realistic nerve model (C) based on histological data

brain dynamics of patients, who underwent wide-awake surgery for Low grade gliomas (LGG). We analysed EEG data of 5 patients, who perfomed an ecological visuomanual task, comparing them to a control group of 8 healthy subjects (fig. 8). We used the motor preparation period to extract power features and phase space features to better characterise changes in EEG signal following surgery and subsequent functional reorganisation. The preparation period was chosen for its stationarity allowing analyses, which were not applicable during the ERP period. Our results clearly identify changes in postoperative brain dynamics of patients, who underwent wide-awake surgery. Both spectral and recurrence quantification analyses suggested imbalances between the injured and healthy hemispheres for patients, whether in terms of spectral power density or temporal structure of EEG signal. These investigations performed on the motor preparation period also provided important information regarding longitudinal recovery of brain dynamics. Although all patients in our study had very different tumours, both in size and location, it is interesting to note that the 2 patients, who underwent the experimental protocol respectively 9 and 12 months after surgery, showed more moderate alterations of spectral content and signal complexity independently of the lesion size. This may be seen as an indicator for EEG signal standardization in time and presumably a resumption of brain dynamics. These findings have potential clinical rehabilitation implications.

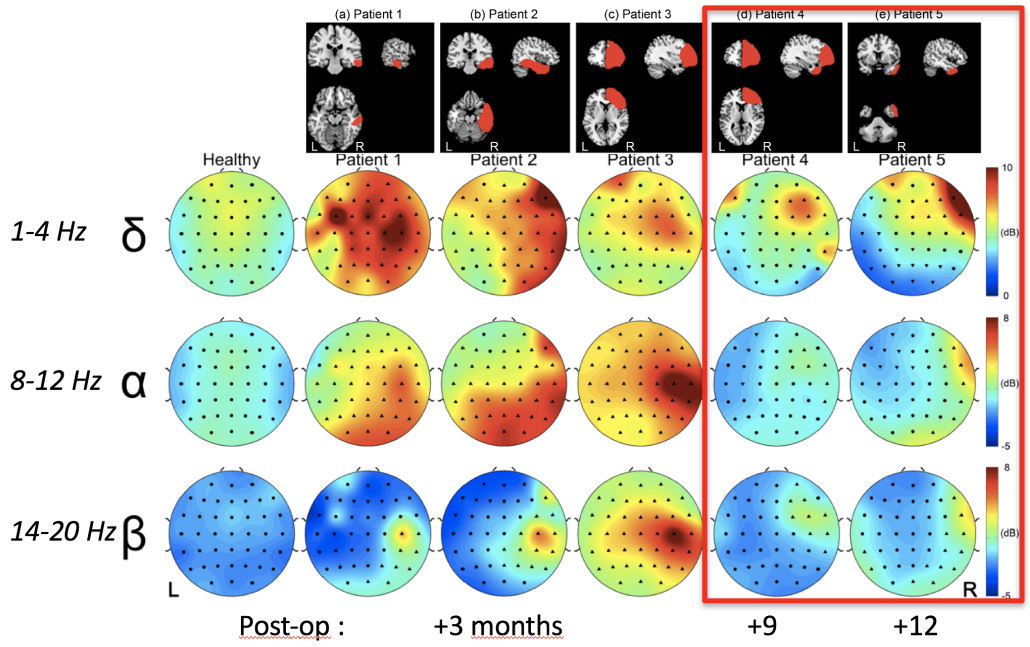


Figure 8. Topographical maps of spectral power: Maps illustrating the spatial distribution of the spectral power contained in a given frequency band over the scalp. We selected bands which showed significant peaks of spectral power for patients in comparison with control group (i.e. δ band: 1 - 4Hz, α/μ band: 9 - 12Hz and β band: 14 - 20Hz) and we then calculated the average power at each electrode for each band. Electrode locations are represented by a set of symbols: ● symbol is used when the corresponding spectral power falls within the 95% confidence interval estimated from healthy subjects. ▲ and ▼ symbols are respectively used when the power is either greater or smaller than the 95% confidence interval. α/μ band

6.4. Electrophysiological brain mapping: measuring evoked potentials induced by electrical stimulation and its physiological spreading in the human brain.

Participants: Marion Vincent, François Bonnetblanc, David Guiraud, Hugues Duffau [CHU Montpellier], Emmanuel Mandonnet [CHU Lariboisière, APHP], Anthony Boyer.

Being able to change or inhibit the activity of a region or population of neurons in the brain is an essential approach in fundamental neuroscience, as it helps the researcher to determine the functional role of neurons. This approach is also important at a more applied level, for brain function mapping during neurosurgical procedures. It is well known that electrical stimulation (ES) affects neural activity by modifying the voltage gradient along the neuronal cell inducing depolarization or hyperpolarization of the membrane. When a current flows in tissues around neuronal cells, it can change their membrane potential and trigger an action potential. However, this general principle can be applied in vivo via several different settings and much is unknown about which neural elements are excited or inhibited locally and how this local perturbation spreads within the brain through physiological pathways [35]. We are now able to record different types of electrophysiological potentials that are evoked by ES in the human brain and we developed some basic methodological considerations required for their correct assessment [26] (fig.9). With our methodology, three different types of evoked potentials can now be measured during brain surgery in the operative room: – Cortical evoked-potential (also called direct cortical response, DCR), when recording the cortex at the stimulation site, – Cortico-axono-cortical evoked-potential, i.e. recording the cortex at a distant site from the stimulating site. These potentials are elicited by physiological propagation through white matter associative pathways from the locally stimulated area towards the distal area, – Axono-cortical evoked potentials, when the cortex is distally recorded from a stimulation site within the white matter. These evoked potentials are technically difficult to observe. Their recording imposes important methodological considerations about the way they can be triggered and measured. In particular, proposed some factors potentially determining the generation of true cortico-axono-cortical evoked potentials, spreading from one stimulated cortical area to another distant one and passing through the white matter pathways. Correctly measuring evoked potentials in the human brain induced by electrical stimulation is important in the clinical domain especially in the neurosurgical context. It remains challenging because of many pitfalls that can occur at the methodological level and few teams in the world are currently able to efficiently record these evoked potentials. Nevertheless, they can give strong real-time in vivo insights into the functional state and connectivity of a patient's brain. In the next years measuring intraoperatively the evoked potentials with ES in the brain will be a new method for mapping the brain in vivo and in real time and taking into account the specificity of each patient's brain.

6.5. Diagnosis evaluation of acute ischemic stroke using new technics

Participants: Victor Vagné, Olivier Rossel, Emmanuelle Le Bars, Stéphane Perrey, Vincent Costalat, David Guiraud.

Cerebral infarctions can now be treated with new techniques using intravenous thrombolysis and thrombectomy. Their proven efficacy is directly correlated to the time lapse between the start of symptoms and the initiation of treatment. Currently, a definitive diagnosis can only be made once the patient has performed a radiological imaging (CT scan or MRI) on a medical center equipped with these expensive devices, thus enabling the medical team to initiate the appropriate treatment. Transit times during the pre-hospitalization phase before diagnosis are therefore often longer and have the greatest negative impact on the patient's prognosis. In collaboration with the interventional neuroradiology department of Gui de Chauliac Hospital, I2FH and Euromov, the EleVANT project is aiming to prospectively evaluate new techniques to assess a diagnosis of acute cerebral ischaemia. This low cost technology could be used in a mobile way for the very early diagnosis of cerebral infarction and thus reduce treatment delays, opening the way to a new generation of diagnostic tools. The concept consist on evaluating the cerebral near-infrared spectroscopy (NIRS) response to different stimulus, and to evaluate its lateralization. Recently, we tested our device on healthy volunteers. Method: Left

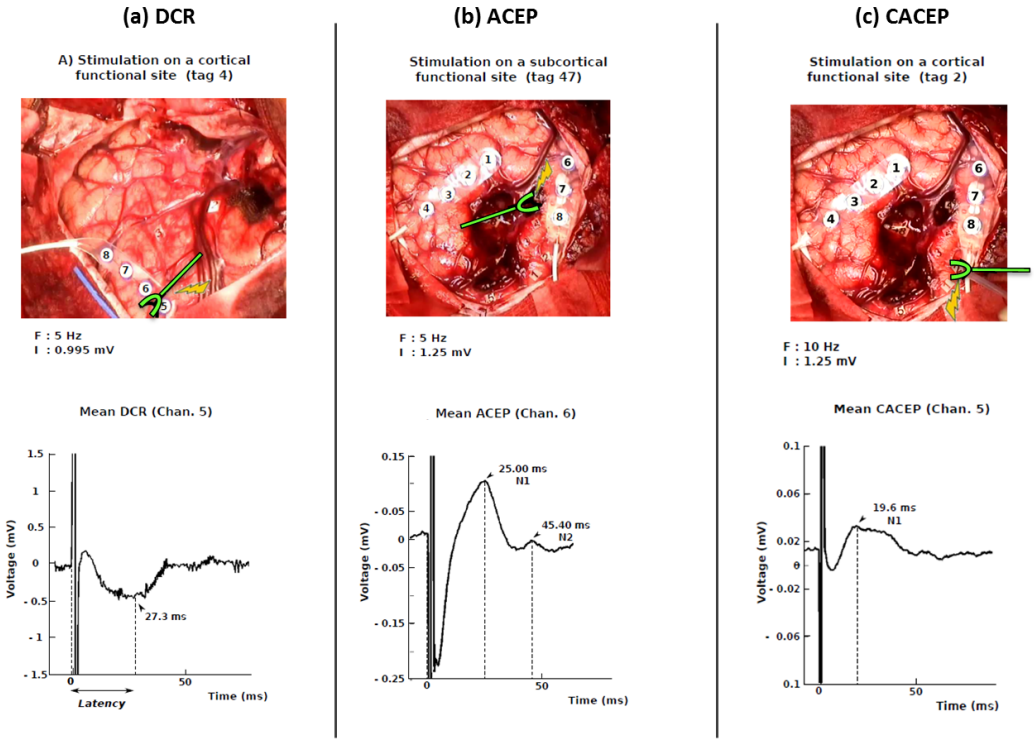


Figure 9. Example of mean evoked potentials induced by DES. (a) Direct cortical responses (DCR) present one primary peak around 27 ms after the stimulation onset. They were measured on 4 patients on cortical sites distant of less than 2 cm from the cortical stimulation site. (b) Axono-cortical evoked potentials (ACEP) presented a primary peak around 25 ms after DES, followed by a second peak 20 ms later. DES is applied subcortically, around 1.7 cm away from the recording site. ACEP were osberved on 2 patients. (c) Cortico-axono-cortical evoked potentials (CACEP) were oberved in 1 patient. Cortical DES induced a one-peak waveform 14 to 35 ms after the artefact onset. CACEP were recorded on cortical site more than 2 cm away from the stimulation site.

and right hemisphere reactivity index are recorded by NIRS and normalized (Figure 10). Result: The experiment present a suitable feasibility and repeatability. In healthy subjects, a good response to the stimulus is recorded, and no significant differences between hemispheres are observed. The confidence level is acceptable since the amplitude response in above the standard deviation level.

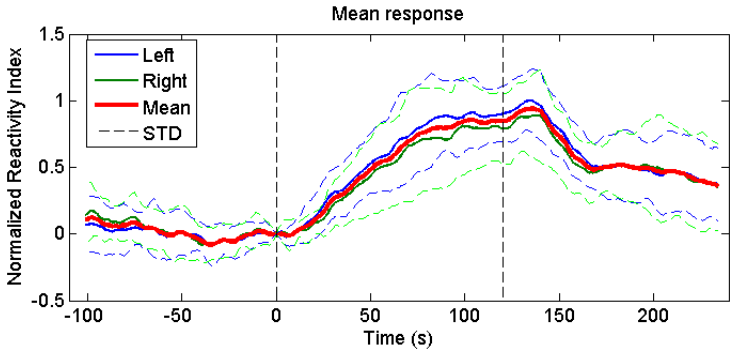


Figure 10. NIRS reactivity Index in response to a stimulus (bounded by the dashed lines)

Discussion: The approach reveal interesting results on healthy subject group. We expect a discriminant difference between hemispheric signals in acute cerebral ischemia.

6.6. A study on Natural user feedback in BCI with peripheral sensory stimulation toward efficient motor learning

Participants: Saugat Bhattacharyya, Maureen Clerc, Mitsuhiro Hayashibe.

Brain-computer Interfacing (BCI) measures the neural activity of the brain to create a direct communication channel to peripheral devices in form of robots, prosthesis, wheelchair or a computer controlled by the user, independent of the peripheral nerves and muscles. To improve the performance of the BCI operation and provide a feedback to the user as an indication of his/her achievement in terms of voluntary brain modulation, a feedback in form of visual, auditory or (vibro-)tactile medium can be provided to the user during training.

One can employ Functional Electrical Stimulation (FES) targeting specific muscle group as a feedback modality in BCI research. Functional Electrical Stimulation (FES) is often applied during rehabilitation to directly engage muscles of the affected side of the body. FES is capable of reconstructing certain daily life skills for physically challenged patients by directly stimulating the targeted muscles group. Thus, it is quite natural to combine FES rehabilitation with BCI systems, where the FES can activate the sensory channel to provide a maximal inflow in the brain and the BCI would provide an efferent outflow of motor commands to close the motor loop. Thus, we aim at studying the effect of electrical stimulation (ES) on the motor imagery EEG and to implement the usage of ES as a natural feedback to BCI. The purpose was to extract all relevant information from the current EEG dataset acquired during BCI experiment with FES based neuro-feedback and to compare the results to the classical visual neuro-feedback paradigm.

The EEG data in this study were recorded from 14 right-handed participants (11 male and 3 female) with a mean age of 28 years and standard deviation of 9 years. The experiments took place at Inria Montpellier and Inria Sophia Antipolis centers. In this experiment, we abide by the norms of the local Inria ethical committee. The participants sat in front of a display placed at eye level and performed the following cued motor imagery tasks: left hand movement, right hand movement, left foot movement and right foot movement. The participants were randomly divided into two groups: one group was provided with only visual feedback (VIS) and the other group was relayed with only FES as feedback during the motor imagery tasks. This step

was taken in the experimentation so that the groups were not influenced by both feebacks and were trained on only one feedback. The VIS feedback group received the feedback in form of a uni-directional bar whereas the FES feedback group received the feedback in form of electrical stimulation on their respective limbs. The group with FES feedback performed two different sets of experiment, which are: 1) the participant performed the motor imagery tasks while receiving electrical stimulation as feedback, which we term as *FES-Active (ACT)* sessions, and 2) the participant performed no motor imagery tasks (relaxation condition) and received the electrical stimulation as a stimuli, which we term as *FES-Passive (PAS)* sessions.

The raw EEG data was first filtered using a notch filter to remove the 50Hz noise from the signal. Then, a 4th order Butterworth filter was applied to the signal. Then, the mean of the signal was removed followed by a spatial filtering using common average referencing technique. Finally, the continuous EEG data were segmented into smaller samples (Epochs) from -1s to 4s, where 0 indicates the onset of the motor imagery task (left hand, right hand, right foot). After filtering and epoching the EEG data, the EEG epochs are spatially filtered using common spatial patterns (CSP) algorithm. Then, the log-variance of 4 discriminating CSP filters (2 for either classes) were selected as features. These features are then used as inputs to a linear discriminating analysis (LDA) classifier to derive the output of the current motor task imagined by the participant. We provide the average classification result across all subjects for three (ACT and VIS) and two sessions (PAS) in Fig.11 . As noted from the figure, the performance of the decoder improves after every session for the VIS and ACT sessions. This can be attributed to the increase in learning occurring to the participant during the progression of the experiment. As the subject was performing no mental tasks during the PAS session, thus, no such improvement on the classification performance is noticed. On the contrary, a decrease in the performance is noted after the second session.

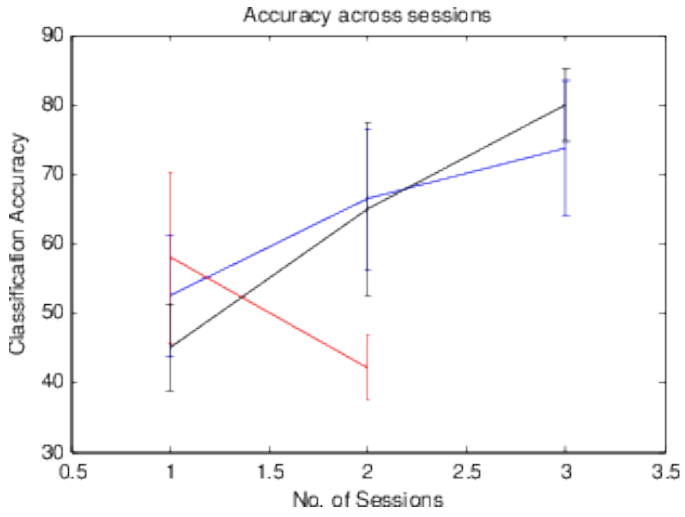


Figure 11. Classification result across all subjects for subsequent sessions.

To further investigate the learning occurring across trials, we have calculated the band power of the epochs at the mu-band (8-12Hz) and the central-beta band (16-25 Hz) at Cz electrode (Fig.12). To calculate the power we have employed Welchs' periodogram at an overlap of 75% and a window size of 250ms. Finally, the average is calculated over all windows of the given trial to determine its power. As seen from the example in Fig.12 , ACT and VIS conditions show a monotonic increase of power which can be quantified as an indication of learning occurring in the participant across the trial.

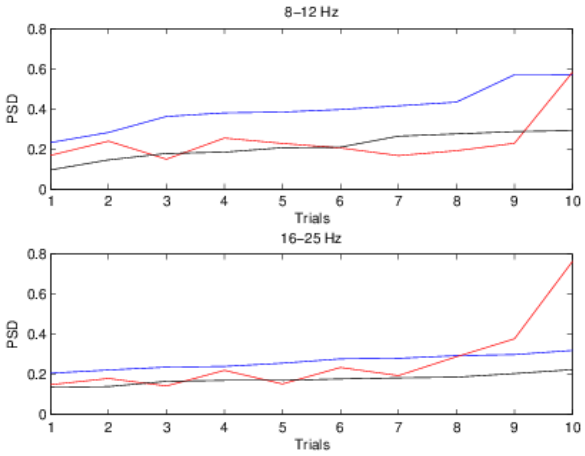


Figure 12. The distribution of power from the first to the tenth correctly classified trial during left hand imagery. Blue is for ACT condition, Red is for PAS and Black is for VIS condition.

6.7. Formal validation for critical digital embedded systems

Participants: Ibrahim Merzoug, Karen Godary-Dejean, David Andreu.

The works addressed here fall under the domain of formal modelling, semantics and verification methods (model checking). We focus on the analysis part of the HILECOP methodology, integrating the specific execution constraints (non-functional properties) into the validation process to guarantee the validation results. Indeed, the state space that is analyzed is that of the model of the system (based on Interpreted Time Petri Nets). It is clear that, if we want to obtain confident validation results, this analyzed state space must include all the possible behaviors of the real system (i.e., considering the execution of the model on the target).

One solution has been studied in the PhD thesis of H. Leroux [34], which lays the foundations of translation rules from the designed model to the analyzed model integrating part of implementation and execution characteristics. These transformations rules allow analyzing the resulting model with classical Petri nets analysis tools (as the Tina toolbox), and to guarantee the inclusion of the real states and traces into the analyzed state space.

However, if the formal model, the Interpreted Time Petri Net in this case (ITPN), is inherently asynchronous, it is nevertheless executed synchronously on the target. In fact, the usual analysis approaches are not adapted in the sense that they construct state graphs that do not conform to the real state evolution within the target. In order to gain confidence in the validity of the results of the formal analysis, we carried on, through the PhD thesis of I. Merzoug, capturing the so-called non-functional characteristics to reify them on the model and finally to consider their impact through a dedicated analysis approach. In other words, we improved the expressiveness of the model and the relevance of the analysis, considering aspects such as clock synchronization, effective parallelism, the risk of blocking induced by the expression of an event (condition) and a time window of occurrence, without omitting the management of exceptions.

To deal with all these aspects, we have proposed a new method of analysis for Synchronously executed ITPN (SITPN), transforming them into an equivalent formalism that could be analyzed ([29]). This formalism is associated with a new formal semantics integrating all the particular aspects of the execution. We also propose and implement a dedicated state space construction algorithm: the Synchronous Behavior Graph (an example

being given on Fig. 13 and Fig. 14). Our work has been applied to an industrial case, more precisely to the validation of the behavior of the digital part of our neuro-stimulator.

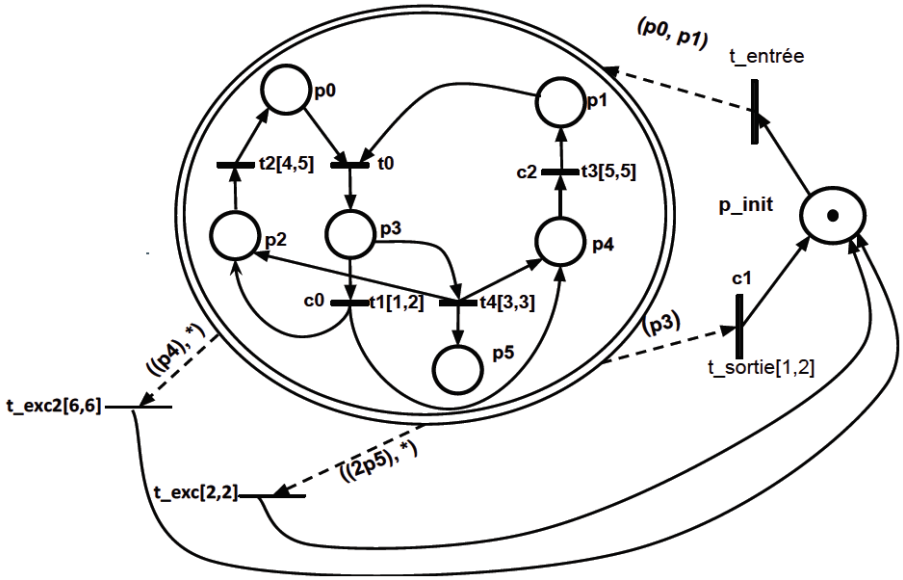


Figure 13. Example of SITPN model with state aggregation and exception handling

6.8. Respiratory detection and monitoring

Participants: Xinyue Lu, Christine Azevedo Coste, David Guiraud, David Andreu, Serge Renaux [Neuroresp], Thomas Similowski [Groupe Hospitalier Pitié-Salpêtrière].

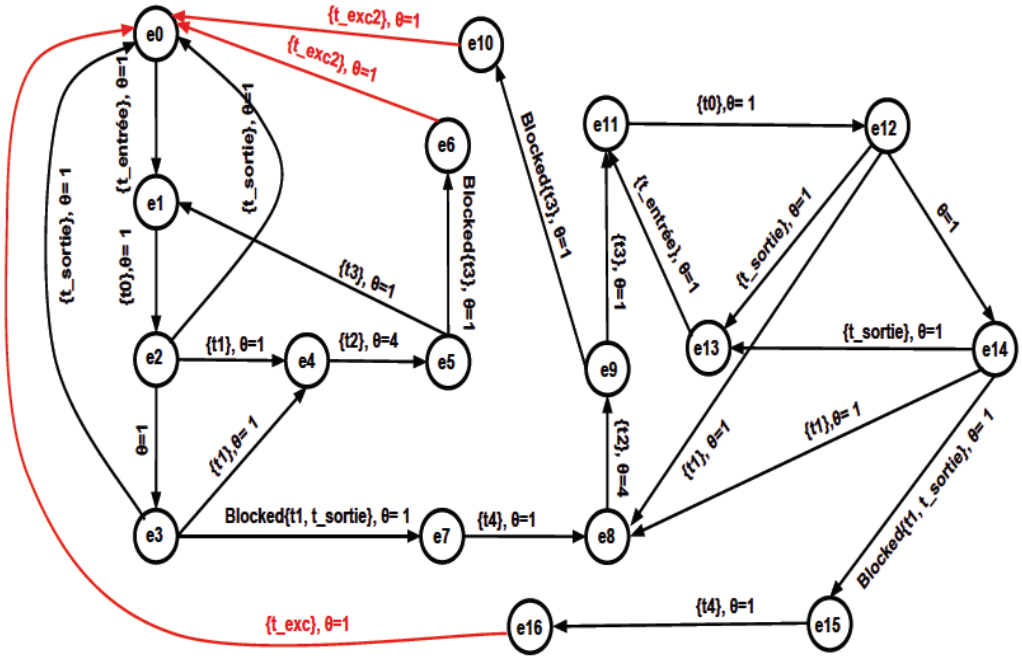
This work is conducted within a CIFRE PhD thesis. The general subject is the respiration induced by implanted stimulation for the tetraplegic and syndrome of Ondine. In France, every year, there is approximately 90 new spinal cord injuries who have a ventilatory dependence due to a high cervical involvement. The prevalence of syndrome of Ondine (central sleep apnea) would be 25.5 per million inhabitants. Because of many disadvantages of mechanical ventilation, the technique of implanted electrical stimulation to restore the respiratory function of the patients can be proposed. But existing systems are based on open-loop controllers, i.e. the phrenic nerve is stimulated with the same intensity, at the same frequency for the whole time, even when the patients can breathe spontaneously. The principle aim of the work is to develop a respiratory detection/monitoring module in this context.

We have developped a first solution bases on a microphone. The signal is processed in order to determine the spectral power between 400Hz and 600Hz (the band of respiratory sounds) and a threshold detection applied to detect respiration.

Preliminary recordings on healthy individuals have been perfomed. Advance signal processing techniques are now under study.

6.9. Evoked EMG correlation with muscle torque

Participants: Adriana Mendes, Mitsuhiro Hayashibe, David Guiraud.



e0: p_init, **B**{}, {t_entrée[1,1]}	**e3:** p3, **B**{}, {t1[1,1], t4[2,2], t_sortie[1,1]}	**e6:** p0,p4, **B**{t3}, {t_exc2[1,1]}	**e9:** p0,p4,p5, **B**{}, {t3[1,1], t_exc2[2,2]}	**e12:** p3,p5,**B**{}, {t1[1,2], t4[3,3], t_sortie[1,2]}	**p15:** p3, p5 **B**{t1, t_sortie}, {t4[1,1]}
e1: p0,p1,**B**{}, {t0[1,1]}	**e4:** p2,p4, **B**{}, {t2[4,4], t3[5,5], t_exc2[6,6]}	**p7:** p3, **B**{t1,t_sortie}, {t4[1,1]}	**e10:** p0,p4, p5 **B**{t3}, {t_exc2[1,1]}	**e13:** p_init,p5 **B**{}, {t_entrée[1,1]}	**p16:** p3, p4, p5(2) **B**{}, {t_exc[1,1]}
e2: p3, **B**{}, {t1[1,2], t4[3,3], t_sortie[1,2]}	**e5:** p0,p4, **B**{}, {t3[1,1], t_exc2[2,2]}	**e8:** p2,p4,p5 **B**{}, {t2[4,4], t3[5,5], t_exc2[6,6]}	**e11:**p0,p1,p5, **B**{}, {t0[1,1]}	**e14:** p3,p5, **B**{}, {t1[1,1], t4[2,2], t_sortie[1,1]}	

Figure 14. Synchronous Behavior Graph of the model given Figure 13

During the internship, we developed software, based on algorithms available in literature, that allow to recover M-Wave induced by surface FES. The algorithm detects the onset of the artifact, the Otsu method is used to determine the length of the contaminated data and finally, the M-wave is interpolated through Cubic Hermite extensions. The results can be seen on Figure 15 where the M-wave, fully reconstructed can then be quantified. We thus verified that the classical torque-EMG relationship can be recovered using MAV, RMS or P2P and the results show a very good correlation. Finally we successfully developed a realtime version of the processing and tested it through a closed loop control of the FES through EMG measurements. The technology used was completely wireless (Delsys for the EMG and Vivaltis for the stimulation). One journal paper is under writing in collaboration with the Technological University of Compiègne.

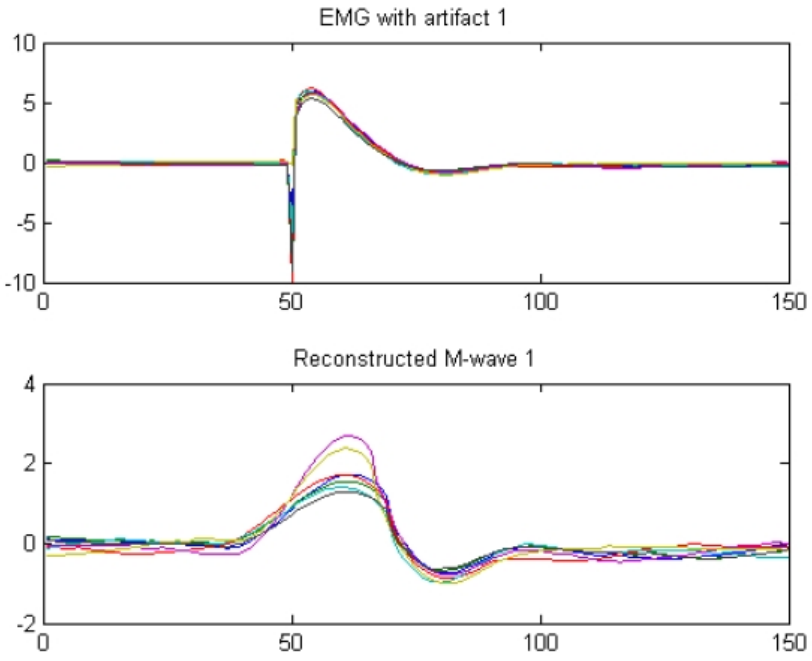


Figure 15. Reconstructed M-Wave

6.10. Pivot transfer assistance in SCI subjects

Participants: Lucas Fonseca [Univ. Brasilia], Antonio Padilha Lanari Bo [Univ. Brasilia], Ana Claudia Lopes [SARA hosp], Christine Azevedo Coste, Emerson Fachin Martins [Univ. Brasilia], Claudia Ochoa-Diaz [Univ. Brasilia].

Spinal cord injured (SCI) patients that have no lower limb motor function perform several transfers during a day. Those transfers are from and to a wheelchair, a car, a hygienic chair, among other situations. These repetitive motions can cause overload on their upper limbs over time. Functional Electrical Stimulation may be used to induce contraction on knee extensors, providing additional support at the joint level during transfer [33]. However, the design of the interface with which to control the onset of stimulation is challenging. The use of some automated system is beneficial, particularly since the user is using both hands to perform the transfer. Therefore, the precise moment of activation is important because, if erroneous, it can cause the user's loss of balance. In the context of CACAO associate team with Brasilia University, a system with which the

users themselves were activating the stimulation with triggers in gloves was used to collect kinematic data from SCI patients during Sitting Pivot Transfers (16). The results show that the trunk angle can be used along a threshold for a reliable assistance device [28].

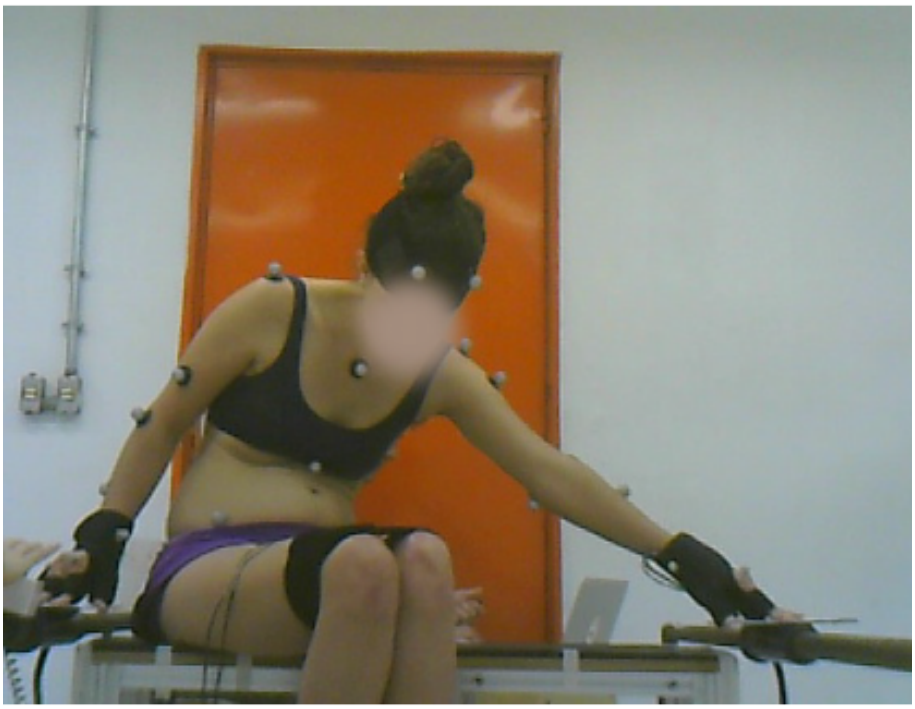


Figure 16. Experimental set-up. The gloves embed pressure sensors. It is possible to see the markers over the subject body, which are captured by the motion capture system.

6.11. Real-time control and scheduling for stimulation systems

Participants: Daniel Simon, David Andreu, Ronan Le Guillou, Benoît Sijobert.

Functional Electrical Stimulation (FES) is used in therapy for rehabilitation or substitution for disabled people. They are control systems using electrodes to interface a digital control system with livings. Hence the whole system gathers continuous-time (muscles and nerves) and discrete-time (controllers and communication links) components. During the design process, realistic simulation remains a precious tool ahead of real experiments to check without danger that the implementation matches the functional and safety requirements [15].

To this aim a real-time open hybrid simulation software has been developed. It is dedicated to the analysis of FES systems deployed over distributed execution resources and wireless links. The simulation tool is especially devoted to the joint design and analysis of control loops and real-time features. Such simulator can be used for the design, testing and preliminary validation of new technologies and implementation. The initial design, working with a simple model of a knee, is currently extended with the dynamic model of a human hand (Figure 17).

A portable controller has been prototyped to run control loops using stimulation and sensing probes. ([32]). It is architectured around a Raspberry Pi3B single board computer, and provides USB ports towards sensing probes from HiKoB and stimulation units from Vivaltis. It uses a dedicated RT_PREEMPT linux kernel to make the system real-time control compliant (Figure 18).

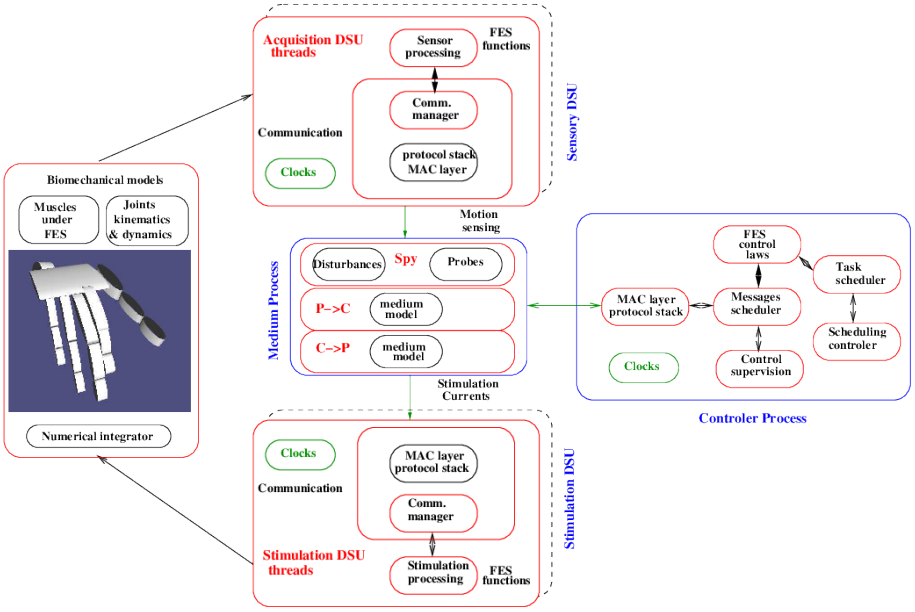


Figure 17. Hybrid simulation of a hand under FES

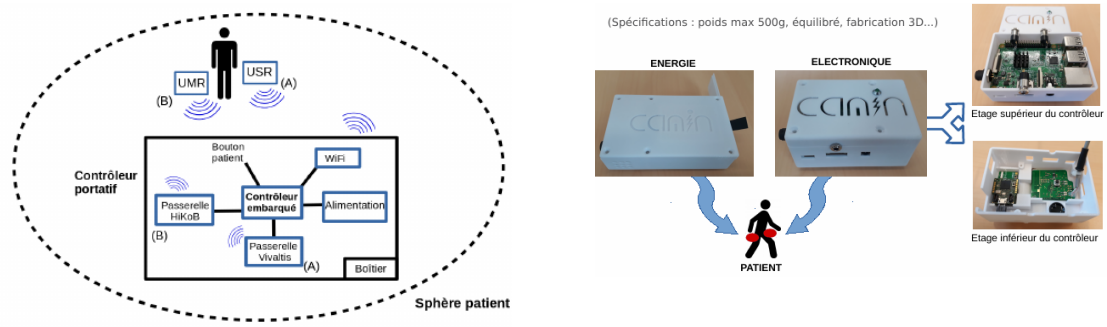


Figure 18. Portable stimulation controller architecture and components

Beyond software-in-the-loop simulation, the controller has been firstly tested connected with the previously developed simulation models to perform a hardware-in-the-loop simulation system and further experiment control/computing co-design algorithms ([25]). Gateways has been developed to connect the Vivaltis and HiKoB probes, together with a small graphical end-user interface. The whole embedded system has been succesfully validated for future applications through a pertinent real-time metrology.

6.12. Sensory feedback for phantom limb pain modulation

Participants: Arthur Hiairrassary, David Andreu, Christine Azevedo Coste, Thomas Guiho, David Guiraud.

In the EPIONE european project, the partners UM and MXM-OBELIA are responsible for the design and manufacturing of the STIMEP stimulator (see figure 1) and in charge of all the software and the experimental follow-up.

During the first round, we were able to quantify the state of each contact of each electrode to prevent misinterpretation of feedback sensation. Indeed, if the patient does not feel anything while stimulating, impedance check may show that it is due to a contact failure and not to a lack of nerve response.

This estimation was done during the “Contacts Check” functionality embedded in the STIMEP. At the same time, a more detailed measure was stored in the STIMEP (but only reachable off-line for further investigation).

For instance, the following figure shows the number of valid contacts during the clinical phase of the 4 TIME-4H electrodes computed by the “Contacts Check” functionality. The electrodes stand almost OK up to February-March on this example (2-3 months) then failures begin to occur.

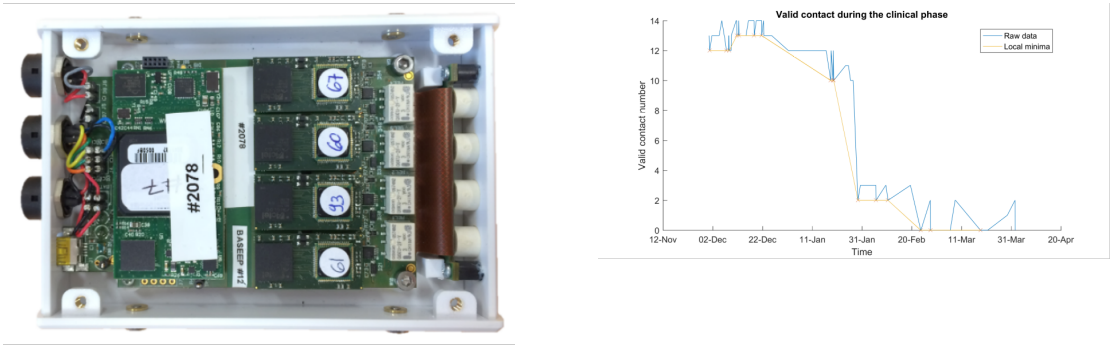


Figure 19. a) the STIMEP b) Ulnar nerve – Proximal electrode

However, these data were not detailed enough so we decided to develop an easy to use software able to control automatically in the same time the STIMEP and the acquisition card (NI-6218). It performed voltage and current measurements to follow and to assess the complex impedance evolution of the TIME-4H in vivo. This software, named “Synergy Acquire” (Figure 3), was used by the clinic of Roma with the second patient.

Synergy Acquire performs safe, really quick stimulation and measurement on an electrode (around 1 minute, less than 5 minutes for the 4 TIME-4H, figure 4), which allow a very regular follow-up (2-3 times by week) by the practitioners. Data were logged, sent to us, processed and then sent back to UCSC for checking. This work is related to clinical trials follow-up. In conclusion, within this project that ended in August, we developed software (following 62304 class B regulation), test in animals, used in humans and we processed all the data

6.13. Spin-off Neurinnov

Participants: David Andreu, David Guiraud, Olivier Climent, Milan Demarcq, Guillaume Souquet.

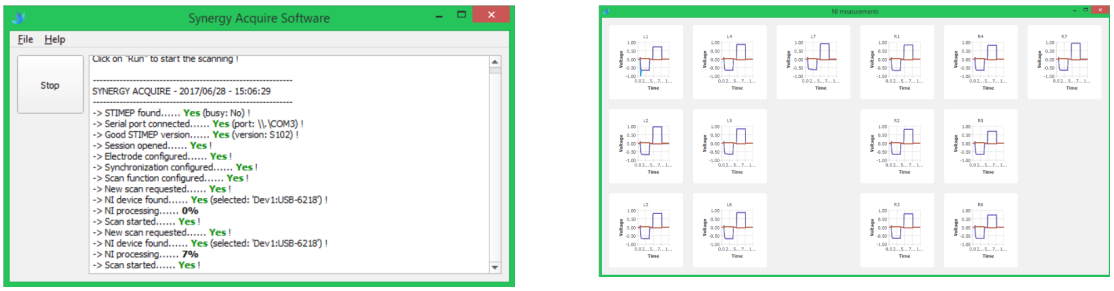


Figure 20. a) Synergy Acquire b) Results of the voltage and current measurements

Thanks to the support of the Inria-DGDT, the spin-off Neurinnov has started the industrialization of the innovative technology developed in CAMIN, a new generation of active implantable medical device (AIMD). Neurinnov has been awarded with the i-Lab 2017 prize by the French Minister of Research and Innovation, that encourage the most innovative and promising startups in France. Moreover, Neurinnov is accredited by the Business Incubation Center of Montpellier and incubated by the Languedoc-Roussillon Incubation Center.

Industrialization means on the one hand the development of an industrial version of the technology with all the regulatory documents required by the Technical Documents part of the CE certification, and on the other hand the setting up of our own system of quality management in accordance with ISO 13485. Since the beginning the spin-off has to consider regulatory aspects, of which the quality management system (QMS). A QMS is a set of policies, processes and procedures designed to help an organization to consistently provide safe and effective medical devices, and to comply with customer and regulatory requirements. Thus the team worked on defining processes and associated procedures regarding for instance the design, the development and the verification of our stimulation device.

The design and verification of the two parts of our AIMD (stimulator), namely the digital part (FPGA) and the analogue part (ASIC), were carried out in accordance with the defined procedure and applicable standards.

The core of the stimulators developed in the CAMIN team are based on an Application Specific Integrated Circuit (ASIC). It includes both the analog part with the generation of 12 current sources that are able to drive a pulticontact electrode. The global ASIC architecture fully implements our patent and allows to spread the current from a unique current source over the 12 outputs through ratios programming. This unique feature is a consequence of researches about the multicontact selective stimulation through neural cuffs. In the new 0.18μ new design, the analog part was entirely revised to enhance power consumption and global analogue features. On the digital aprt the concept of virtual electrode was fully implemented within this ASIC (named CORAIL) to embed all the low level programming parts dedicated to the spreding of the current. It enhances the safety but also the efficacy of the code developed to control this ASIC as it virtualize the concept of ratios. Moreover CORAIL stores the Virtual Electrode so that the needed bandwidth but also the transfer time between CORAIL and the hiigh level control is much lowered compared to the previous version we developed. We tested and implemented this original digital part in collaboration with the micro-electronics department at LIRMM and SL3J company. A first version of the ASIC was made and is under investigation to prepare the next version.

The digital part of the device embeds a set of functionalities (described section 5.1.1) allowing the stimulator to be programmable, communicating and fully controllable remotely. The formal design and verification of this digital part is based on the HILECOP software developed within CAMIN (see sections 5.1.1 and 6.7). All constituent components have been developed, verified and documented in accordance with the defined procedure.

In addition, Neurinnov has focused on setting up the necessary industrial collaborations on the one hand to complement its device (e.g., electrodes, connectors) and on the other hand to manufacture it.

GALEN Project-Team

7. New Results

7.1. Graph Based Slice-to-Volume Deformable Registration

Participants: Enzo Ferrante, Nikos Paragios

Deformable image registration is a fundamental problem in computer vision and medical image computing. In this contribution [9], we investigate the use of graphical models in the context of a particular type of image registration problem, known as slice-to-volume registration, while we introduced the first comprehensive survey [10] of the literature about slice-to-volume registration, presenting a categorical study of the algorithms according to an ad-hoc taxonomy and analyzing advantages and disadvantages of every category. We introduce a scalable, modular and flexible formulation that can accommodate low-rank and high order terms, that simultaneously selects the plane and estimates the in-plane deformation through a single shot optimization approach. The proposed framework is instantiated into different variants seeking either a compromise between computational efficiency (soft plane selection constraints and approximate definition of the data similarity terms through pair-wise components) or exact definition of the data terms and the constraints on the plane selection. Simulated and real-data in the context of ultrasound and magnetic resonance registration (where both framework instantiations as well as different optimization strategies are considered) demonstrate the potentials of our method.

7.2. Deformable Registration Through Learning of Context-Specific Metric Aggregation

Participants: Enzo Ferrante, Rafael Marini, Punnet K. Dokania, Nikos Paragios

We propose a novel weakly supervised discriminative algorithm [21] for learning context specific registration metrics as a linear combination of conventional similarity measures. Conventional metrics have been extensively used over the past two decades and therefore both their strengths and limitations are known. The challenge is to find the optimal relative weighting (or parameters) of different metrics forming the similarity measure of the registration algorithm. Hand-tuning these parameters would result in sub optimal solutions and quickly become infeasible as the number of metrics increases. Furthermore, such hand-crafted combination can only happen at global scale (entire volume) and therefore will not be able to account for the different tissue properties. We propose a learning algorithm for estimating these parameters locally, conditioned to the data semantic classes. The objective function of our formulation is a special case of non-convex function, difference of convex function, which we optimize using the concave convex procedure. As a proof of concept, we show the impact of our approach on three challenging datasets for different anatomical structures and modalities.

7.3. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology

Participants: Roger Sun, Evangelia I. Zacharaki, Nikos Paragios (in collaboration with Gustave Roussy and Paris Sud University)

Computational medical imaging (also known as radiomics) is a promising and rapidly growing discipline that consists in the analysis of high-dimensional data extracted from medical imaging, to further describe tumour phenotypes. The end goal of radiomics is to determine imaging biomarkers as decision support tools for clinical practice and to facilitate better understanding of cancer biology, allowing the assessment of the changes throughout the evolution of the disease and the therapeutic sequence. We have reviewed [12], [17] the critical issues necessary for proper development of radiomics as a biomarker and for its implementation in clinical practice.

7.4. Multi-atlas segmentation in medical imagery

Participants: Stavros Alchatzidis, Evangelia I. Zacharaki, Nikos Paragios (in collaboration with University of Pennsylvania)
Multi-atlas segmentation has emerged in recent years as a simple yet powerful approach in medical image segmentation. It commonly comprises two steps: (1) a series of pairwise registrations that establish correspondences between a query image and a number of atlases, and (2) the fusion of the available segmentation hypotheses towards labeling objects of interest. In [5], we introduce a novel approach that solves simultaneously for the underlying segmentation labels and the multi-atlas registration. We propose a pairwise Markov Random Field approach,where registration and segmentation nodes are coupled towards simultaneously recovering all atlas deformations and labeling the query image.

7.5. Protein function prediction

Participants: Evangelia I. Zacharaki, Nikos Paragios (in collaboration with University of Patras)
The massive expansion of the worldwide Protein Data Bank (PDB) provides new opportunities for computational approaches which can learn from available data and extrapolate the knowledge into new coming instances. The aim of our work in [6], [18] was to exploit experimentally acquired structural information of enzymes through machine learning techniques in order to produce models that predict enzymatic function.

7.6. Deformable group-wise registration using a physiological model: Application to diffusion-weighted MRI

Participants: Evgenios Kornaropoulos, Evangelia I. Zacharaki, Nikos Paragios (in collaboration with Centre Hospitalier Universitaire Henri-Mondor and Chang Gung Memorial Hospital)
In this contribution [2] we develop a novel group-wise deformable registration method for motion correction in Diffusion-Weighted MRI towards computing a more accurate Apparent Diffusion Coefficient parametric map (ADC map). Calculation of the ADC has been performed without motion correction in the previous studies. It is reported though that ADC is a parameter susceptible to artifacts, the most frequent of all being patient's motion and breathing, resulting in misregistration of the images obtained with different b-values. Being group-wise designed, the image registration method we propose has no need of choosing a reference template while in the same time it is computationally efficient. We aim at finding the optimal deformation fields of the diffusion-weighted (DW) images using a temporal constraint, related to the diffusion process, as well as a smoothness penalty on the deformations. To this end, we address the deformation fields estimation problem with an Markov Random Fields formulation, in which the latent variables are the deformations (B-spline polynomials) of the images. The latent variables are connected with the observations towards ensuring meaningful temporal correspondence among the DW images. They are also inter-connected in order to decrease the cost of pairwise comparisons between individual images. Linear programming and duality are used to determine the optimal solution of the problem. Finally, as an image similarity criterion in the MRF framework, we used a metric that was based on a physiological model describing the image acquisition process. Quantitative evaluation of the method was performed, in which it was compared against two state-of-the-art methods that use other modelling criteria, It outperformed both of them, while the ADC map derived by our method appeared to preserve structure, that was not observable by the other methods.

7.7. Variational Bayesian Approach for Image Restoration

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Y. Marnissi, SAFRAN TECH and Y. Zheng, IBM Research China)

In the work [13], a methodology is investigated for signal recovery in the presence of non-Gaussian noise. In contrast with regularized minimization approaches often adopted in the literature, in our algorithm the regularization parameter is reliably estimated from the observations. As the posterior density of the unknown parameters is analytically intractable, the estimation problem is derived in a variational Bayesian framework where the goal is to provide a good approximation to the posterior distribution in order to compute posterior mean estimates. Moreover, a majorization technique is employed to circumvent the difficulties raised by the intricate forms of the non-Gaussian likelihood and of the prior density. We demonstrate the potential of the proposed approach through comparisons with state-of-the-art techniques that are specifically tailored to signal recovery in the presence of mixed Poisson-Gaussian noise. Results show that the proposed approach is efficient and achieves performance comparable with other methods where the regularization parameter is manually tuned from an available ground truth.

7.8. Non-Modular Loss Functions

Participant: Jiaqian Yu

Defining Non-modular loss functions and their optimization procedure present an interesting direction for many classes of problems. Jiaqian Yu has completed her PhD Thesis on Non-Modular Loss Functions this year. The PhD Thesis has included several yet unpublished results regarding approximate losses for Jaccard index and DICE coefficients commonly used in evaluating segmentation algorithms.

7.9. Graph Structure Discovery

Participant: Eugene Belilovsky

Discovering the interaction structure amongst variables, particularly from few observations, has important implications in many fields including neuroimaging, genetics and finance. Eugene Belilovsky in collaboration with Gael Varoquaux (Inria Parietal), Kyle Kastner (University of Montreal) and Matthew Blaschko has published a new approach for graph structure discovery in high dimensional gaussian markov random fields. The work has been presented in [19].

7.10. Structured and Efficient Convolutional Networks

Participant: Eugene Belilovsky

Convolutional Neural Networks have revolutionized the computer vision field. Yet, they are not well understood and do not well leverage basic geometric structures known by the computer vision community. In recent work in collaboration with the Ecole Normale Superier and the École des Ponts ParisTech we have tried to address some of these issues. We use as a starting point the recently introduced Scattering Transform and show that we can use this to build Convolutional Networks that are more interpertable and can generalize faster in the few sample regime. This work has been presented in [25].

7.11. Stochastic Majorize-Minimize Subspace Algorithm

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet

Stochastic optimization plays an important role in solving many problems encountered in machine learning or adaptive processing. In this context, the second-order statistics of the data are often unknown a priori or their direct computation is too intensive, and they have to be estimated on-line from the related signals. In the context of batch optimization of an objective function being the sum of a data fidelity term and a penalization (e.g. a sparsity promoting function), Majorize-Minimize (MM) subspace methods have recently attracted much interest since they are fast, highly flexible and effective in ensuring convergence. The goal of the work [8] is to show how these methods can be successfully extended to the case when the cost function is replaced by a sequence of stochastic approximations of it. Simulation results illustrate the good practical performance of the proposed MM Memory Gradient (3MG) algorithm when applied to 2D filter identification

7.12. Deconvolution and Deinterlacing of Video Sequences

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with F. Abboud, PhD student, J.-H. Chenot and L. Laborelli, research engineers, Institut National de l'Audiovisuel)

Optimization methods play a central role in the solution of a wide array of problems encountered in various application fields, such as signal and image processing. Especially when the problems are highly dimensional, proximal methods have shown their efficiency through their capability to deal with composite, possibly non smooth objective functions. The cornerstone of these approaches is the proximity operator, which has become a quite popular tool in optimization. In the work [31], we propose new dual forward-backward formulations for computing the proximity operator of a sum of convex functions involving linear operators. The proposed algorithms are accelerated thanks to the introduction of a block coordinate strategy combined with a preconditioning technique. Numerical simulations emphasize the good performance of our approach for the problem of jointly deconvoluting and deinterlacing video sequences.

7.13. PALMA, an improved algorithm for DOSY signal processing

Participants: Emilie Chouzenoux (in collaboration with M.-A. Delsuc, IGBMC, Strasbourg, and A. Cherni, PhD student, Univ. Strasbourg)

NMR is a tool of choice for the measure of diffusion coefficients of species in solution. The DOSY experiment, a 2D implementation of this measure, has proven to be particularly useful for the study of complex mixtures, molecular interactions, polymers, etc. However, DOSY data analysis requires to resort to inverse Laplace transform, in particular for polydisperse samples. This is a known difficult numerical task, for which we present here a novel approach. A new algorithm based on a splitting scheme and on the use of proximity operators is introduced in [7]. Used in conjunction with a Maximum Entropy and ℓ_1 hybrid regularisation, this algorithm converges rapidly and produces results robust against experimental noise. This method has been called PALMA. It is able to reproduce faithfully monodisperse as well as polydisperse systems, and numerous simulated and experimental examples are presented. It has been implemented on the server http://palma.labo.igbmc.fr where users can have their datasets processed automatically.

7.14. Proximal Approaches for Solving Matrix Optimization Problems

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with A. Benfenati, Univ. Paris Est)

In recent years, there has been a growing interest in problems where the underlining mathematical model involves the minimization in a matrix space of a Bregman divergence function coupled with a regularization term. We consider a general framework where the regularization term is decoupled in two parts, one acting only on the eigenvalues of the matrix and the other on the whole matrix. We propose in [26], [32] a new minimization approach to address problem of this type, by providing a list of proximity operators allowing us to consider various choices for the fit–to–data functional and for the regularization term. The numerical experience show that this approach gives better results in term of computational time with respect to some state of the arts algorithms.

7.15. Fast Algorithm for Least-Squares Regression with GMRF Prior

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with J.Y. Tourneret, IRIT, Toulouse, and Q. Wei, Duke Univ.)

The paper [29] presents a fast approach for penalized least squares (LS) regression problems using a 2D Gaussian Markov random field (GMRF) prior. More precisely, the computation of the proximity operator of the LS criterion regularized by different GMRF potentials is formulated as solving a Sylvester-like matrix equation. By exploiting the structural properties of GMRFs, this matrix equation is solved column-wise in an analytical way. The proposed algorithm can be embedded into a wide range of proximal algorithms to solve LS regression problems including a convex penalty. Experiments carried out in the case of a constrained

LS regression problem arising in a multichannel image processing application, provide evidence that an alternating direction method of multipliers performs quite efficiently in this context.

7.16. Optimization Approach for Deep Neural Network Training

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Vyacheslav Dudar (in collaboration with G. Chierchia, Univ. Paris Est and V. Semenov, Univ. of Kiev)

In paper [28], we develop a novel second-order method for training feed-forward neural nets. At each iteration, we construct a quadratic approximation to the cost function in a low-dimensional subspace. We minimize this approximation inside a trust region through a two-stage procedure: first inside the embedded positive curvature subspace, followed by a gradient descent step. This approach leads to a fast objective function decay, prevents convergence to saddle points, and alleviates the need for manually tuning parameters. We show the good performance of the proposed algorithm on benchmark datasets.

7.17. Auxiliary Variable Method for MCMC Algorithms in High Dimension

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Y. Marnissi, SAFRAN TECH and A. Benazza-Benhayia, SUP'COM, COSIM, Tunis)

When the parameter space is high dimensional, the performance of stochastic sampling algorithms is very sensitive to existing dependencies between parameters. For instance, this problem arises when one aims to sample from a high dimensional Gaussian distribution whose covariance matrix does not present a simple structure. Then, one often resorts to sampling algorithms based on a perturbation-optimization technique that requires to minimize a cost function using an iterative algorithm. This makes the sampling process time consuming, especially when used within a Gibbs sampler. Another challenge is the design of Metropolis-Hastings proposals that make use of information about the local geometry of the target density in order to speed up the convergence and improve mixing properties in the parameter space, while being not too computationally expensive. These two contexts are mainly related to the presence of two heterogeneous sources of dependencies stemming either from the prior or the likelihood in the sense that the related covariances matrices cannot be diagonalized in the same basis. In paper [34], we are interested in inverse problems where either the data fidelity term or the prior distribution is Gaussian or driven from a hierarchical Gaussian model. We propose to add auxiliary variables to the model in order to dissociate the two sources of dependencies. In the new augmented space, only one source of correlation remains directly related to the target parameters, the other sources of correlations being captured by the auxiliary variables. Experiments conducted on two image restoration problems show the good performance of the proposed strategy.

7.18. Block Coordinate Approach for Sparse Logistic Regression

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with G. Chierchia, Univ. Paris Est, L. M. Briceno-Arias, CMM - Univ. Chile, and A. Cherni, PhD student, Univ. Strasbourg)

We propose in [20], [33] stochastic optimization algorithms for logistic regression based on a randomized version of Douglas–Rachford splitting method. Our approach sweeps the training set by randomly selecting a mini-batch of data at each iteration, and it performs the update step by leveraging the proximity operator of the logistic loss, for which a closed-form expression is derived. Experiments carried out on standard datasets compare the efficiency of our algorithm to stochastic gradient-like methods.

7.19. An Alternating Proximal Approach for Blind Video Deconvolution

Participants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Feriel Abboud, WITBE, Jean-Hugues Chenot, Louis Laborelli, INA)

Blurring occurs frequently in video sequences captured by consumer devices, as a result of various factors such as lens aberrations, defocus, relative camera scene motion, and camera shake. When it comes to the contents of archive documents such as old films and television shows, the degradations are even more serious due to several physical phenomena happening during the sensing, transmission, recording, and storing processes. We propose in [31] a versatile formulation of blind video deconvolution problems that seeks to estimate both the sharp unknown video sequence and the underlying blur kernel from an observed video. This inverse problem is severely ill-posed, and an appropriate solution can be obtained by modeling it as a nonconvex minimization problem. We provide a novel iterative algorithm to solve it, grounded on the use of recent advances in convex and nonconvex optimization techniques, and having the ability of including numerous well-known regularization strategies

7.20. BRANE Clust: Cluster-Assisted Gene Regulatory Network Inference Refinement

Participants: Jean-Christophe Pesquet (in collaboration with Aurélie Pirayre, IFP Energies nouvelles, Camille Couprie, Facebook Research, Laurent Duval, IFP Energies nouvelles)

Discovering meaningful gene interactions is crucial for the identification of novel regulatory processes in cells. Building accurately the related graphs remains challenging due to the large number of possible solutions from available data. Nonetheless, enforcing a priori on the graph structure, such as modularity, may reduce network indeterminacy issues. BRANE Clust (Biologically-Related A priori Network Enhancement with Clustering) refines gene regulatory network (GRN) inference thanks to cluster information. It works as a post-processing tool for inference methods (i.e. CLR, GENIE3). In BRANE Clust, the clustering is based on the inversion of a system of linear equations involving a graph-Laplacian matrix promoting a modular structure. Our approach [16] is validated on DREAM4 and DREAM5 datasets with objective measures, showing significant comparative improvements. We provide additional insights on the discovery of novel regulatory or co-expressed links in the inferred Escherichia coli network evaluated using the STRING database. The comparative pertinence of clustering is discussed computationally (SIMoNe, WGCNA, X-means) and biologically (RegulonDB).

7.21. Proximity Operators of Discrete Information Divergences

Participants: Jean-Christophe Pesquet (in collaboration with Mireille El Gheche, EPFL, Giovanni Chierchia, ESIEE Paris)

Information divergences allow one to assess how close two distributions are from each other. Among the large panel of available measures, a special attention has been paid to convex ϕ-divergences, such as Kullback-Leibler, Jeffreys-Kullback, Hellinger, Chi-Square, Renyi, and I_α divergences. While ϕ-divergences have been extensively studied in convex analysis, their use in optimization problems often remains challenging. In this regard, one of the main shortcomings of existing methods is that the minimization of ϕ-divergences is usually performed with respect to one of their arguments, possibly within alternating optimization techniques. In this paper, we overcome this limitation by deriving new closed-form expressions for the proximity operator of such two-variable functions. This makes it possible to employ standard proximal methods for efficiently solving a wide range of convex optimization problems involving ϕ-divergences. In addition, we show that these proximity operators are useful to compute the epigraphical projection of several functions of practical interest. The proposed proximal tools are numerically validated in the context of optimal query execution within database management systems, where the problem of selectivity estimation plays a central role. Experiments are carried out on small to large scale scenarios.

7.22. Stochastic Quasi-Fejér Block-Coordinate Fixed Point Iterations With Random Sweeping: Mean-Square and Linear Convergence

Participants: Jean-Christophe Pesquet (in collaboration with Patrick L. Combettes, North Caroline State University)

In one of our previous works, we investigated the almost sure weak convergence of a block-coordinate fixed point algorithm and discussed its application to nonlinear analysis and optimization. This algorithm features random sweeping rules to select arbitrarily the blocks of variables that are activated over the course of the iterations and it allows for stochastic errors in the evaluation of the operators. The present paper establishes results on the mean-square and linear convergence of the iterates. Applications to monotone operator splitting and proximal optimization algorithms are presented.

7.23. Human Joint Angle Estimation and Gesture Recognition for Assistive Robotic Vision

Participants: Riza Alp Guler, Siddhartha Chandra, Iasonas Kokkinos (in collaboration with National Technical University of Athens)

In this work, we explore new directions for automatic human gesture recognition and human joint angle estimation as applied for human-robot interaction in the context of an actual challenging task of assistive living for real-life elderly subjects. Our contributions include state-of-the-art approaches for both low- and mid-level vision, as well as for higher level action and gesture recognition. The first direction investigates a deep learning based framework for the challenging task of human joint angle estimation on noisy real world RGB-D images. The second direction includes the employment of dense trajectory features for online processing of videos for automatic gesture recognition with real-time performance. Our approaches are evaluated both qualitative and quantitatively on a newly acquired dataset that is constructed on a challenging real-life scenario on assistive living for elderly subjects.

7.24. Fast, Exact and Multi-Scale Inference for Semantic Image Segmentation with Deep Gaussian CRFs

Participants: Siddhartha Chandra, Iasonas Kokkinos

In this work we propose a structured prediction technique that combines the virtues of Gaussian Conditional Random Fields (G-CRF) with Deep Learning: (a) our structured prediction task has a unique global optimum that is obtained exactly from the solution of a linear system (b) the gradients of our model parameters are analytically computed using closed form expressions, in contrast to the memory-demanding contemporary deep structured prediction approaches that rely on back-propagation-through-time, (c) our pairwise terms do not have to be simple hand-crafted expressions, as in the line of works building on the DenseCRF, but can rather be 'discovered' from data through deep architectures, and (d) out system can trained in an end-to-end manner. Building on standard tools from numerical analysis we develop very efficient algorithms for inference and learning, as well as a customized technique adapted to the semantic segmentation task. This efficiency allows us to explore more sophisticated architectures for structured prediction in deep learning: we introduce multi-resolution architectures to couple information across scales in a joint optimization framework, yielding systematic improvements. We demonstrate the utility of our approach on the challenging VOC PASCAL 2012 image segmentation benchmark, showing substantial improvements over strong baselines.

7.25. Dense and Low-Rank Gaussian CRFs Using Deep Embeddings

Participants: Siddhartha Chandra, Iasonas Kokkinos

In this work we introduce a structured prediction model that endows the Deep Gaussian Conditional Random Field (G-CRF) with a densely connected graph structure. We keep memory and computational complexity under control by expressing the pairwise interactions as inner products of low-dimensional, learnable embeddings. The G-CRF system matrix is therefore low-rank, allowing us to solve the resulting system in a few milliseconds on the GPU by using conjugate gradient. As in G-CRF, inference is exact, the unary and pairwise terms are jointly trained end-to-end by using analytic expressions for the gradients, while we also develop even faster, Potts-type variants of our embeddings. We show that the learned embeddings capture pixel-to-pixel affinities in a task-specific manner, while our approach achieves state of the art results on three challenging benchmarks, namely semantic segmentation, human part segmentation, and saliency estimation. This work was published in [30].

7.26. DenseReg: Fully Convolutional Dense Shape Regression In-the-Wild

Participants: Riza Alp Guler, Iasonas Kokkinos (in collaboration with Imperial College London)

In this work we propose to learn a mapping from image pixels into a dense template grid through a fully convolutional network. We formulate this task as a regression problem and train our network by leveraging upon manually annotated facial landmarks “in-the-wild". We use such landmarks to establish a dense correspondence field between a three-dimensional object template and the input image, which then serves as the ground-truth for training our regression system. We show that we can combine ideas from semantic segmentation with regression networks, yielding a highly-accurate ‘quantized regression’ architecture.

Our system, called DenseReg allows us to estimate dense image-to-template correspondences in a fully convolutional manner. As such our network can provide useful correspondence information as a stand-alone system, while when used as an initialization for Statistical Deformable Models we obtain landmark localization results that largely outperform the current state-of-the-art on the challenging 300W benchmark. We thoroughly evaluate our method on a host of facial analysis tasks, and also demonstrate its use for other correspondence estimation tasks, such as modelling of the human ear. This work was publised in [22].

7.27. Structured Output Prediction and Learning for Deep Monocular 3D Human Pose Estimation

Participants: Stefan Kinauer, Riza Alp Guler, Siddhartha Chandra, Iasonas Kokkinos

In this work we address the problem of estimating 3D human pose froma single RGB image by blending a feed-forward Convolutional Neural Network (CNN) with a graphical model thatcouples the 3D positions of parts. The CNN populates a volumetric output space that represents the possible positions of 3D human joints, and also regresses the estimated displacements between pairs of parts. These constitute the ‘unary’ and ‘pairwise’ terms of the energy of a graphical model that resides in a 3D label spaceand delivers an optimal 3D pose configuration at its output. The CNN is trained onthe 3D human pose dataset 3.6M, the graphical model is trained jointly with the CNN in an end-to-end manner, allowing us to exploit both the discriminative powerof CNNs and the top-down information pertaining to human pose. We introduce(a) memory efficient methods for getting accurate voxel estimates for parts byblending quantization with regression (b) employ efficient structured prediction algorithms for 3D pose estimation using branch-and-bound and (c) develop a framework for qualitative and quantitative comparison of competing graphical models. We evaluate our work on the Human 3.6M dataset, demonstrating that exploiting the structure of the human pose in 3D yields systematic gains.

7.28. Newton-type Methods for Inference in Higher-Order Markov Random Fields

Participants: Hariprasad Kannan, Nikos Paragios

Linear programming relaxations are central to MAP inference in discrete Markov Random Fields. The ability to properly solve the Lagrangian dual is a critical component of such methods. In this paper, we study the benefit of using Newton-type methods to solve the Lagrangian dual of a smooth version of the problem. We investigate their ability to achieve superior convergence behavior and to better handle the ill-conditioned nature of the formulation, as compared to first order methods. We show that it is indeed possible to efficiently apply a trust region Newton method for a broad range of MAP inference problems. In this paper we propose a provably convergent and efficient framework that includes (i) excellent compromise between computational complexity and precision concerning the Hessian matrix construction, (ii) a damping strategy that aids efficient optimization , (iii) a truncation strategy coupled with a generic pre-conditioner for Conjugate Gradients, (iv) efficient sum-product computation for sparse clique potentials. Results for higher-order Markov Random Fields demonstrate the potential of this approach. This work was published in [23].

7.29. Alternating Direction Graph Matching

Participants: D. Khuê Lê-Huu, Nikos Paragios

In this work, we introduce a graph matching method that can account for constraints of arbitrary order, with arbitrary potential functions. Unlike previous decomposition approaches that rely on the graph structures, we introduce a decomposition of the matching constraints. Graph matching is then reformulated as a non-convex non-separable optimization problem that can be split into smaller and much-easier-to-solve subproblems, by means of the alternating direction method of multipliers. The proposed framework is modular, scalable, and can be instantiated into different variants. Two instantiations are studied exploring pairwise and higher-order constraints. Experimental results on widely adopted benchmarks involving synthetic and real examples demonstrate that the proposed solutions outperform existing pairwise graph matching methods, and competitive with the state of the art in higher-order settings. This work was published in [24].

7.30. Prediction and classification in biological and information networks

Participants: Fragkiskos Malliaros (in collaboration with Duong Nguyen, UC San Diego)
We investigate how network representation learning algorithms can be applied to deal with the problem of link prediction and classification in protein-protein interaction networks as well as in social and information networks. In particular, we have proposed BiasedWalk, a scalable, unsupervised feature learning algorithm that is based on biased random walks to sample context information about each node in the network.

MATHNEURO Team

4. New Results

4.1. Neural Networks as dynamical systems

4.1.1. Latching dynamics in neural networks with synaptic depression

Participants: Elif Köksal Ersöz, Carlos Aguilar [Université de Nice - BCL], Pascal Chossat [Université de Nice - LJAD, Inria MathNeuro], Martin Krupa [UCA, Inria MathNeuro], Frédéric Lavigne [Université de Nice - BCL].

Prediction is the ability of the brain to quickly activate a target concept in response to a related stimulus (prime). Experiments point to the existence of an overlap between the populations of the neurons coding for different stimuli, and other experiments show that prime-target relations arise in the process of long term memory formation. The classical modelling paradigm is that long term memories correspond to stable steady states of a Hopfield network with Hebbian connectivity. Experiments show that short term synaptic depression plays an important role in the processing of memories. This leads naturally to a computational model of priming, called latching dynamics; a stable state (prime) can become unstable and the system may converge to another transiently stable steady state (target). Hopfield network models of latching dynamics have been studied by means of numerical simulation, however the conditions for the existence of this dynamics have not been elucidated. In this work we use a combination of analytic and numerical approaches to confirm that latching dynamics can exist in the context of a symmetric Hebbian learning rule, however lacks robustness and imposes a number of biologically unrealistic restrictions on the model. In particular our work shows that the symmetry of the Hebbian rule is not an obstruction to the existence of latching dynamics, however fine tuning of the parameters of the model is needed.

This work has been published in PLoS one and is available as [13].

A natural follow-up of the work which has lead to the article [13] has been initiated through the postdoc project of Elif Köksal Ersöz. The objective is to extend the previous results in several ways. First, to gain more robustness in the heteroclinic chains sustained by the network model. Second, to be able to simulate much larger networks and exhibit heteroclinic dynamics in them. Third, to link with experimental data. The postdoc of Elif Köksal Ersöz is funded by the "tail" of the ERC Advanced Grant NerVi held by Olivier Faugeras.

4.1.2. Special issue for Martin Golubitsky

Participants: Pietro-Luciano Buono University Of Ontario Institute Of Technology, Canada ,, Martin Krupa [UCA, Inria MathNeuro], Ian Stewart [University of Warwick, UK].

The work is the introducion of this special issue, co-edited by Martin Krupa. It has been published in Dynamical Systems: An International Journal and is available as [17].

4.1.3. Consecutive and non-consecutive heteroclinic cycles in Hopfield networks

Participants: Pascal Chossat [Université de Nice - LJAD, Inria MathNeuro], Martin Krupa [UCA, Inria MathNeuro].

We review and extend the previous work [38] where a model was introduced for Hopfield-type neural networks, which allows for the existence of heteroclinic dynamics between steady patterns. This dynamics is a mathematical model of periodic or aperiodic switching between stored information items in the brain, in particular, in the context of sequential memory or cognitive tasks as observed in experiments. The basic question addressed in this work is whether, given a sequence of steady patterns, it is possible by applying classical learning rules to build a matrix of connections between neurons in the network, such that a heteroclinic dynamics links these patterns. It has been shown previously that the answer is positive in the case where the sequence is a so-called simple consecutive cycle. We show that on the contrary the answer is negative for a non-simple cycle: heteroclinic dynamics does still exist; however, it cannot follow the sequence of patterns from which the connectivity matrix was derived.

This work has been published in Dynamical Systems: An International Journal and is available as [21].

4.1.4. Asymptotic stability of pseudo-simple heteroclinic cycles in $\mathbb{R}^4$

Participants: Pascal Chossat [Université de Nice - LJAD, Inria MathNeuro], Olga Podvigina [Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russia].

Robust heteroclinic cycles in equivariant dynamical systems in $\mathbb{R}^4$ have been a subject of intense scientific investigation because, unlike heteroclinic cycles in $\mathbb{R}^3$, they can have an intricate geometric structure and complex asymptotic stability properties that are not yet completely understood. In a recent work [51], we have compiled an exhaustive list of finite subgroups of $O(4)$ admitting the so-called simple heteroclinic cycles, and have identified a new class which we have called pseudo-simple heteroclinic cycles. By contrast with simple heteroclinic cycles, a pseudo-simple one has at least one equilibrium with an unstable manifold which has dimension 2 due to a symmetry. Here, we analyze the dynamics of nearby trajectories and asymptotic stability of pseudo-simple heteroclinic cycles in $\mathbb{R}^4$.

This work has been published in Journal of Nonlinear Science and is available as [26].

4.1.5. The period adding and incrementing bifurcations: from rotation theory to applications

Participants: Albert Granados [Polytechnic University of Catalonia, Barcelona, Spain], Lluís Alsedà [Autonomous University of Barcelona, Spain], Martin Krupa [UCA, Inria MathNeuro].

This survey article is concerned with the study of bifurcations of piecewise-smooth maps. We review the literature in circle maps and quasi-contractions and provide paths through this literature to prove sufficient conditions for the occurrence of two types of bifurcation scenarios involving rich dynamics. The first scenario consists of the appearance of periodic orbits whose symbolic sequences and "rotation" numbers follow a Farey tree structure; the periods of the periodic orbits are given by consecutive addition. This is called the *period adding* bifurcation, and its proof relies on results for maps on the circle. In the second scenario, symbolic sequences are obtained by consecutive attachment of a given symbolic block and the periods of periodic orbits are incremented by a constant term. It is called the *period incrementing* bifurcation, in its proof relies on results for maps on the interval. We also discuss the expanding cases, as some of the partial results found in the literature also hold when these maps lose contractiveness. The higher dimensional case is also discussed by means of *quasi-contractions*. We also provide applied examples in control theory, power electronics and neuroscience where these results can be applied to obtain precise descriptions of their dynamics.

This work has been published in SIAM Review and is available as [24].

4.1.6. Inverse correlation processing by neurons with active dendrites

Participants: Tomasz Górski [UNIC, CNRS, France], Romain Veltz, Mathieu Galtier [UNIC, CNRS, France], Helissande Fragnaud [UNIC, CNRS, France], Bartosz Teleńczuk [UNIC, CNRS, France], Alain Destexhe [UNIC, CNRS, France].

In many neuron types, the dendrites contain a significant density of sodium channels and are capable of generating action potentials, but the significance and role of dendritic sodium spikes are unclear. Here, we use simplified computational models to investigate the functional effect of dendritic spikes. We found that one of the main features of neurons equipped with excitable dendrites is that the firing rate of the neuron measured at soma decreases with increasing input correlations, which is an inverse relation compared to passive dendrite and single-compartment models. We first show that in biophysical models the collision and annihilation of dendritic spikes causes an inverse dependence of firing rate on correlations. We then explore this in more detail using excitable dendrites modeled with integrate-and-fire type mechanisms. Finally, we show that the inverse correlation dependence can also be found in very simple models, where the dendrite is modeled as a discrete-state cellular automaton. We conclude that the cancellation of dendritic spikes is a generic mechanism that allows neurons to process correlations inversely compared to single-compartment models. This qualitative effect due to the presence of dendrites should have strong consequences at the network level, where networks of neurons with excitable dendrites may have fundamentally different properties than networks of point neuron models.

This work has been submitted for publication and is available as [33].

4.2. Mean field theory and stochastic processes

4.2.1. Emergence of collective phenomena in a population of neurons

Participants: Benjamin Aymard, Fabien Campillo, Romain Veltz.

In this work, we propose a new model of biological neural network, combining a two-dimensional integrate-and-fire neuron model with a deterministic model of electrical synapse, and a stochastic model of chemical synapse. We describe the dynamics of a population of neurons in interaction as a piecewise deterministic Markov process. We prove the weak convergence of the associated empirical process, as the population size tends to infinity, towards a McKean-Vlasov type process and we describe the associated PDE. We are also interested in the simulation of these dynamics, in particular by comparing "detailed" simulations of a finite population of neurons with a simulation of the system with infinite population. Benjamin Aymard has the adapted toolkit to attack these questions numerically. The mean field equations studied by Benjamin are of transport type for which numerical methods are technical. However, they are the domain of expertise of Benjamin. His postdoc is funded by the Flagship Human Brain Project.

4.2.2. Off-line numerical Bayes identification of dynamical systems for life sciences

Participants: Fabien Campillo, Vivien Rossi [CIRAD].

In this project, we develop Monte Carlo algorithms for the identification of parameters and hidden components for dynamic systems used in the life sciences. The peculiarity of these systems and they do not require online processing and they call for data of various natures and sometimes low quality. We use particle filtering techniques so that we try to improve the prediction phases using MCMC techniques.

4.2.3. On the variations of the principal eigenvalue with respect to a parameter in growth-fragmentation models

Participants: Fabien Campillo, Nicolas Champagnat [Inria, project-team TOSCA, Nancy], Coralie Fritsch [Inria, project-team TOSCA, Nancy].

We study the variations of the principal eigenvalue associated to a growth-fragmentation-death equation with respect to a parameter acting on growth and fragmentation. To this aim, we use the probabilistic individual-based interpretation of the model. We study the variations of the survival probability of the stochastic model, using a generation by generation approach. Then, making use of the link between the survival probability and the principal eigenvalue established in a previous work, we deduce the variations of the eigenvalue with respect to the parameter of the model.

This work has been published in Communications in Mathematical Sciences and is available as [18].

4.2.4. Hopf bifurcation in a nonlocal nonlinear transport equation stemming from stochastic neural dynamics

Participants: Audric Drogoul [Thales, France], Romain Veltz.

In this work, we provide three different numerical evidences for the occurrence of a Hopf bifurcation in a recently derived mean field limit of a stochastic network of excitatory spiking neurons [40], [46]. The mean field limit is a challenging nonlocal nonlinear transport equation with boundary conditions. The first evidence relies on the computation of the spectrum of the linearized equation. The second stems from the simulation of the full mean field. Finally, the last evidence comes from the simulation of the network for a large number of neurons. We provide a "recipe" to find such bifurcation which nicely complements the works in [40], [46]. This suggests in return to revisit theoretically these mean field equations from a dynamical point of view. Finally, this work shows how the noise level impacts the transition from asynchronous activity to partial synchronization in excitatory globally pulse-coupled networks.

This work has been published in Chaos and is available as [22].

4.2.5. Mathematical statistical physics applied to neural populations

Participants: Émilie Soret, Olivier Faugeras, Étienne Tanré [Inria, project-team TOSCA, Sophia-Antipolis].

This project focuses on Mean-Field descriptions or thermodynamics limits of large populations of neurons. They study a system of Stochastic Differential Equations (SDEs) which describes the evolution of membrane potential of each neuron over the time when the synaptic weights are random variables (not assumed to be independent). This setup is well suited to Émilie, who has worked during her PhD and first postdoc on mathematical statistical physics and stochastic processes. Her postdoc is funded by the Flagship Human Brain Project.

4.2.6. A numerical approach to determine mutant invasion fitness and evolutionary singular strategies

Participants: Coralie Fritsch [Inria, project-team TOSCA, Nancy], Fabien Campillo, Otso Ovaskainen [University of Helsinki, Finland].

We propose a numerical approach to study the invasion fitness of a mutant and to determine evolutionary singular strategies in evolutionary structured models in which the competitive exclusion principle holds. Our approach is based on a dual representation, which consists of the modelling of the small size mutant population by a stochastic model and the computation of its corresponding deterministic model. The use of the deterministic model greatly facilitates the numerical determination of the feasibility of invasion as well as the convergence-stability of the evolutionary singular strategy. Our approach combines standard adaptive dynamics with the link between the mutant survival criterion in the stochastic model and the sign of the eigenvalue in the corresponding deterministic model. We present our method in the context of a mass-structured individual-based chemostat model. We exploit a previously derived mathematical relationship between stochastic and deterministic representations of the mutant population in the chemostat model to derive a general numerical method for analyzing the invasion fitness in the stochastic models. Our method can be applied to the broad class of evolutionary models for which a link between the stochastic and deterministic invasion fitnesses can be established.

This work has been published in Theoretical Population Biology and is available as [23].

4.3. Neural fields theory

4.3.1. Spatiotemporal canards in neural field equations

Participants: Daniele Avitabile [University of Nottingham, UK], Mathieu Desroches, Edgar Knobloch [University of California Berkeley, USA].

Canards are special solutions to ordinary differential equations that follow invariant repelling slow manifolds for long time intervals. In realistic biophysical single-cell models, canards are responsible for several complex neural rhythms observed experimentally, but their existence and role in spatially extended systems is largely unexplored. We identify and describe a type of coherent structure in which a spatial pattern displays temporal canard behavior. Using interfacial dynamics and geometric singular perturbation theory, we classify spatiotemporal canards and give conditions for the existence of folded-saddle and folded-node canards. We find that spatiotemporal canards are robust to changes in the synaptic connectivity and firing rate. The theory correctly predicts the existence of spatiotemporal canards with octahedral symmetry in a neural field model posed on the unit sphere.

This work has been published in Physical Review E and is available as [14].

4.3.2. Standing and travelling waves in a spherical brain model: the Nunez model revisited

Participants: Sid Visser [University of Nottingham, UK], Rachel Nicks [University of Nottingham, UK], Olivier Faugeras, Stephen Coombes [University of Nottingham, UK].

The Nunez model for the generation of electroencephalogram (EEG) signals is naturally described as a neural field model on a sphere with space-dependent delays. For simplicity, dynamical realisations of this model either as a damped wave equation or an integro-differential equation, have typically been studied in idealised one dimensional or planar settings. Here we revisit the original Nunez model to specifically address the role of spherical topology on spatio-temporal pattern generation. We do this using a mixture of Turing instability analysis, symmetric bifurcation theory, centre manifold reduction and direct simulations with a bespoke numerical scheme. In particular we examine standing and travelling wave solutions using normal form computation of primary and secondary bifurcations from a steady state. Interestingly, we observe spatio-temporal patterns which have counterparts seen in the EEG patterns of both epileptic and schizophrenic brain conditions.

This work has been published in Physica D and is available as [27].

4.4. Slow-fast dynamics in Neuroscience

4.4.1. Ducks in space: from nonlinear absolute instability to noise-sustained structures in a pattern-forming system

Participants: Daniele Avitabile [University of Nottingham, UK], Mathieu Desroches, Edgar Knobloch [University of California Berkeley, USA], Martin Krupa [UCA, Inria MathNeuro].

A subcritical pattern-forming system with nonlinear advection in a bounded domain is recast as a slow-fast system in space and studied using a combination of geometric singular perturbation theory and numerical continuation. Two types of solutions describing the possible location of stationary fronts are identified, whose origin is traced to the onset of convective and absolute instability when the system is unbounded. The former are present only for non-zero upstream boundary conditions and provide a quantitative understanding of noise-sustained structures in systems of this type. The latter correspond to the onset of a global mode and are present even with zero upstream boundary conditions. The role of canard trajectories in the nonlinear transition between these states is clarified and the stability properties of the resulting spatial structures are determined. Front location in the convective regime is highly sensitive to the upstream boundary condition, and its dependence on this boundary condition is studied using a combination of numerical continuation and Monte Carlo simulations of the partial differential equation. Statistical properties of the system subjected to random or stochastic boundary conditions at the inlet are interpreted using the deterministic slow-fast spatial dynamical system.

This work has been published in Proceedings of the Royal Society A and is available as [15].

4.4.2. Canard dynamics and anticipated synchronisation in spiking models

Participants: Elif Köksal Ersöz, Mathieu Desroches, Claudio Mirasso [University of the Balearic Islands, Palma, Spain], Serafim Rodrigues [Ikerbasque, BCAM, Bilbao, Spain].

This project is on the phenomenon of anticipated synchronisation, studied theoretically in a number of models of excitable systems over the past fifteen years or so, and observed experimentally in laser systems. The idea is that when coupling two identical excitable system unidirectionally from a “master” system to a “slave” system with a delayed term of the slave’s signal in its own differential equation, one may observe that the slave reacts to an external stimulus before the master, and this is referred to as *anticipation* or *anticipated synchronisation*. Even though a number of studies have reported and analysed this effect in various systems, its main underpinning mechanisms remain elusive. In the current project, we show that in the case where the systems have an explicit slow-fast nature, then the canard regime can induce anticipation and explain its feature. Our objective is to go beyond the theoretical explanation, on which we are currently preparing an article, and to propose an electrophysiological protocol so as observe this phenomenon in real neurons. This is very much related to the PhD work of Elif Köksal Ersöz on the synchronisation properties of canard oscillators, in particular to the paper [25] (see Section 4.4.5 below). This postdoc is funded by the “tail” of the ERC Advanced Grant NerVi held by Olivier Faugeras.

4.4.3. Spike-adding in a canonical three time scale model: superslow explosion & folded-saddle canards

Participants: Mathieu Desroches, Vivien Kirk [University of Auckland, New-Zealand].

We examine the origin of complex bursting oscillations in a phenomenological ordinary differential equation model with three time scales. We show that bursting solutions in this model arise from a Hopf bifurcation followed by a sequence of spike-adding transitions, in a manner reminiscent of spike-adding transitions previously observed in systems with two time scales. However, the details of the process can be much more complex in this three-time-scale context than in two-time-scale systems. In particular, we find that spike-adding can involve canard explosions occurring on two different time scales and is associated with passage near a folded-saddle singularity. We show that the form of spike-adding transition that occurs depends on the geometry of certain singular limit systems, specifically the relative positions of the critical and superslow manifolds. We also show that, unlike the case of spike-adding in two-time-scale systems, the onset of a new spike in our model is not typically associated with a local maximum in the period of the bursting oscillation.

This work has been submitted for publication and is available as [31].

4.4.4. Piecewise-linear (PWL) canard dynamics: Simplifying singular perturbation theory in the canard regime using piecewise-linear systems

Participants: Mathieu Desroches, Soledad Fernández-García [University of Sevilla, Spain], Martin Krupa [UCA, Inria MathNeuro], Rafel Prohens [University of the Balearic Islands, Spain], Antonio Teruel [University of the Balearic Islands, Spain].

In this chapter we gathered recent results on piecewise-linear (PWL) slow-fast dynamical systems in the canard regime. By focusing on minimal systems in $\mathbb{R}^2$ (one slow and one fast variables) and $\mathbb{R}^3$ (two slow and one fast variables), we proved the existence of (maximal) canard solutions and show that the main salient features from smooth systems is preserved. We also highlighted how the PWL setup carries a level of simplification of singular perturbation theory in the canard regime, which makes it more amenable to present it to various audiences at an introductory level. Finally, we presented a PWL version of Fenichel theorems about slow manifolds, which are valid in the normally hyperbolic regime and in any dimension, which also offers a simplified framework for such persistence results.

This work has been accepted for publication as a chapter in a book titled *Nonlinear Systems; Vol. 1: Mathematical Theory and Computational Methods* (Springer, in press) and is available as [28].

4.4.5. Synchronization of weakly coupled canard oscillators

Participants: Elif Köksal Ersöz, Mathieu Desroches, Martin Krupa [UCA, Inria MathNeuro].

Synchronization has been studied extensively in the context of weakly coupled oscillators using the so-called phase response curve (PRC) which measures how a change of the phase of an oscillator is affected by a small perturbation. This approach was based upon the work of Malkin, and it has been extended to relaxation oscillators. Namely, synchronization conditions were established under the weak coupling assumption, leading to a criterion for the existence of synchronous solutions of weakly coupled relaxation oscillators. Previous analysis relies on the fact that the slow nullcline does not intersect the fast nullcline near one of its fold points, where canard solutions can arise. In the present study we use numerical continuation techniques to solve the adjoint equations and we show that synchronization properties of canard cycles are different than those of classical relaxation cycles. In particular, we highlight a new special role of the maximal canard in separating two distinct synchronization regimes: the Hopf regime and the relaxation regime. Phase plane analysis of slow-fast oscillators undergoing a canard explosion provides an explanation for this change of synchronization properties across the maximal canard.

This work has been published in Physica D and is available as [25].

4.5. Models of neural excitability

4.5.1. Modeling cortical spreading depression induced by the hyperactivity of interneurons

Participants: Mathieu Desroches, Olivier Faugeras, Martin Krupa [UCA, Inria MathNeuro], Massimo Mantegazza [IMPC, Sophia Antipolis].

Cortical spreading depression (CSD) is a wave of transient intense neuronal firing leading to a long lasting depolarization block of neuronal activity. It is a proposed pathological mechanism of migraine with aura. Some molecular/cellular mechanisms of migraine with aura and of CSD have been identified studying a rare mendelian form: familial hemiplegic migraine (FHM). FHM type 1 & 2 are caused by mutations of the CaV2.1 Ca^{2+} channel and the glial Na^{+} / K^{+} pump, respectively, leading to facilitation of CSD in mouse models mainly because of increased glutamatergic transmission/extracellular glutamate build-up. FHM type 3 mutations of the SCN1A gene, coding for the voltage gated sodium channel NaV1.1, cause gain of function of the channel and hyperexcitability of GABAergic interneurons. This leads to the counterintuitive hypothesis that intense firing of interneurons can cause CSD ignition. To test this hypothesis in silico, we developed a computational model of an E-I pair (a pyramidal cell and an interneuron), in which the coupling between the cells in not just synaptic, but takes into account also the effects of the accumulation of extracellular potassium caused by the activity of the neurons and of the synapses. In the context of this model, we show that the intense firing of the interneuron can lead to CSD. We have investigated the effect of various biophysical parameters on the transition to CSD, including the levels of glutamate or GABA, frequency of the interneuron firing and the efficacy of the KCC2 co-transporter. The key element for CSD ignition in our model was the frequency of interneuron firing and the related accumulation of extracellular potassium, which induced a depolarization block of the pyramidal cell. Our model can be used to study other types of activities in microcircuits and of couplings between excitatory and inhibitory neurons.

This work has been submitted for publication and is available as [30].

MIMESIS Team

7. New Results

7.1. Augmented Reality in Surgical Navigation

7.1.1. Organ Pose Estimation for Augmented Reality in Hepatic Surgery

Participants: Y. Adagolodjo, R. Trivisonne, H. Courtecuisse, S. Cotin

A contribution focusing on intra-operative organ pose estimation was publihsed at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2017) [19]. A novel method for semi-automatic registration of 3D deformable models using 2D shape outlines (silhouettes) extracted from a monocular camera view was introduced. The proposed framework is based on the combination of a biomechanical model of the organ with a set of projective constraints influencing the deformation of the model. To enforce convergence towards a global minimum for this ill-posed problem we interactively provide a rough (rigid) estimation of the pose. We show that our approach allows for the estimation of the non-rigid 3D pose while relying only on 2D information. The method is evaluated experimentally on a soft silicone gel model of a liver, as well as on real surgical data, providing augmented reality of the liver and the kidney using a monocular laparoscopic camera. Results show that the final elastic registration can be obtained in just a few seconds, thus remaining compatible with clinical constraints. We also evaluate the sensitivity of our approach according to both the initial alignment of the model and the silhouette length and shape.

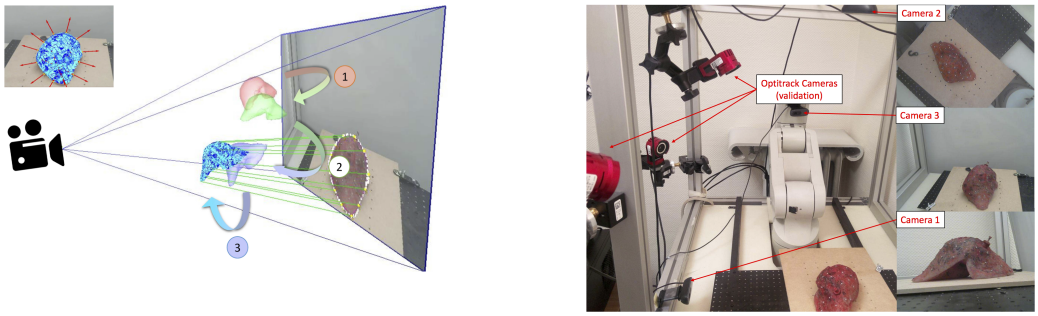


Figure 6. Left: 1) a direct simulation is applied to transform the reconstructed model obtain from the segmentation (red) in a shape close to the 3D position observed in the image (green). 2) A Rigid transformation (blue) is provided by the user to roughly align the model with the contour of the organ segmented in the image (yellow). 3) Projective constraints are applied to the biomechanical model to fit the organ contour and provide the 3D shape w.r.t. the camera position. Right: validation setup.

7.1.2. Image-driven Stochastic Estimation of Boundary Conditions

Participants: N. Haouchine, I. Peterlik, S. Cotin

A novel method was proposed in the context of image-driven stochastic simulation employed in the intra-operative navigation [25]. In the proposed approach, the boundary conditions are modeled as stochastic parameters. The method employs the reduced-order unscented Kalman filter to transform in real-time the probability distributions of the parameters, given observations extracted from intra-operative images. The method is evaluated using synthetic, phantom and real data acquired in vivo on a porcine liver. A quantitative assessment is presented and it is shown that the method significantly increases the predictive power of the biomechanical model employed by a framework implemented the augmented reality for surgical navigation.

7.2. Advanced Numerical Modeling and Simulation

7.2.1. Face-based Smoothed Finite Element Method for Real-time Simulation of Soft Tissue

Participants: A. Mendizabal, C. Paulus, R. Bessard-Duparc, I. Peterlik, S. Cotin

A method based on face-based smoothed finite element method was proposed and applied in the context of modeling of brain shift in [23]. This numerical technique has been introduced recently to overcome the overly stiff behavior of the standard FEM and to improve the solution accuracy and the convergence rate in solid mechanics problems. In this paper, a face-based smoothed finite element method (FS-FEM) using 4-node tetrahedral elements is presented. We show that in some cases, the method allows for reducing the number of degrees of freedom, while preserving the accuracy of the discretization. The method is evaluated on a simulation of a cantilever beam loaded at the free end and on a simulation of a 3D cube under traction and compression forces. Further, it is applied to the simulation of the brain shift and of the kidney's deformation. The results demonstrate that the method outperforms the standard FEM in a bending scenario and that has similar accuracy as the standard FEM in the simulations of brain shift and kidney deformation.

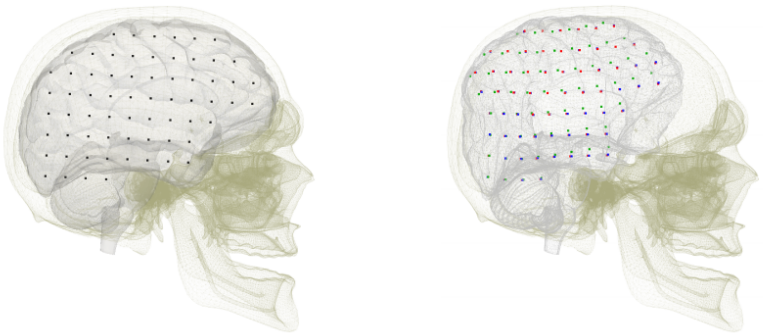


Figure 7. Grid representing tumor positions using a mesh of 7924 elements for linear FEM in blue, FS-FEM in red and non-linear FEM in green after the brain-shift. Left: rest position. Right: position after the deformation due to brain-shift.

7.2.2. Immersed Boundary Method for Real-time

Participants: C. Paulus, S. Cotin

Although the finite element method is widely used as a numerical approach in this area, it is often hindered by the need for an optimal meshing of the domain of interest. The derivation of meshes from imaging modalities such as CT or MRI can be cumbersome and time-consuming. In our contribution [24], we employed the Immersed Boundary Method (IBM) to bridge the gap between these imaging modalities and the fast simulation of soft tissue deformation on complex shapes represented by a surface mesh directly retrieved from binary images. A high resolution surface, that can be obtained from binary images using a marching cubes approach, is embedded into a hexahedral simulation grid. The details of the surface mesh are properly taken into account in the hexahedral mesh by adapting the Mirtich integration method. In addition to not requiring a dedicated meshing approach, our method results in higher accuracy for less degrees of freedom when compared to other element types. Examples on brain deformation demonstrate the potential of our method.

7.2.3. Error Control in Surgical Simulations

Participants: H. Courtecuisse, S. Cotin

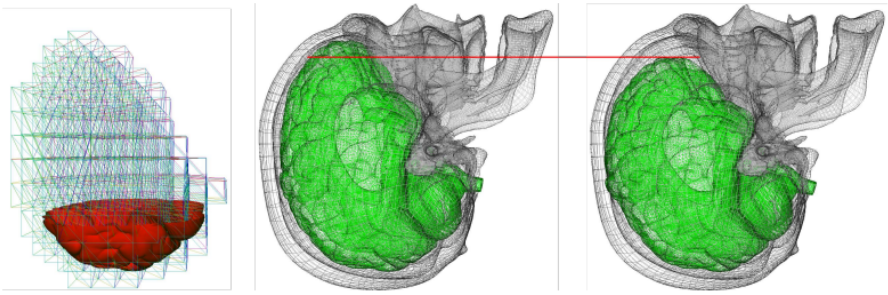


Figure 8. Simulation of brain shift using a detailed surface mesh embedded into an hexahedral grid. Boundary conditions are applied onto the exact surface, not the grid (left).

A contribution [16] presents the first real-time a posteriori error-driven adaptive finite element approach for real-time simulation and demonstrates the method on a needle insertion problem.

We use corotational elasticity and a frictional needle–tissue interaction model. The problem is solved using finite elements and the refinement strategy relies upon a hexahedron-based finite element method, combined with a posteriori error estimation driven local *h-refinement*, for simulating soft tissue deformation. We proposed to control the local and global error level in the mechanical fields (e.g. displacement or stresses) during the simulation. We show the convergence of the algorithm on academic examples, and demonstrate its practical usability on a percutaneous procedure involving needle insertion in a liver. For the latter case, we compare the force displacement curves obtained from the proposed adaptive algorithm with that obtained from a uniform refinement approach. Error control guarantees that a tolerable error level is not exceeded during the simulations. Local mesh refinement accelerates simulations. The work provides a first step to discriminate between discretization error and modeling error by providing a robust quantification of discretization error during simulations.

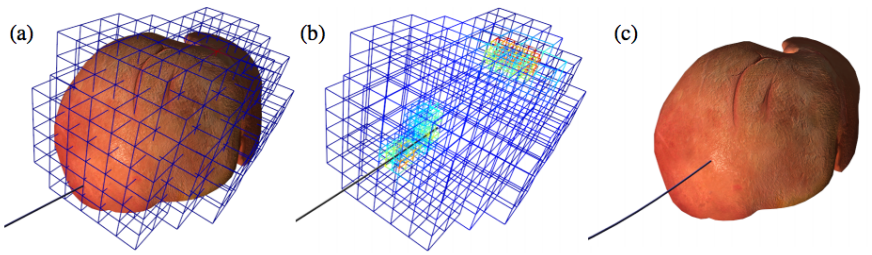


Figure 9. (a) Simulation of needle insertion in a liver; (b) Using dynamic mesh refinement scheme driven by error estimate; (c) Visual depiction. The simulation runs at 22 Hz using a PC with 4 GHz CPU.

7.3. Model-based Image Registration

7.3.1. Intraoperative Biomechanical Registration of the Liver

Participants: R. Plantefève, I. Peterlik, S. Cotin

Different aspects of model-based registration in the context of surgical navigation employing the augmented reality were analyzed in an invited contribution [17] published in the context of the attributed Prix de thèse de former Ph.D. student Rosalie Plantefève. Preoperative images such as computed tomography scans or magnetic resonance imaging contain lots of valuable information that are not easily available for surgeons during an operation. To help the clinicians better target the structures of interest during an intervention, many registration methods that align preoperative images onto the intra-operative view of the organs have been developed. For important organ deformation, biomechanical model-based registration has proven to be a method of choice. Using an existing model-based registration algorithm for laparoscopic liver surgery we investigated the influence of the heterogeneity of the liver on the registration result. It was found that the use of an heterogeneous model does not improve significantly the registration result but increases the computation time necessary to perform the registration.

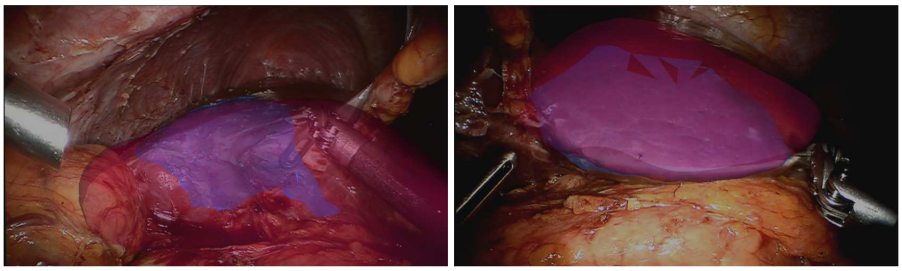


Figure 10. Registration results on in vivo data on two different views of a human liver. The registered mesh is shown in red while the partial reconstructed patch is depicted in blue.

7.3.2. Registration of Cell Nuclei in Cell Microscopy

Participants: I. Peterlik

A contribution *Registration of Cell Nuclei in 2D Live Cell Microscopy* was published in a collaboration with Centre of Biomedical Image Analysis at Masaryk University, Czech Republic [18]. The analysis of the pure motion of sub-nuclear structures without influence of the cell nucleus motion and deformation is essential in live cell imaging. We proposed a 2D contour-based image registration approach for compensation of nucleus motion and deformation in fluorescence microscopy time-lapse sequences. The proposed approach extends our previous approach which uses a static elasticity model to register cell images. Compared to that scheme, the new approach employs a dynamic elasticity model for forward simulation of nucleus motion and deformation based on the motion of its contours. The contour matching process is embedded as a constraint into the system of equations describing the elastic behavior of the nucleus. This results in better performance in terms of the registration accuracy. Our approach was successfully applied to real live cell microscopy image sequences of different types of cells including image data that was specifically designed and acquired for evaluation of cell image registration methods.

7.4. Reconstruction of Geometries from Images

7.4.1. Automatic Skeletonization of Vascular Trees in Pre-operative CT Images

Participants: R. Plantefève, I. Peterlik

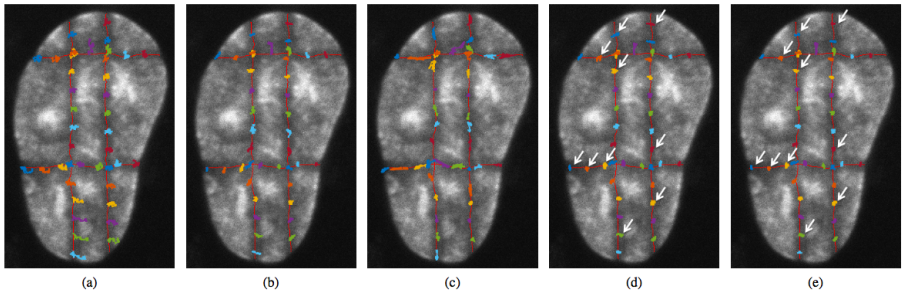


Figure 11. Tracks of line features overlaid with the first image of the sequence. The tracks represent the motion of the points of the line features sampled with 30 pixel interval for better visibility. The tracks are shown for (a) unregistered data, (b) after registration with the contour-based approach [19], (c) after registration with the intensity-based approach [9], (d) after registration with the static version of our approach, and (e) after registration with the proposed dynamic approach. White arrows indicate tracks with the most visible difference between (d) and (e).

An algorithm of an automatic skeletonization of vascularization based on Dijkstra minimum-cost spanning tree was published in [27]. The result is an extension of an existing graph-based method where the vascular topology is constructed by computation of shortest paths in a minimum-cost spanning tree obtained from binary mask of the vascularization. We suppose that the binary mask is extracted from a 3D CT image using automatic segmentation and thus suffers from important artifacts and noise. When compared to the original algorithm, the proposed method (i) employs a new weighting measure which results in smoothing of extracted topology and (ii) introduces a set of tests based on various geometric criteria which are executed in order to detect and remove spurious branches. The method is evaluated on vascular trees extracted from abdominal contrast-enhanced CT scans and MR images. The method is quantitatively compared to the original version of the algorithm showing the importance of proposed modifications. Since the branch testing depends on parameters, the para-metric study of the proposed method is presented in order to identify the optimal parametrization.

7.4.2. Template-based Recovery of Elastic Shapes from Monocular Video

Participants: N. Haouchine, S. Cotin

A method of template-based 3D recovery of elastic shapes using Lagrange multiplied was presented at a top computed-vision conference [21]. By exploiting the object's elasticity, in contrast to isometric methods that use inextensibility constraints, a large range of deformations can be handled. Our method is expressed as a saddle point problem using Lagrangian multipliers resulting in a linear system which unifies both mechanical and optical constraints and integrates Dirichlet boundary conditions, whether they are fixed or free. We experimentally show that no prior knowledge on material properties is needed, which exhibit the generic usability of our method with elastic and inelastic objects with different kinds of materials. Comparison with existing techniques are conducted on synthetic and real elastic objects with strains ranging from 25% to 130% resulting to low errors.

7.5. Simulation for Intra-operative Rehearsal

Participants: N. Haouchine, F. Roy, S. Cotin

DejaVu, a novel surgical simulation approach for intra-operative surgical gesture rehearsal was published in [22] in collaboration with UCL London. With DejaVu we aim at bridging the gap between pre-operative surgical simulation and crucial but not yet robust intra-operative surgical augmented reality. By exploiting

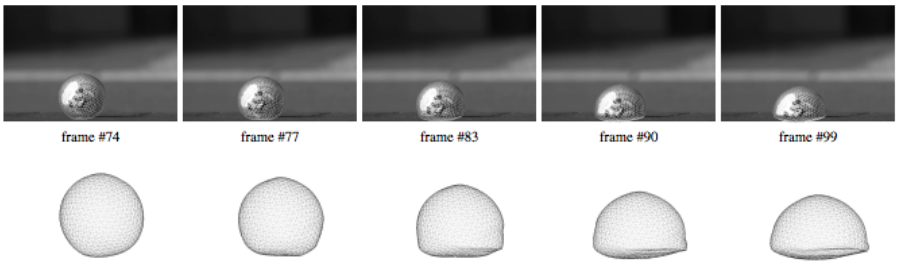


Figure 12. The proposed method illustrated on an example with a soft ball colliding the ground in slow motion. No prior knowledge of material properties in considered. The spherical volume model is composed of 512 linear P1 tetrahedral elements. The recovery and augmentation is performed in real-time at 25 FPS.

intra-operative images we produce a simulation that faithfully matches the actual procedure without visual discrepancies and with an underlying physical modeling that performs real-time deformation of organs and surrounding tissues, surgeons can interact with the targeted organs through grasping, pulling or cutting to immediately rehearse their next gesture. We present results on different in vivo surgical procedures and demonstrate the feasibility of practical use of our system.

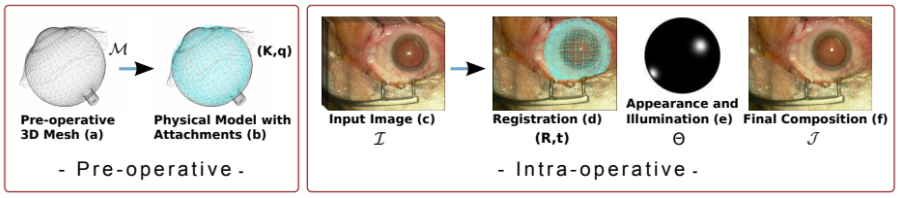


Figure 13. Schematic illustration of DejaVu Simulation. (a) preoperative model is built from tomographic images; (b) material law, tissue properties and attachments, constitute the physical model; (c) an intra-operative image is selected; (d) 3D/2D registration is performed between the physical model in (b) and the selected frame in (c); (e) appearance and illumination are estimated corresponding to specular and diffuse components and light position; (f) the final composition is build to enable surgical gesture rehearsal.

MNEMOSYNE Project-Team

7. New Results

7.1. Overview

This year we have explored two main cortico-basal loops of cerebral architecture, the limbic and motor loops, and their associated memory mechanisms. The limbic loop (*cf.* § 7.2) concerns the taking into account of the emotional and motivational aspects by the respondant and operant conditioning and their relations with the semantic and episodic memories. The motor loop (*cf.* § 7.3) considers the evolution of sensorimotor learning, from goal-directed behaviors to habitual behaviors.

We have also worked on the systemic integration of our models (*cf.* § 7.4), raising the question of the conditions of autonomous learning and certain global characteristics such as neuromodulation.

Finally, we study the links between our bio-inspired modeling work and Machine Learning (*cf.* § 7.5), revisiting this latter domain in the light of the principles highlighted by our models.

7.2. The limbic loop

We explore the limbic loop by studying a series of neural mechanisms that propose how respondant conditioning results from interactions between the amygdala, the nucleus accumbens and the limbic pole of the frontal cortex. In our models, this learning is also fed by exchanges with the hippocampus (episodic memory) [6] and the sensory cortex (semantic memory). We have also addressed the difficult question of the articulation between the respondant and operant conditioning in particular in the nucleus accumbens.

Also in connection with this loop, we studied the dynamics of dopamine release in the midbrain, considered to play an essential role in the coding of the prediction error. This model [12] developed in the framework of our collaboration with India (*cf.* § 9.3) proposes to introduce into the classical circuit, new actors (such as the pedunculopontine tegmental nucleus in the brainstem) and new functions (dissociation of amplitude and timing of the reward), that we will seek to corroborate in the future.

7.3. The motor loop

The nervous system structures involved in decision making constitute a circuit formed by the basal ganglia, the cortex, the thalamus and their numerous interconnections. This circuit can be described as a set of loops operating in parallel and interacting at different points. The decisions and therefore the actions of an individual emerge from the interactions between these loops and the plasticity of their connections. These emerging behaviors and arising learning processes are addressed through a closed-loop approach in which the theoretical model is in constant interaction with the environment of the task. To this end, neural modeling and dedicated analysis software tools were developed in the laboratory, at the level of the neuronal circuit.

7.4. Systemic integration

Systemic integration promotes the idea of developing large models that associate several cortico-basal loops and even other cerebral structures and more generally takes into account the influence of the body on this network [19]. This requires to propose a global picture for the organization and functional association between all these elements [18] and to analyze its consequences from a representational point of view [1] and also concerning autonomous learning [7].

It also requires to evaluate the properties of such systems from their interactions with the body and the environment, as we have done this year using the VirtualEnaction platform.

7.5. Machine Learning

In this section, we report on some neuronal adaptive mechanisms, that we develop at the frontier between Machine Learning and Computational Neuroscience. Our goal is to consider and adapt models in Machine Learning for their integration in a bio-inspired framework. We were interested this year in three paradigms of computation.

The first paradigm concerns the manipulation of temporal sequences. In a perspective of better understanding how the brain learns structured sequences we work on a model on syntax acquisition and Human-Robot Interaction using the Reservoir Computing framework (using random recurrent networks) [24], [15], [17] with our collaborators at the University of Hamburg (*cf.* § 9.3). A syntactic re-analysis system [15], which corrects syntax errors in speech recognition hypotheses, was built in order to enhance vocal Human-Robot Interaction and to enhance the previously developped model [40]. Additionally, the ability to deal with several languages (from different language families) of this later model of sentence parsing [40] was evaluated. We showed that it can successfully learn to parse sentences related to home scenarios in fifteen languages originating from Europe and Asia [24]. In a different perspective, in order to try to overcome word misrecognition at a more basic level, we tested whether the same architecture was able to process directly phonemes instead of grammatical constructions [17]. Applied on a small corpus, we see that the model has similar performance.

In an industrial application for the representation of electrical diagrams (*cf.* § 8.1), we also study how recurrent layered models can be trained to run through these schemes for prediction and sequence representation tasks [10].

The second paradigm concerns the extraction of characteristics and the use of hierarchical networks, as in the case of deep networks. An industrial application (*cf.* § 9.2) allows us to revisit these models to make them more easily usable in constrained frameworks, for example with limited size corpuses, and more interpretable introducing a new notion of prototypes and exploring the capability to learn the network architecture itself (using shortcuts) [11]. In order to push the sate of the art, the next step is going to consider not only feed-forward but also recurrent architecture, and to this end neural network recurrent weight estimation through backward tuning has been revisited [21].

The third paradigm is about spatial computation. We have designed a graphical method originating from the computer graphics domain that is used for the arbitrary and intuitive placement of cells over a two-dimensional manifold. Using a bitmap image as input, where the color indicates the identity of the different structures and the alpha channel indicates the local cell density, this method guarantees a discrete distribution of cell position respecting the local density function. This method scales to any number of cells, allows to specify several different structures at once with arbitrary shapes and provides a scalable and versatile alternative to the more classical assumption of a uniform non-spatial distribution. This preliminary work will be used in the design of a new class of model where explicit topography allows to connect structure according to known pathways.

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Axel Hutt, Francesco Giovannini, Mélanie Aussel, Ivan Kotiuchi.
In collaboration with Radu Ranta (university of Lorraine), Beate Knauer and Motoharu Yoshida (Ruhr university) and LieJune Shiau (university of Houston)

7.1.1. Memory and anesthesia

7.1.1.1. Modeling effects of propofol anesthesia

Neural oscillations are thought to be correlated with the execution of cognitive functions. Indeed, gamma oscillations are often recorded in functionally-coupled brain regions for cooperation during memory tasks, and this rhythmic behavior is thought to result from synaptic GABAergic interactions between interneurons. Interestingly, GABAergic synaptic and extrasynaptic receptors have been shown to be the preferred target of the most commonly used anesthetic agents. We presented a in-depth computational study [0] [1] of the action of anesthesia on neural oscillations by introducing a new mathematical model which takes into account the four main effects of the anesthetic agent propofol on GABAergic hippocampal interneurons. These are: the action on synaptic $GABA_A$ receptors, which includes an amplification and an extension of the duration of the synaptic currents, as well as an increase in current baseline, and the action on extrasynaptic $GABA_A$ receptors mediating a tonic inhibitory current. Our results indicate that propofol-mediated tonic inhibition contributes to an unexpected enhancement of synchronization in the activity of a network of hippocampal interneurons. This enhanced synchronization could provide a possible mechanism supporting the occurrence of intraoperative awareness, explicit memory formation, and even paradoxical excitation under general anesthesia, by transiently facilitating the communication between brain structures which should supposedly be not allowed to do so when anesthetized.

7.1.1.2. Stability Analysis in a model of hippocampal place cells

Ring networks, a particular form of Hopfield neural networks, can be used to model the activity of place cells, a type of cells in the hippocampus that are involved in the building and memorization of a cognitive map of one's environment. The behavior of these models is highly dependent on their recurrent synaptic connectivity matrix and on individual neurons' activation function, which must be chosen appropriately to obtain physiologically meaningful conclusions. In [4], we proposed several simpler ways to adjust this synaptic connectivity matrix compared to existing literature so as to achieve stability in a ring attractor network with a piece-wise affine activation functions, and we link these results to the possible stable states the network can converge to.

7.1.1.3. Modeling of the hippocampal formation over the sleep-wake cycle :

The hippocampus can exhibit different oscillatory rhythms within the sleep-wake cycle, each of them being involved in cognitive processes. For example, theta-nested gamma oscillations, consisting of the coupling of theta (4-12Hz) and gamma (40-100Hz) rhythms, are produced during wakefulness and are associated with spatial navigation tasks, whereas Sharp-Wave-Ripple (SWR) complexes, consisting of fast (140-200Hz) oscillatory events occurring at low ($\leq$ 0.5Hz) frequencies, are produced during slow-wave sleep and play an important role in memory consolidation. The mechanisms underlying the generation and switch between each of these rhythms is not yet fully understood, but Acetylcholine is thought to play a key role in it.

[0] F. Giovannini and L. Buhry, Tonic inhibition mediates a synchronization enhancement during propofol anesthesia in a network of hippocampal interneurons: a modeling study Journal of computational neuroscience (Submitted) 2017

In an article in preparation, we propose a computational model of the hippocampal formation based on a realistic topology and synaptic connectivity, influenced by the changing concentration of Acetylcholine between wakefulness and sleep. By using a detailed estimation of intracerebral recordings, we show that this model is able to reproduce both the theta-nested gamma oscillations that are seen in awake brains and the sharp-wave ripple complexes that appear during slow-wave sleep. The results of our simulations support the idea that the functional connectivity of the hippocampus is a key factor in controlling its rhythms.

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Tamara Tošić, Cecilia Lindig-León, Romain Orhand, Sébastien Rimbert, Oleksii Avilov, Rahaf Al-Chwa.
In collaboration with Stéphanie Fleck (Univ. Lorraine)

7.2.1. Motor system

In collaboration with Stéphanie Fleck (Univ. Lorraine)
Kinesthetic motor imagery (KMI) tasks induce brain oscillations over specific regions of the primary motor cortex within the contralateral hemisphere of the body part involved in the process. This activity can be measured through the analysis of electroencephalographic (EEG) recordings and is particularly interesting for Brain-Computer Interface (BCI) applications.

7.2.1.1. Continuous and discrete

In most BCI experimental paradigms based on Motor Imagery (MI), subjects perform continuous motor imagery (CMI), i.e., a repetitive and prolonged intention of movement, for a few seconds. To improve efficiency such as detecting faster a motor imagery and thus avoid fatigue and boredom, we proposed to show the difference between discrete motor imagery (DMI), i.e., a single short MI, and CMI. The results of the experiment involving 13 healthy subjects suggest that DMI generates a robust post-MI event-related synchronization (ERS). Moreover event-related desynchronization (ERD) produced by DMI seems less variable in certain cases compared to CMI [10], [12]. We showed the difference, in term of classification, between a DMI and a CMI. The results of the experiment involving 16 healthy subjects show that a BCI based on DMI is as effective as a BCI based on CMI and could be used to allow a faster detection [6].

7.2.1.2. Profiling

The most common approach for classification consists of analyzing the signal during the course of the motor task within a frequency range including the alpha band, which attempts to detect the Event-Related Desynchronization (ERD) characteristics of the physiological phenomenon. However, to discriminate right-hand KMI and left-hand KMI, this scheme can lead to poor results on subjects for which the lateralization is not significant enough. To solve this problem, we proposed to analyze the signal at the end of the motor imagery within a higher frequency range, which contains the Event-Related Synchronization (ERS). We showed that 6 out of 15 subjects have a higher classification rate after the KMI than during the KMI, due to a higher lateralization during this period. Thus, for this population we obtained a significant improvement of 13% in classification taking into account the users lateralization profile [9].

7.2.1.3. Combined motor imageries

Combined motor imageries can be detected to deliver more commands in a Brain-Computer Interface for controlling a robotic arm. Nevertheless only a few systems use more than three motor imageries: right hand, left hand and feet. Combining them allows to get four additional commands. We presented an electrophysiological study to show that i) simple motor imageries have mainly an electrical modulation over the cortical area related the body part involved in the imagined movement and that ii) combined motor imageries reflect a superposition of the electrical activity of simple motor imageries. A shrinkage linear discriminant analysis has been used to test as a first step how a resting state and seven motor imageries can be detected. 11 healthy subjects participated in the experiment for which an intuitive assignment has been done to associate motor imageries and movements of the robotic arm with 7 degrees of freedom [2], [5].

7.2.1.4. Anesthesia

Each year, several million of general anesthesia are realized in France. A recent study shows that, between 0.1-0.2 % of patients are victims of intraoperative awareness. This kind of awakening could cause post-traumatic syndromes for the patient. Unfortunately, today, no monitoring system is able to avoid the intraoperative awareness phenomenon. Interestingly, if there is no subject movement due to curare, an electroencephalographical study of the motor cortex can help to detect an intention of movement. The dynamic study of motor cerebral activity during general anesthesia is essential if we want to create a brain-computer interface adapted to the detection of intraoperative awareness. We wrote a clinical protocol to allow EEG data recording during general anesthesia with propofol. Then, the development of temporal analysis specific methods allows us to quantify patterns of desynchronization and synchronization phases observed in delta, alpha and beta frequency bands to prevent intraoperative awareness [8].

PARIETAL Project-Team

7. New Results

7.1. Joint prediction of multiple scores captures better individual traits from brain images

To probe individual variations in brain organization, population imaging relates features of brain images to rich descriptions of the subjects such as genetic information or behavioral and clinical assessments. Capturing common trends across these measurements is important: they jointly characterize the disease status of patient groups. In particular, mapping imaging features to behavioral scores with predictive models opens the way toward more precise diagnosis. Here we propose to jointly predict all the dimensions (behavioral scores) that make up the individual profiles, using so-called multi-output models. This approach often boosts prediction accuracy by capturing latent shared information across scores. We demonstrate the efficiency of multi-output models on two independent resting-state fMRI datasets targeting different brain disorders (Alzheimer's Disease and schizophrenia). Furthermore, the model with joint prediction generalizes much better to a new cohort: a model learned on one study is more accurately transferred to an independent one. Finally, we show how multi-output models can easily be extended to multi-modal settings, combining heterogeneous data sources for a better overall accuracy.

More information can be found in Fig. 3 in [46].

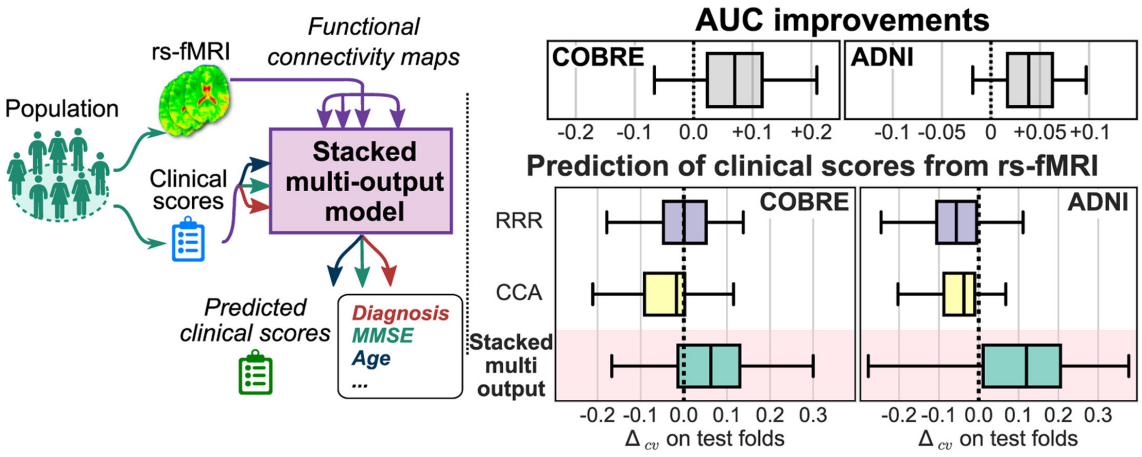


Figure 3. Joint prediction of multiple scores captures better individual traits from brain images

7.2. Population-shrinkage of covariance to estimate better brain functional connectivity

Brain functional connectivity, obtained from functional Magnetic Resonance Imaging at rest (r-fMRI), reflects inter-subject variations in behavior and characterizes neuropathologies. It is captured by the covariance matrix between time series of remote brain regions. With noisy and short time series as in r-fMRI, covariance estimation calls for penalization, and shrinkage approaches are popular. Here we introduce a new covariance

estimator based on a non-isotropic shrinkage that integrates prior knowledge of the covariance distribution over a large population. The estimator performs shrinkage tailored to the Riemannian geometry of symmetric positive definite matrices, coupled with a probabilistic modeling of the subject and population covariance distributions. Experiments on a large-scale dataset show that such estimators resolve better intra-and inter-subject functional connectivities compared existing co-variance estimates. We also demonstrate that the estimator improves the relationship across subjects between their functional-connectivity measures and their behavioral assessments. More information can be found in Fig. 4 in [47].

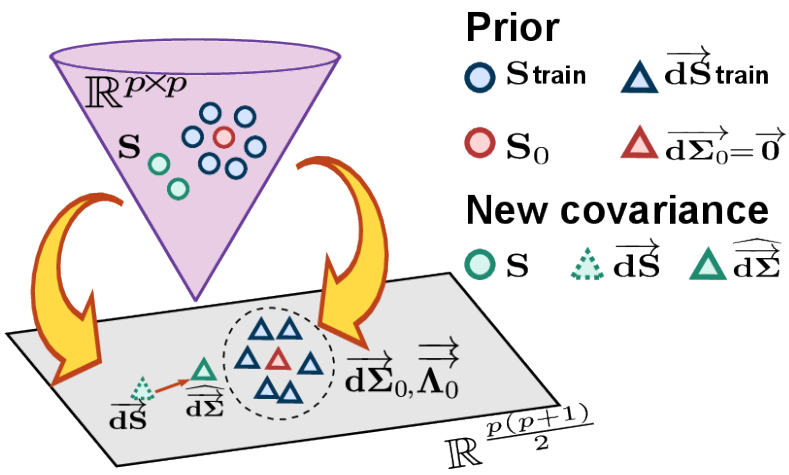


Figure 4. (a) Shrunk embedding estimation workflow: the empirical covariance is estimated from r-fMRI time-series; it is projected onto a tangent space built from a prior population; the embedding is then shrunk towards the prior $(\overrightarrow{\mathbf{d}\Sigma_0}, \overrightarrow{\Lambda}_0)$. *(b) Principle of tangent embedding shrinkage towards population distribution.*

7.3. Fast Regularized Ensembles of Models

Brain decoding relates behavior to brain activity through predictive models. These are also used to identify brain regions involved in the cognitive operations related to the observed behavior. Training such multivariate models is a high-dimensional statistical problem that calls for suitable priors. State of the art priors –eg small total-variation– enforce spatial structure on the maps to stabilize them and improve prediction. However, they come with a hefty computational cost. We build upon very fast dimension reduction with spatial structure and model ensembling to achieve decoders that are fast on large datasets and increase the stability of the predictions and the maps. Our approach, fast regularized ensemble of models (FReM), includes an implicit spatial regularization by using a voxel grouping with a fast clustering algorithm. In addition, it aggregates different estimators obtained across splits of a cross-validation loop, each time keeping the best possible model. Experiments on a large number of brain imaging datasets show that our combination of voxel clustering and model ensembling improves decoding maps stability and reduces the variance of prediction accuracy. Importantly, our method requires less samples than state-of-the-art methods to achieve a given level of prediction accuracy. Finally, FreM is highly parallelizable, and has lower computation cost than other spatially-regularized methods.

More information can be found in Fig. 5 in [23].

7.4. time decoding

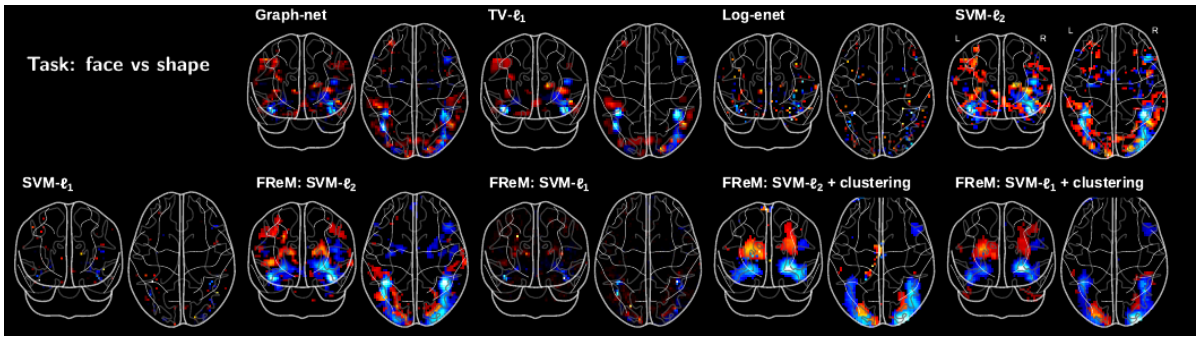


Figure 5. Qualitative comparison of decoder weight maps: Weight maps for different discriminative tasks on the HCP dataset. The maps are thresholded at the 99 percentile for visualization purposes. These correspond to a face-recognition task. The weight maps obtained with TV-L1 and FReM methods with clustering outline the organization of the functional areas of the visual mosaic, such as: primary visual areas, lateral occipital complex, the face and place specific regions in the fusiform gyrus.

Most current functional Magnetic Resonance Imaging (fMRI) decoding analyses rely on statistical summaries of the data resulting from a deconvolution approach: each stimulation event is associated with a brain response. This standard approach leads to simple learning procedures, yet it is ill-suited for decoding events with short inter-stimulus intervals. In order to overcome this issue, we propose a novel framework that separates the spatial and temporal components of the prediction by decoding the fMRI time-series continuously, i.e. scan-by-scan. The stimulation events can then be identified through a deconvolution of the reconstructed time series. We show that this model performs as well as or better than standard approaches across several datasets, most notably in regimes with small inter-stimuli intervals (3 to 5s), while also offering predictions that are highly interpretable in the time domain. This opens the way toward analyzing datasets not normally thought of as suitable for decoding and makes it possible to run decoding on studies with reduced scan time.

More information can be found in Fig. 6 in [28].

7.5. Hierarchical Region-Network Sparsity for High-Dimensional Inference in Brain Imaging

Structured sparsity penalization has recently improved statistical models applied to high-dimensional data in various domains. As an extension to medical imaging, the present work incorporates priors on network hierarchies of brain regions into logistic-regression to distinguish neural activity effects. These priors bridge two separately studied levels of brain architecture: functional segregation into regions and functional integration by networks. Hierarchical region-network priors are shown to better classify and recover 18 psychological tasks than other sparse estimators. Varying the relative importance of region and network structure within the hierarchical tree penalty captured complementary aspects of the neural activity patterns. Local and global priors of neurobiological knowledge are thus demonstrated to offer advantages in generalization performance, sample complexity, and domain interpretability.

More information can be found in Fig. 7 in [48].

7.6. Cross-validation failure: small sample sizes lead to large error bars

Predictive models ground many state-of-the-art developments in statistical brain image analysis: decoding, MVPA, searchlight, or extraction of biomarkers. The principled approach to establish their validity and

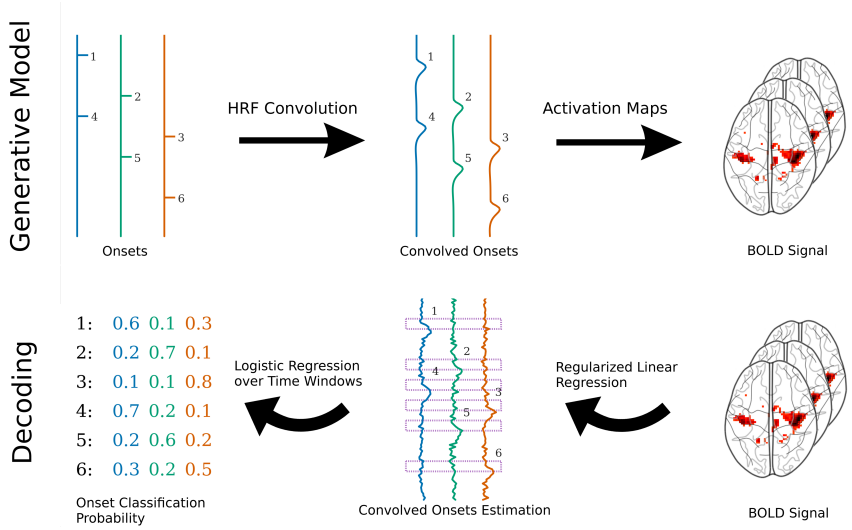


Figure 6. Schema of the time-domain decoding model. Straight arrows represent generative steps, while curved ones represent estimation steps.

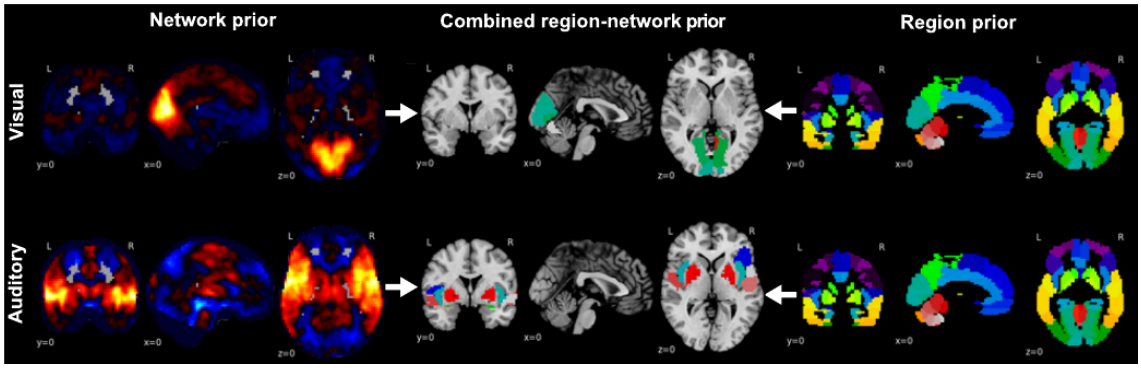


Figure 7. Building blocks of the hierarchical region-network tree. Displays the a-priori neurobiological knowledge introduced into the classification problem by hierarchical structured sparsity. Left: Continuous, partially overlapping brain network priors (hot-colored) accommodate the functional integration perspective of brain organization. Right: Discrete, non-overlapping brain region priors (single-colored) accommodate the functional segregation perspective. Middle: These two types of predefined voxel groups are incorporated into a joint hierarchical prior of parent networks with their descending region child nodes. Top to bottom: Two exemplary region-network priors are shown, including the early cortices that process visual and sound information from the environment.

usefulness is cross-validation, testing prediction on unseen data. Here, we raise awareness on error bars of cross-validation, which are often underestimated. Simple experiments show that sample sizes of many neuroimaging studies inherently lead to large error bars, eg $\pm 10\%$ for 100 samples. The standard error across folds strongly underestimates them. These large error bars compromise the reliability of conclusions drawn with predictive models, such as biomarkers or methods developments where, unlike with cognitive neuroimaging MVPA approaches, more samples cannot be acquired by repeating the experiment across many subjects. Solutions to increase sample size must be investigated, tackling possible increases in heterogeneity of the data.

More information can be found in Fig. 8 in [33].

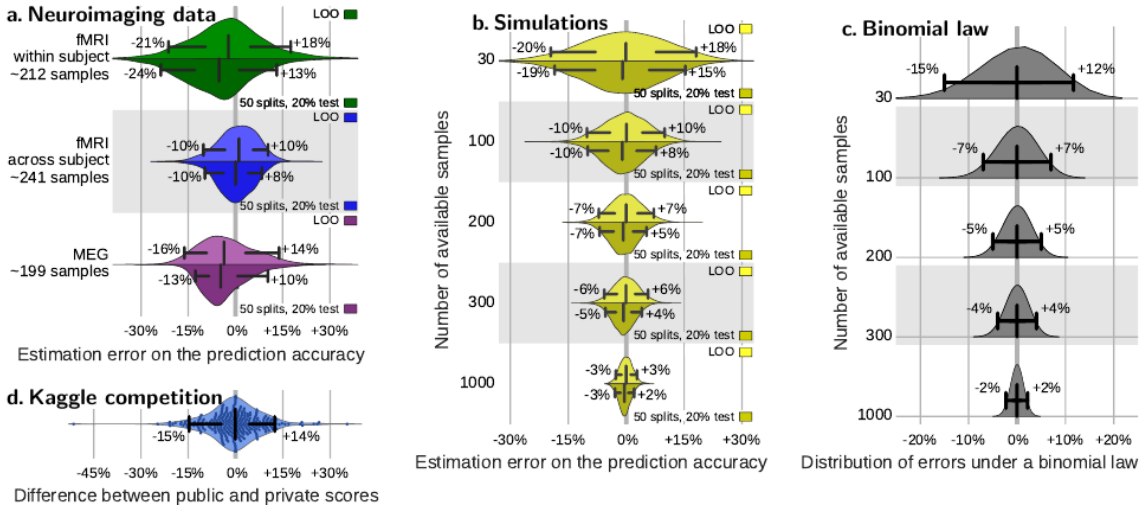


Figure 8. Cross-validation errors. a – Distribution of errors between the prediction accuracy as assessed via cross-validation (average across folds) and as measured on a large independent test set for different types of neuroimaging data. b – Distribution of errors between the prediction accuracy as assessed via cross-validation on data of various sample sizes and as measured on 10 000 new data points for simple simulations. c – Distribution of errors as given by a binomial law: difference between the observed prediction error and the population value of the error, p = 75%, for different sample sizes. d – Discrepancies between private and public score. Each dot represents the difference between the accuracy of a method on the public test data and the private one. The scores are retrieved from http://www.kaggle.com/c/mlsp-2014-mri, in which 144 subjects were used total, 86 for training predictive model, 30 for the public test set, and 28 for the private test set. The bar and whiskers indicate the median and the 5th and 95th percentile. Measures on cross-validation (a and b) are reported for two reasonable choices of cross-validation strategy: leave one out (leave one run out or leave one subject out in data with multiple runs or subjects), or 50-times repeated splitting of 20% of the data.

7.7. Autoreject: Automated artifact rejection for MEG and EEG data

We present an automated algorithm for unified rejection and repair of bad trials in magnetoencephalography (MEG) and electroencephalography (EEG) signals. Our method capitalizes on cross-validation in conjunction with a robust evaluation metric to estimate the optimal peak-to-peak threshold – a quantity commonly used for identifying bad trials in M/EEG. This approach is then extended to a more sophisticated algorithm which estimates this threshold for each sensor yielding trial-wise bad sensors. Depending on the number of bad sensors, the trial is then repaired by interpolation or by excluding it from subsequent analysis. All

steps of the algorithm are fully automated thus lending itself to the name Autoreject. In order to assess the practical significance of the algorithm, we conducted extensive validation and comparisons with state-of-the-art methods on four public datasets containing MEG and EEG recordings from more than 200 subjects. The comparisons include purely qualitative efforts as well as quantitatively benchmarking against human supervised and semi-automated preprocessing pipelines. The algorithm allowed us to automate the preprocessing of MEG data from the Human Connectome Project (HCP) going up to the computation of the evoked responses. The automated nature of our method minimizes the burden of human inspection, hence supporting scalability and reliability demanded by data analysis in modern neuroscience.

More information can be found in Fig. 9 and in [24].

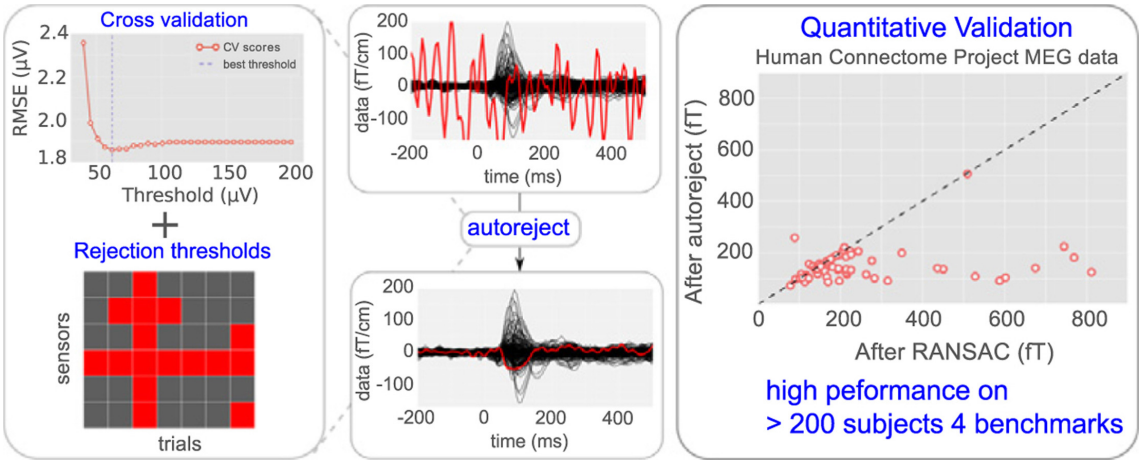


Figure 9. Autoreject: Automated artifact rejection for MEG and EEG data.

7.8. Learning Neural Representations of Human Cognition across Many fMRI Studies

Cognitive neuroscience is enjoying rapid increase in extensive public brain-imaging datasets. It opens the door to large-scale statistical models. Finding a unified perspective for all available data calls for scalable and automated solutions to an old challenge: how to aggregate heterogeneous information on brain function into a universal cognitive system that relates mental operations/cognitive processes/psychological tasks to brain networks? We cast this challenge in a machine-learning approach to predict conditions from statistical brain maps across different studies. For this, we leverage multi-task learning and multi-scale dimension reduction to learn low-dimensional representations of brain images that carry cognitive information and can be robustly associated with psychological stimuli. Our multi-dataset classification model achieves the best prediction performance on several large reference datasets, compared to models without cognitive-aware low-dimension representations; it brings a substantial performance boost to the analysis of small datasets, and can be introspected to identify universal template cognitive concepts.

More information can be found in Fig. 10 in [45].

7.9. SPARKLING: Novel Non-Cartesian Sampling Schemes for Accelerated 2D Anatomical Imaging at 7T Using Compressed Sensing

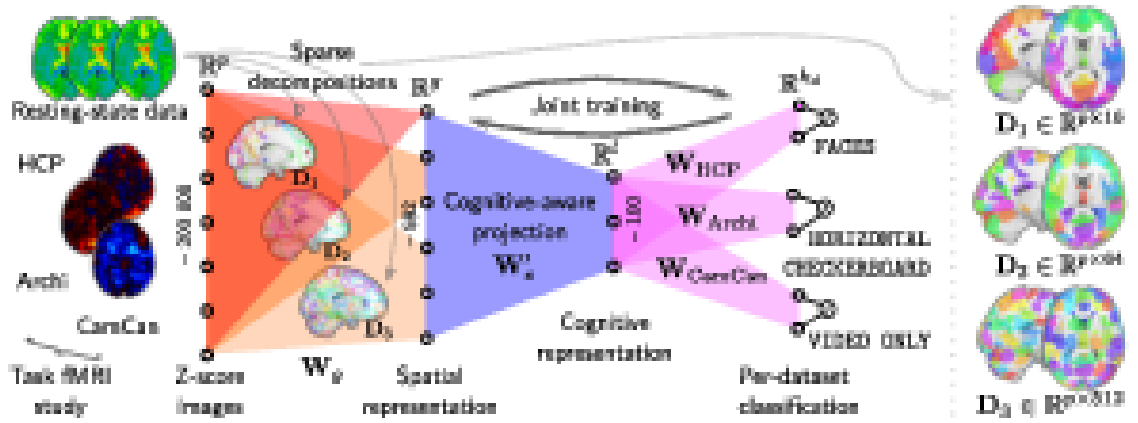


Figure 10. ***Model architecture: Three-layer multi-dataset classification.*** *The first layer (orange) is learned from data acquired outside of cognitive experiments and captures a spatially coherent signal at multiple scales, the second layer (blue) embeds these representations in a space common to all datasets, from which the conditions are predicted (pink) from multinomial models.*

We have presented for the first time the implementation of non-Cartesian trajectories on a 7T scanner for 2D anatomical imaging. The proposed SPARKLING curves (Segmented Projection Algorithm for Random K-space sampLING) are a new type of non-Cartesian segmented sampling trajectories which allow fast and efficient coverage of the k-space according to a chosen variable density. To demonstrate their potential, a high-resolution (0.4×0.4×3.0mm 3) T2*-weighted image was acquired with an 8-fold undersampled SPARKLING trajectory. Images were reconstructed using non-linear iterative reconstructions derived from the Compressed Sensing theory.

More information can be found in Fig. 11 in [43], [42].

7.10. A Projection Method on Measures Sets

We consider the problem of projecting a probability measure π on a set $\mathcal{M}_N$ of Radon measures. The projection is defined as a solution of the following variational problem:

$$\inf_{\mu\in\mathcal{M}_N} \|h * (\mu - \pi)\|_2^2,$$

where $h \in L^2(\Omega)$ is a kernel, $\Omega \subset \mathbb{R}^d$ and denotes the convolution operator. To motivate and illustrate our study, we show that this problem arises naturally in various practical image rendering problems such as stippling (representing an image with N dots) or continuous line drawing (representing an image with a continuous line). We provide a necessary and sufficient condition on the sequence $(\mathcal{M}_N)_{N\in\mathbb{N}}$ that ensures weak convergence of the projections $(\mu_N^*)_{N\in\mathbb{N}}$ to π. We then provide a numerical algorithm to solve a discretized version of the problem and show several illustrations related to computer-assisted synthesis of artistic paintings/drawings.

More information can be found in [19].

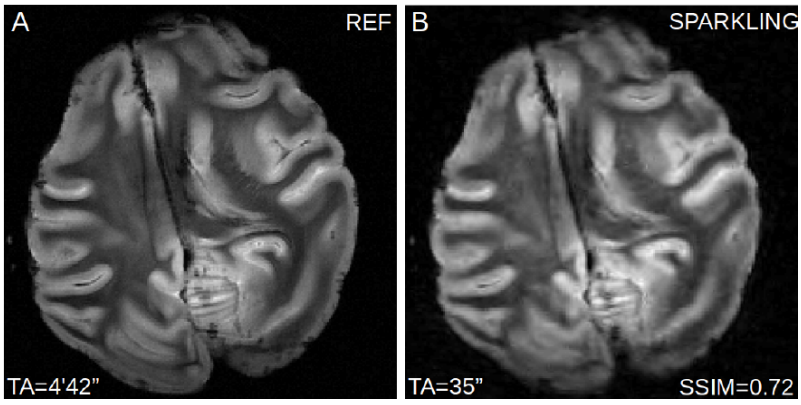


Figure 11. T2-weighted Gradient-Echo transversal slice of the ex vivo baboon brain is displayed for A) the full Cartesian reference sampling lasting 4 minutes and 42 seconds and B) the presented 8-fold undersampled SPARKLING trajectories lasting only 35 seconds. A good visual quality of the subsampled reconstruction was observed with the preservation of major folded patterns. The structural similarity score between the SPARKLING reconstructions and the reference was also very satisfactory.*

VISAGES Project-Team

7. New Results

7.1. Research axis 1: Medical Image Computing in Neuroimaging

Extraction and exploitation of complex imaging biomarkers involve an imaging processing workflow that can be quite complex. This goes from image physics and image acquisition, image processing for quality control and enhancement, image analysis for features extraction and image fusion up to the final application which intends to demonstrate the capability of the image processing workflow to issue sensitive and specific markers of a given pathology. In this context, our objectives in the recent period were directed toward 4 major methodological topics:

7.1.1. Diffusion imaging

7.1.1.1. L2 Similarity Metrics for Diffusion Multi-Compartment Model Images Registration

Participants: Renaud Hédouin, Olivier Commowick, Emmanuel Caruyer, Christian Barillot.

Diffusion multi-compartment models (MCM) allow for a fine and comprehensive study of the white matter microstructure. Non linear registration of MCM images may provide valuable information on the brain for example through population comparison. State-of-the-art MCM registration however relies on pairing-based similarity measures where the one-to-one mapping of MCM compartments is required. This approach leads to non differentiabilties or discontinuities, which may turn into poorer registration. Moreover, these measures are often specific to one MCM compartment model. We proposed [34] two new MCM similarity measures based on the space of square integrable functions, applied to MCM characteristic functions. These measures are pairing-free and agnostic to compartment types. We derived their analytic expressions for multi-tensor models and proposed a spherical approximation for more complex models. Evaluation was performed on synthetic deformations and inter-subject registration, demonstrating the robustness of the proposed measures.

7.1.1.2. Block-Matching Distortion Correction of Echo-Planar Images with Opposite Phase Encoding Directions

Participants: Renaud Hédouin, Olivier Commowick, Élise Bannier, Christian Barillot.

By shortening the acquisition time of MRI, Echo Planar Imaging (EPI) enables the acquisition of a large number of images in a short time, compatible with clinical constraints as required for diffusion or functional MRI. However such images are subject to large, local distortions disrupting their correspondence with the underlying anatomy. The correction of those distortions is an open problem, especially in regions where large deformations occur. We have proposed a new block-matching registration method to perform EPI distortion correction based on the acquisition of two EPI with opposite phase encoding directions (PED). It relies on new transformations between blocks adapted to the EPI distortion model, and on an adapted optimization scheme to ensure an opposite symmetric transformation. We have produced qualitative and quantitative results of the block-matching correction using different metrics on a phantom dataset and on in-vivo data. We have shown the ability of the block-matching to robustly correct EPI distortion even in strongly affected areas. This work has been published in IEEE Transactions on Medical Imaging [21].

7.1.1.3. Diffusion MRI processing for multi-compartment characterization of brain pathology

Participants: Renaud Hédouin, Olivier Commowick, Christian Barillot.

Diffusion weighted imaging (DWI) is a specific type of MRI acquisition based on the direction of diffusion of the brain water molecules. It allows, through several acquisitions, to model the brain microstructure, as white matter, which is significantly smaller than the voxel-resolution. To acquire a large number of images in a clinical setting, very-fast acquisition techniques are required as single-shot imaging. However these acquisitions suffer locally large distortions. We have proposed a block-matching registration method based on the acquisition of images with opposite phase-encoding directions (PED). This technique specially designed for Echo-Planar Images (EPI) robustly correct images and provides a deformation field. This field is applicable to an entire DWI series from only one reversed EPI allowing distortion correction with a minimal acquisition time cost. This registration algorithm has been validated both on phantom and on *in vivo* data and is available in our source medical image processing toolbox Anima. From these diffusion images, we are able to construct multi-compartments models (MCM) which can represent complex brain microstructure. Doing registration, averaging and atlas creation on these MCM images is required to perform studies and statistic analyses. We propose a general method to interpolate MCM as a simplification problem based on spectral clustering. This technique, which is adaptable for any MCM, has been validated on both synthetic and real data. Then, from a registered dataset, we performed a patient to population analysis at a voxel-level computing statistics on MCM parameters. Specifically designed tractography can also be used to make analysis, following tracks, based on individual anisotropic compartments. All these tools are designed and used on real data and contribute to the search of biomakers for brain diseases such as multiple sclerosis.

7.1.1.4. The challenge of mapping the human connectome based on diffusion tractography

Participant: Emmanuel Caruyer.

Tractography based on non-invasive diffusion imaging is central to the study of human brain connectivity. To date, the approach has not been systematically validated in ground truth studies. Based on a simulated human brain data set with ground truth tracts, we organized an open international tractography challenge, which resulted in 96 distinct submissions from 20 research groups. Here, we report the encouraging finding that most state-of-the-art algorithms produce tractograms containing 90 percent of the ground truth bundles (to at least some extent). However, the same tractograms contain many more invalid than valid bundles, and half of these invalid bundles occur systematically across research groups. Taken together, our results demonstrate and confirm fundamental ambiguities inherent in tract reconstruction based on orientation information alone, which need to be considered when interpreting tractography and connectivity results. Our approach provides a novel framework for estimating reliability of tractography and encourages innovation to address its current limitations [26].

7.1.1.5. Comparison of inhomogeneity distortion correction methods in diffusion MRI of the spinal cord

Participants: Haykel Snoussi, Emmanuel Caruyer, Christian Barillot.

Diffusion MRI (dMRI) is a modality that describes the geometry of neural architecture. Diffusion images suffer from various artifacts originating from subject and physiological motion, eddy currents and B0-field inhomogeneity. These can severely affect image quality particularly in the spine region. However, strategies exist to correct these distortions, including co-registration, point spread function, phase field map and reversed gradient polarity method (RGPM). We evalute various correction methods using RGPM which provides best results. More precisely, we compared Voss plus two other recent methods: Topup (FSL) and HySCO (ACID/SPM). This work was presented at the ESMRMB conference [38].

7.1.2. Arterial Spin Labeling:

Our contributions on this topic are illustrated in Fig. 2 . Arterial Spin Labeling (ASL) enables measuring cerebral blood flow in MRI without injection of a contrast agent. Perfusion measured by ASL carries relevant information for patients suffering from pathologies associated with singular perfusion patterns.

However this technique suffers from drawbacks such as low signal to noise ratio and poor resolution.

7.1.2.1. Patch-based super-resolution for arterial spin labeling MRI

Participants: Cédric Meurée, Pierre Maurel, Christian Barillot.

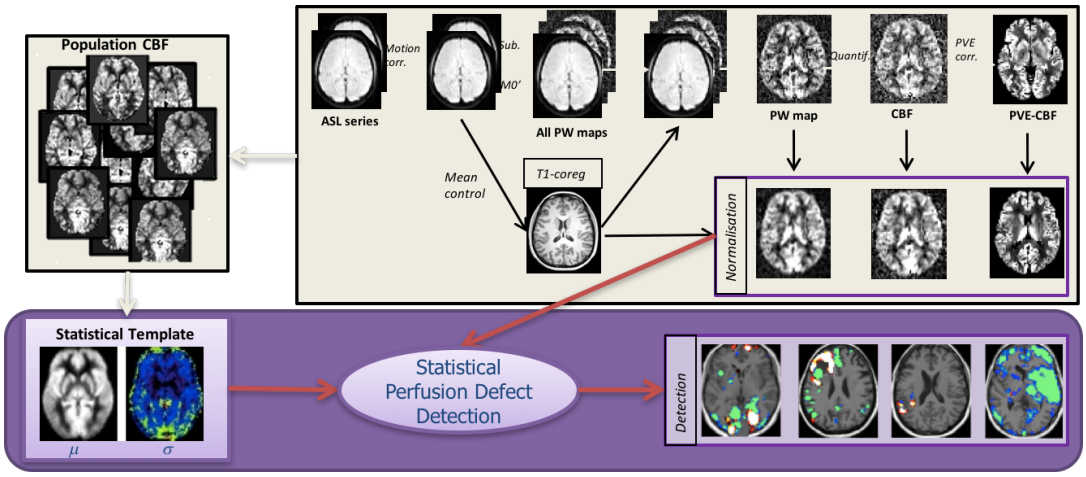


Figure 2. Summary of the image processing workflow that allows the quantification of brain perfusion and detection of potential perfusion defect on patients or populations

In this context, our contributions focused on a super resolution approach to reduce the influence of Partial Volume Effects (PVE) and obtain images close to the ones that would be acquired at a high resolution, but in a shorter scan duration. PVE are an important limitation of arterial spin labeling (ASL) acquisitions, impacting the validity of quantitative cerebral blood flow (CBF) estimations. This work consists of a super-resolution algorithm, which includes information of high resolution (HR) structural images to reconstruct HR CBF maps from low resolution ASL series, without increasing the acquisition time. Compared with nearest neighbor, trilinear and 3rd order spline interpolations, the proposed algorithm is found to generate a CBF image closer to the one obtained with a reference HR ASL acquisition. CBF calculations can therefore be improved by using this algorithm, which reduces the PVE [36].

7.1.2.2. Resting-state functional ASL

Participants: Corentin Vallée, Isabelle Corouge, Pierre Maurel, Christian Barillot.

We have started to work on resting-state functional ASL (rs-fASL). Rs-fASL in clinical daily practice and academic research stay discreet compared to resting-state BOLD. However, by giving direct access to cerebral blood flow maps, rs-fASL could lead to significant clinical subject scaled application as CBF can be considered as a biomarker in common neuropathology. As a new topic, we started by building a viable long sequence for rs-fASL. We take advantage of the long duration of the sequence to assess the link between overall quality of rs-fASL and duration of acquisition. To this end, we consider typical functional areas of the brain, and assess their quality compared to gold standards depending on the duration of acquisition. While some more work remain to be done, we tend to show there is an optimal duration of acquistion for rs-fASL. This work was submitted for the next ISMRM Conference.

7.1.2.3. Longitudinal atlas creation and brain development analysis

Participants: Antoine Legouhy, Olivier Commowick, Christian Barillot.

The study of brain development provides insights in the normal trend of brain evolution and enables early detection of abnormalities. We propose a method to quantify growth in three arbitrary orthogonal directions of the brain through linear registration. We introduce a 9 degrees of freedom transformation that gives the opportunity to extract scaling factors describing brain growth along those directions by registering a data base of subjects in a common basis. We apply this framework to create a longitudinal curve of scaling ratios along fixed orthogonal directions from 0 to 16 years highlighting anisotropic brain development. In pediatric

image analysis, the study of brain development provides insights in the normal trend of brain evolution and enables early detection of abnormalities. Tools like longitudinal atlases allow to compute statistics on populations, understand brain variability at different ages to highlight changes in growth, shape, structure etc. We experimented different methods to perform longitudinal atlases. This work was submitted for the next ISMRM Conference.

7.1.3. Quantitative relaxation times estimation and processing:

The VisAGeS team has proposed new methodologies to exploit new relaxometry sequences, able to provide direct information on tissue properties (T1, T2, T2* relaxation times) and their alteration in diseases. Such sequences have a great potential in diagnostic and evolution study of patients suffering from various neurological diseases.

7.1.3.1. Gaining Insights Into Multiple Sclerosis Lesion Characteristics from Brain Tissue Microstructure Information: A Multi-Compartment T2 Relaxometry Approach:
Participants: Sudhanya Chatterjee, Olivier Commowick, Christian Barillot.

In addition to raw relaxation times, we have also studied other estimation methods able, from T2 relaxometry sequences, to estimate the fraction of myelin (myelin water fraction) inside each voxel, a quantity that may be largely impacted in neurological diseases. To this end, we have proposed new multi-compartment T2 estimation methods [42] with a new water three-compartment T2 model of tissue bounded water (free water, axons and cells, and myelin), using variable projection to make the estimation faster and more robust. Clinical trends and pathogenetic ways of onset and progression of Multiple Sclerosis (MS) in patients suggest that MS is a highly heterogeneous disease. MS is predominantly a White Matter (WM) disease, which is mainly composed of myelinated axons and neuroglia type cells. Demyelination and axonal loss characterize the condition of MS in a patient. However, they follow varying trends in patients. In this work, we propose a method in which T2 relaxometry data is used to obtain a quantitative brain tissue microstructure information. This information is then studied to check its corroborations with pathogenetic understanding of MS in literature [41].

7.1.3.2. Multi-Compartment T2 Relaxometry Model Using Gamma Distribution Representations: A Framework for Quantitative Estimation of Brain Tissue Microstructures:
Participants: Sudhanya Chatterjee, Olivier Commowick, Christian Barillot.

Advanced MRI techniques (e.g., d-MRI, MT, relaxometry etc.) can provide quantitative information of brain tissues. Image voxels are often heterogeneous in terms of microstructure information due to physical limitations and imaging resolution. Quantitative assessment of the brain tissue microstructure can provide valuable insights into neurodegenerative diseases (e.g., Multiple Sclerosis). In this work, we propose a multicompartment model for T2-Relaxometry to obtain brain microstructure information in a quantitative framework. The proposed method allows simultaneous estimation of the model parameters [42].

7.1.4. Multi-modal EEG and fMRI Source Estimation Using Sparse Constraints:
Participants: Saman Noorzadeh, Pierre Maurel, Christian Barillot.

In this work, a multi-modal approach is presented and validated on real data to estimate the brain neuronal sources based on EEG and fMRI. Combining these two modalities can lead to source estimations with high spatio-temporal resolution. The joint method is based on the idea of linear model already presented in the literature where each of the data modalities are first modeled linearly based on the sources. Afterwards, they are integrated in a joint framework which also considers the sparsity of sources. The sources are then estimated with the proximal algorithm. The results are validated on real data and show the efficiency of the joint model compared to the uni-modal ones. We also provide a calibration solution for the system and demonstrate the effect of the parameter values for uni- and multi-modal estimations on 8 subjects [37].

7.2. Research axis 2: Applications in Neuroradiology and Neurological Disorders

7.2.1. Arterial Spin Labeling:

Participants: Jean-Christophe Ferré, Maia Proisy, Isabelle Corouge, Élise Bannier, Christian Barillot.

Arterial Spin Labeling is an attractive perfusion MRI technique due to its complete non-invasiveness. However it still remains confidential in clinical practice. Over the years, we have developed several applications to evaluate its potential in different contexts. In 2017, in the context of the MALTA project, we focused on the application of ASL to activation-fMRI. Functional Arterial Spin Labeling (fASL) has demonstrated its greater specificity as a marker of neuronal activity than the reference BOLD fMRI for motor activation mapping in healthy volunteers. Motor fASL was yet to be investigated in the context of tumors, under the assumption that fASL would be less sensitive to venous contamination induced by the hemodynamics remodeling in the tumor vicinity than BOLD fMRI. As the arterial transit time may be shortened in activation areas, we explored the ability of fASL to map the motor areas at different post-labeling delays (PLD) in healthy subjects and patient with brain tumor. As part of the PhD of Maia Proisy, we have also been working on processing and analyse MR perfusion images using arterial spin labeling in neonates and children for several purposes:

- ASL and TOF-MRA are two totally non-invasive, easy-to-use MRI sequences for children in emergency settings. Hypoperfusion associated with homolateral vasospasm may suggest a diagnosis of migraine with aura (published in Cephalagia and presented in 3 congresses including RSNA)
- Investigation of brain perfusion evolution between 6 month and 15 years using ASL sequence in order to provide reference values in this age range (Measurement of pediatric regional cerebral blood flow from 6 months to 15 years of age article under revision, presented in one national congress)
- Work in Progress: ASL perfusion images in 20 neonates with hypoxic-ischemic encephalopathy that underwent MRI on day-of-life 3 and day-of-life 10.

7.2.2. Hybrid EEG-fMRI Neurofeedback:

Participants: Lorraine Perronnet, Marsel Mano, Élise Bannier, Mathis Fleury, Giulia Lioi, Christian Barillot.

Over the last 4 years, we developed a whole new range of activities around hybrid EEG-MR imaging and neurofeedback for brain rehabilitation. We propose to combine advanced instrumental devices (Hybrid EEG and MRI platforms), with new man-machine interface paradigms (Brain computer interface and serious gaming) and new computational models (source separation, sparse representations and machine learning) to provide novel therapeutic and neuro-rehabilitation paradigms in some of the major neurological and psychiatric disorders of the developmental and the aging brain. We first performed 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 design a NF protocol based on motor imagery and compatible with EEG and fMRI. We implemented different types of feedback and compared for the first time the effects of unimodal EEG-NF and fMRI-NF versus bimodal EEG-fMRI-NF by looking both at EEG and fMRI activations. We also introduced a new feedback metaphor for bimodal EEG-fMRI-neurofeedback that integrates both EEG and fMRI signal in a single bi-dimensional feedback (a ball moving in 2D). The participants to this study were able to regulate activity in their motor regions in all NF conditions. Our results also suggest that that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-NF alone [31].

All the experiments were performed on the Neurinfo platform which is equipped with an EEG MR compatible 64-channel device in 2014 to perform joint EEG and BOLD or ASL fMRI. We developed, installed and successfully tested a hybrid EEG-fMRI platform for bimodal NF experiments. Our system is based on the integration and the synchronization of an MR-compatible EEG and fMRI acquisition subsystems. We developed two real-time pipelines for EEG and fMRI that handle all the necessary signal processing, the joint NF block that calculates and fuses the NF and a visualization block that displays the NF to the subject. The control and the synchronization of both subsystems with each other and with the experimental protocol

is handled by the NF Control. Our platform showed very good real-time performance with various pre-processing, filtering, and NF estimation and visualization methods. Its modular architecture is easily adaptable to different experimental environments, and offers high efficiency for optimal real-time NF applications [27].

These developments came as part of the HEMISFER project which is conducted through a very complementary set of competences over the different teams involved (Visages Inserm U1228, HYBRID and PANAMA Teams from Inria/IRISA, EA 4712 team from University of Rennes I and ATHENA team from Inria Sophia-Antipolis). The overall principle of this project is illustrated in Fig. 3 .

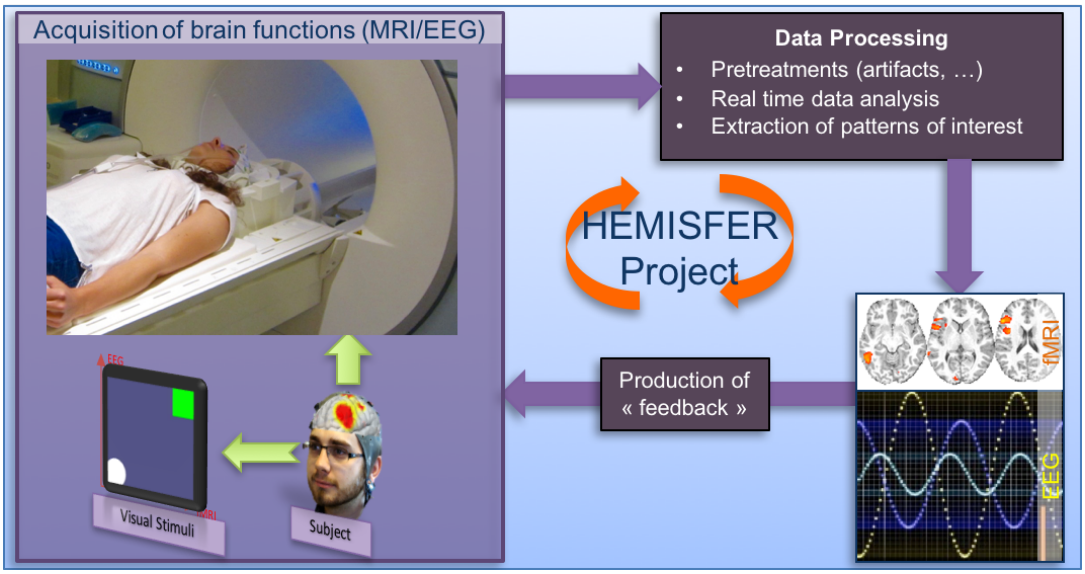


Figure 3. Principle of the Hybrid EEG:fMRI environement set up and used by the HEMISFER project

7.2.3. Multiple sclerosis:

Participants: Anne Kerbrat, Gilles Edan, Jean-Christophe Ferré, Benoit Combès, Olivier Commowick, Élise Bannier, Sudhanya Chatterjee, Haykel Snoussi, Emmanuel Caruyer, Christian Barillot.

The VisAGeS research team has a strong focus on applying the developed methodologies (illustrated in research axis 1) to multiple sclerosis (MS) understanding and the prediction of its evolution. Related to the EMISEP project on spinal cord injury evolution in MS, a first work investigated the magnetization transfer reproducibility across centers in the spinal cord and was accepted for presentation at ESMRMB [33]. Based on this work, a second work investigated the sensitivity of magnetization transfer to assess diffuse and focal burden in MS patients [43]. In parallel, methodological developments have addressed spinal cord diffusion data analysis, starting with a comparaison of several distortion correction methods [38].

Finally, we investigated myelin water fraction (MWF) estimation on multiple sclerosis and demonstrated in longitudinal studies [41] how these figures can be related with lesion evolution, paving the way towards myelin oriented MS evaluation of patient future evolution prediction (and thus treatment adaptation) and joint studies between different quantitative imaging modalities (e.g., diffusion).

7.2.4. Recovery imaging:

Participants: Isabelle Bonan, Stephanie Leplaideur, Élise Bannier, Jean-Christophe Ferré, Christian Barillot.

More common after a right hemispheric brain injury, misperception of body in space, impacting moves and posture is often associated with disturbance of spatial attention (behavioural symptoms of a failure in spontaneously reorienting attention to stimulus information in the left field). While different subjects use different references in their elaboration of spatial representation, body-centered coordinate systems are the most prevalent. As part of an fMRI substudy of a national research study on balance disorder rehabilitation, we investigated differences in activations during body-centered spatial tasks in corporeal and in extracorporeal space. Healthy controls and stroke patients were included in this fMRI sub study comprising 2 egocentric spatial tasks: perception of the midsagittal plane in extracorporeal space (straight-ahead task) and in corporeal space (longitudinal axis task). Results obtained on healthy control data were presented at the SOFMER conference and the journal paper is under review. For both tasks, cerebral activations largely dominated in the right hemisphere and essentially involved the right frontoparietal network. In addition, the straight-ahead task presented specific activations in the temporoparieto-insular cortex and thalamic areas. Patient data processing is ongoing in the context of an MD-PhD. In parallel, a master study investigated the brain structural connections between the cortical areas obtained from the fMRI study using diffusion MRI and the white matter query language.

7.2.5. White matter connectivity analysis in patients suffering from depression:

Participants: Julie Coloigner, Jean-Marie Batail, Jean-Christophe Ferré, Isabelle Corouge, Christian Barillot.

The mood depressive disorder (MDD) is a common chronically psychiatric disorder with an estimated lifetime prevalence reported to range from 10 percent to 15 percent worldwide. This disease is characterized by an intense dysregulation of affect and mood as well as additional abnormalities including cognitive dysfunction, insomnia, fatigue and appetite disturbance. Despite the extensive therapy options available for depression, up to 80 percent of patients will suffer from a relapse [1]. Consequently, exhibiting imaging biomarkers of this disease will support both a better understanding of the neural correlates underlying the depression, and a better diagnosis and treatment of individual depressed patients. Previous studies of structural and functional magnetic resonance imaging have reported several microstructural abnormalities in the prefrontal cortex, anterior cingulate cortex, hippocampus and thalamus [2]. These observations suggest a dysfunction of the circuits connecting frontal and subcortical brain regions, leading to a "disconnection syndrome" [3]. Given the small sample size used in the past studies, we proposed a more robust analysis using a larger cohort of patients suffering from depression, called LONGIDEP. The latter is a routine care cohort of patients suffering from mood depressive disorder who underwent a clinical evaluation, neuropsychological testing and brain MRI. The population sample consists of 125 patients suffering from depression and 65 healthy age and gender-matched, control subjects. A composite measure of medication load for each patient was assessed using a previously established method [4]. We investigated alterations of white matter integrity using a voxel-based analysis based on fractional anisotropy (FA) and the apparent diffusion coefficient (ADC) in patients with depression. Using graph theory-based analysis, we also examined white matter changes in the organization of networks in patients suffering from depression. Our findings provide robust evidence that the reduction of white-matter integrity in the interhemispheric connections and fronto-limbic neuronal circuits may play an important role in MDD pathogenesis. These results are consistent with an overall hypothesis that depression involves a disconnection of prefrontal, striatal, and limbic emotional areas.

7.2.6. Knowing and Remembering: Cognitive and Neural Influences of Familiarity on Recognition Memory in Early Alzheimer's Disease (EPMR-MA):

Participants: Pierre-Yves Jonin, Quentin Duché, Élise Bannier, Christian Barillot.

Inclusion of the 20 healthy participants in the "EPMR-MA" study (clinical trials ID NCT02492529) has been achieved, the inclusion phase will be achieved before 30th, december, 2017. Healthy controls data are pre-processed and the first analysis workflow proved promising, it should allow submitting a first paper at the beginning of 2018.

7.2.7. Semantic Dementia Imaging:

Participants: Jean-Christophe Ferré, Isabelle Corouge, Elise Bannier, Christian Barillot.

After demonstrating the relative preservation of fruit and vegetable knowledge in patients with semantic dementia (SD), we sought to identify the neural substrate of this unusual category effect. Nineteen patients with SD performed a semantic sorting task and underwent a morphometric 3T MRI scan. The grey-matter volumes of five regions within the temporal lobe were bilaterally computed, as well as those of two recently described areas (FG1 and FG2) within the posterior fusiform gyrus. In contrast to the other semantic categories we tested, fruit and vegetable scores were only predicted by left FG1 volume. We therefore found a specific relationship between the volume of a subregion within the left posterior fusiform gyrus and performance on fruits and vegetables in SD. We argue that the left FG1 is a convergence zone for the features that might be critical to successfully sort fruits and vegetables. We also discuss evidence for a functional specialization of the fusiform gyrus along two axes (lateral medial and longitudinal), depending on the nature of the concepts and on the level of processing complexity required by the ongoing task [28].

7.3. Research axis 3: Management of Information in Neuroimaging

Participants: Élise Bannier, Christian Barillot, Yao Chi, Isabelle Corouge, Olivier Commowick, Inès Fakhfakh, Michael Kain, Florent Leray, Julien Louis, Aneta Morawin, Mathieu Simon, Arnaud Touboulic.

The major topic that has been reached in the period concerns the sharing of data and processing tools in neuroimaging (through the ANR Neurolog and VIP projects, and more recently the "Programme d'Investissement d'Avenir" project such as OFSEP and FLI-IAM) that led to build a suitable architecture to share images and processing tools). Our overall goal within these projects was to set up a computational infrastructure to facilitate the sharing of neuroimaging data, as well as image processing tools, in a distributed and heterogeneous environment. These consortiums gathered expertises coming from several complementary domains: image processing in neuroimaging, workflows and grid computing, ontology development and ontology-based mediation. Shanoir (SHAring NeurOImaging Resources) is one of the major outcome of these projects. Shanoir uses semantics for concepts organization that are defined by the OntoNeuroLOG ontology. OntoNeuroLOG reuses and extends the OntoNeuroBase ontology. Both were designed using the same methodological framework, based on the use of the foundational ontology DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), and the use of a number of core ontologies, that provide generic, basic and minimal concepts and relations in specific domains such as Artefacts. Shanoir aims at establishing the conditions allowing, through the Internet, to share distributed information sources in neuroimaging, whether these sources are located in various centers of experimentation, clinical departments of neurology, or research centers in cognitive neurosciences or image processing. This enables a large variety of users to diffuse, exchange or reach neuroimaging information with appropriate access means, in order to be able to retrieve information almost as easily as if the data were stored locally by means of the "cloud computing" Storage as a Service (SaaS) concept. The Shanoir environment has been succesfully deployed to the Neurinfo platform were it is routinely used to manage images of the research studies. It is also currently being deployed for two large projets: OFSEP ("*Observatoire Français de la Sclérose en Plaques*") where up to 30000 patients will be acquired on a ten years frame, and the Image Analysis and Management (IAM) node of the France Life Imaging national infrastructure (FLI-IAM). Our team VisAGeS fulfills multiple roles in this nation-wide FLI project. Christian Barillot is the chair of the IAM node, Olivier Commowick is participating in the working group workflow and image processing and Michael Kain is the technical manager of the node. Apart from the team members, software solutions like MedInria and Shanoir are part of the final infrastructure software solutions.

AIRSEA Project-Team

7. New Results

7.1. Modeling for Oceanic and Atmospheric flows

7.1.1. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Laurent Debreu, Florian Lemarié, Christopher Eldred, Farshid Nazari.

The increase of model resolution naturally leads to the representation of a wider energy spectrum. As a result, in recent years, the understanding of oceanic submesoscale dynamics has significantly improved. However, dissipation in submesoscale models remains dominated by numerical constraints rather than physical ones. Effective resolution is limited by the numerical dissipation range, which is a function of the model numerical filters (assuming that dispersive numerical modes are efficiently removed). A review paper on coastal ocean models has been written with German colleagues and will be published in Ocean Modelling early 2018 ([34]). Ocean models usually rely on a mode splitting procedure which separates the fast external gravity waves with the slower internal waves. A paper on the stability of the mode splitting has been submitted to Journal of Computational Physics ([21])
The team is involved in the HEAT (Highly Efficient ATmospheric Modelling) ANR project. This project aims at developing a new atmospheric dynamical core (DYNAMICO) discretized on an icosahedral grid. This project is in collaboration with Ecole Polytechnique, Meteo-France, LMD, LSCE and CERFACS. This year we worked on dispersion analysis of compatible Galerkin schemes for a 1D shallow water model ([8]).

7.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Laurent Debreu, Florian Lemarié, Charles Pelletier, Antoine Rousseau, Sophie Thery.

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approximation of one [61]. For the last few years we have been actively working on the analysis of ocean-atmosphere coupling both in terms of its continuous and numerical formulation.Our activities over the last few years can be divided into four general topics

1. *Stability and consistency analysis of existing coupling methods*: in [61] we showed that the usual methods used in the context of ocean-atmosphere coupling are prone to splitting errors because they correspond to only one iteration of an iterative process without reaching convergence. Moreover, those methods have an additional condition for the coupling to be stable even if unconditionally stable time stepping algorithms are used. This last remark was further studied recently in [1] and it turned out to be a major source of instability in atmosphere-snow coupling.
2. *Study of physics-dynamics coupling*: during the PhD-thesis of Charles Pelletier (funded by Inria) the scope is on including the formulation of physical parameterizations in the theoretical analysis of the coupling, in particular the parameterization schemes to compute air-sea fluxes [18]. To do so, a metamodel representative of the behavior of the full parameterization but with a continuous form easier to manipulate has been derived thanks to a sensitivity analysis based on Sobol' indexes. This metamodel has the advantage to be more adequate to conduct the mathematical analysis of the coupling while being physically satisfactory. This work is in revision for publication in Quarterly Journal of the Royal Meteorological Society and has been presented in various conferences [69], [24], [20], [17]. In parallel we have contributed to a general review gathering the main international specialists on the topic [53].

3. *Design of a coupled single column model*: in order to focus on specific problems of ocean-atmosphere coupling, a work on simplified equation sets has been started. The aim is to implement a one-dimensional (in the vertical direction) coupled model with physical parameterizations representative of those used in realistic models. Thanks to this simplified coupled model the objective is to develop a benchmark suite for coupled models evaluation. Last year the single column oceanic and atmospheric components have been developed and coupled during the PhD-thesis of Rémi Pellerej and in the framework of the SIMBAD project. A publication describing this model and its interfacing with the OOPS software to allow the implementation of various data assimilation techniques is currently in preparation for the Geoscientific Model Development journal.
4. *Analysis of air-sea interactions in realistic high-resolution realistic simulations*: part of our activity has been in collaboration with atmosphericists and physical oceanographers to study the impact on some modeling assumptions (e.g. [62]) in large-scale realistic ocean-atmosphere coupled simulations [70], [66], [12].

These four topics are addressed through strong collaborations between the applied mathematicians and the climate community.

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 (from march 2015 to march 2018). 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. First results have been presented during international conferences [22], [23] and a publication is currently in preparation. Another industrial contract named ALBATROS is also funded by (from June 2016 to June 2019) to couple SIMBAD with the NEMO global ocean model and a wave model called WW3.

An ANR project COCOA (COmprehensive Coupling approach for the Ocean and the Atmosphere, P.I.: E. Blayo) has been funded in 2016 and has officially start in January 2017.

7.1.3. Data assimilation for coupled models

In the context of operational meteorology and oceanography, forecast skills heavily rely on proper combination of model prediction and available observations via data assimilation techniques. Historically, numerical weather prediction is made separately for the ocean and the atmosphere in an uncoupled way. However, in recent years, fully coupled ocean-atmosphere models are increasingly used in operational centers to improve the reliability of seasonal forecasts and tropical cyclones predictions. For coupled problems, the use of separated data assimilation schemes in each medium is not satisfactory since the result of such assimilation process is generally inconsistent across the interface, thus leading to unacceptable artefacts. Hence, there is a strong need for adapting existing data assimilation techniques to the coupled framework. As part of our ERACLIM2 contribution, R. Pellerej started a PhD on that topic late 2014 and will defend it early 2018. So far, three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a single column coupled model. The dynamical equations of the considered problem are coupled using an iterative Schwarz domain decomposition method. The aim is to properly take into account the coupling in the assimilation process in order to obtain a coupled solution close to the observations while satisfying the physical conditions across the air-sea interface. Preliminary results shows significant improvement compared to the usual approach on this simple system [68], [25]. The aforementioned system has been coded within the OOPS framework (Object Oriented Prediction System) in order to ease the transfer to more complex/realistic models.

The second contribution to ERACLIM2 was to investigate the importance of the quality of the data assimilation scheme in the ocean in the coupled system. It led to the proposition of cost effective approximations either in term of resolution reduction or equation simplifications, along with a metric to asses the quality of said approximations [35]

Finally, CASIS, a new collaborative project with Mercator Océan has started late 2017 in order to extend developments to iterative Kalman smoother data assimilation scheme, in the framework of a coupled ocean-atmospheric boundary layer context.

7.1.4. Parameterizing subgrid scale eddy effects

Participant: Eugene Kazantsev.

Basing on the maximum entropy production principle, the influence of subgrid scales on the flow is presented as the harmonic dissipation accompanied by the backscattering of the dissipated energy. This parametrization is tested on the shallow water model in a square box. Two possible solutions of the closure problem are compared basing on the analysis of the energy dissipation-backscattering balance. Results of this model on the coarse resolution grid are compared with the reference simulation at four times higher resolution. It is shown that the mean flow is correctly recovered, as well as variability properties, such as eddy kinetic energy fields and its spectrum [33].

7.2. Model reduction / multiscale algorithms

7.2.1. Intrusive sensitivity analysis, reduced models

Participants: Maëlle Nodet, Clémentine Prieur.

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 [57], 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 [58]. In [55], 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 [59].

When considering parameter-dependent PDE, it happens that the quantity of interest is not the PDE’s solution but a linear functional of it. In [56], we have proposed a probabilistic error bound for the reduced output of interest (goal-oriented error bound). By probabilistic we mean that this bound may be violated with small probability. The bound is efficiently and explicitly computable, and we show on different examples that this error bound is sharper than existing ones.

A collaboration has been started with Christophe Prieur (Gipsa-Lab) on the very challenging issue of sensitivity of a controlled system to its control parameters [59]. In [60], we propose a generalization of the probabilistic goal-oriented error estimation in [56] to parameter-dependent nonlinear problems. One aims at applying such results in the previous context of sensitivity of a controlled system.

7.3. Dealing with uncertainties

7.3.1. Sensitivity Analysis

Participants: Eric Blayo, Laurent Gilquin, François-Xavier Le Dimet, Elise Arnaud, Maëlle Nodet, Clémentine Prieur, Laurence Viry.

7.3.1.1. 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.

7.3.1.2. 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 [54] in the case where input parameters are correlated. Clémentine Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [41]. We obtained first results [42], 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 [40], [43]. We also considered the estimation of groups Sobol' indices, with a procedure based on replicated designs [52]. 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 supervised the PhD thesis of Simon Nanty (funded by CEA Cadarache, and defended in October, 2015). 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 implied functional dependent inputs. A first step was the modeling of these inputs [64]. The whole methodology proposed during the PhD is presented in [65].

More recently, the Shapley value, from econometrics, was proposed as an alternative to quantify the importance of random input variables to a function. Owen [67] derived Shapley value importance for independent inputs and showed that it is bracketed between two different Sobol' indices. Song et al. [72] recently advocated the use of Shapley value for the case of dependent inputs. In a very recent work [13], in collaboration with Art Owen (Standford's University), we show that Shapley value removes the conceptual problems of functional ANOVA for dependent inputs. We do this with some simple examples where Shapley value leads to intuitively reasonable nearly closed form values. We also investigated further the properties of Shapley effects in [31].

7.3.2. Non-Parametric Estimation for Kinetic Diffusions

Participants: Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Chile and Uruguay. More precisely, we started working with Venezuela. Due to the crisis in Venezuela, our main collaborator on that topic moved to Uruguay.

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 [37] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [38] (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 accepted [38], in collaboration with J.R. Leon (Montevideo, Uruguay) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [39], also recently accepted. In a recent collaboration with Adeline Samson from the statistics department in the Lab, we considered adaptive estimation, that is we proposed a data-driven procedure for the choice of the bandwidth parameters.

In [5], we focused on damping Hamiltonian systems under the so-called fluctuation-dissipation condition. Idea in that paper were re-used with applications to neuroscience in [63].

Note that Professor Jose R. Leon (Caracas, Venezuela, Montevideo, Uruguay) was funded by an international Inria Chair, allowing to collaborate further on parameter estimation.

We recently proposed a paper on the use of the Euler scheme for inference purposes, considering reflected diffusions.This paper could be extended to the hypoelliptic framework.

We started a collaboration with Karine Bertin (Valparaiso, Chile) funded by a MATHAMSUD project. We are interested in new adaptive estimators for invariant densities on bounded domains, and would like to extend that results to hypo-elliptic diffusions.

7.3.3. 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 [44] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [71] 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 [45] 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 [46], [48], 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 [29].

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [47]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervised 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 and was defended in February 2017. A first paper on data reconstruction has been accepted [73]. It was a necessary step as the initial series contained many missing data. A second paper is in revision, considering the modeling of precipitation amount with semi-parametric models, modeling both the bulk of the distribution and the tails, but avoiding the arbitrary choice of a threshold. We work in collaboration with Philippe Naveau (LSCE, Paris).

7.3.4. Extensions of the replication method for the estimation of Sobol' indices

Participants: Elise Arnaud, Laurent Gilquin, Clémentine Prieur.

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 n becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. However, if the input space is not properly explored, that is if n is too small, the Sobol' indices estimates may not be accurate enough. Gaining in efficiency and assessing the estimate precision still remains an issue, all the more important when one is dealing with limited computational budget.

We designed approaches to render the replication method recursive, enabling the required number of evaluations to be controlled. With these approaches, more accurate Sobol' estimates are obtained while recycling previous sets of model evaluations. The estimation procedure is therefore stopped when the convergence of estimates is considered reached. One of these approaches corresponds to a recursive version of the replication method and is based on the iterative construction of stratified designs, latin hypercubes and orthogonal arrays [50]. A second approach combines the use of quasi-Monte Carlo sampling and the construction of a new stopping criterion [9] [32] .

In [30] a new strategy to estimate the full set of first-order and second-order Sobol' indices with only two replicated designs based on orthogonal arrays of strength two. Such a procedure increases the precision of the estimation for a given computation budget. A bootstrap procedure for producing confidence intervals, that are compared to asymptotic ones in the case of first-order indices, is also proposed.

7.3.5. Parameter control in presence of uncertainties: robust estimation of bottom friction

Participants: Victor Trappler, Elise Arnaud, Laurent Debreu, Arthur Vidard.

Many physical phenomena are modelled numerically in order to better understand and/or to predict their behaviour. However, some complex and small scale phenomena can not be fully represented in the models. The introduction of ad-hoc correcting terms, can represent these unresolved processes, but they need to be properly estimated.

A good example of this type of problem is the estimation of bottom friction parameters of the ocean floor. This is important because it affects the general circulation. This is particularly the case in coastal areas, especially for its influence on wave breaking. Because of its strong spatial disparity, it is impossible to estimate the bottom friction by direct observation, so it requires to do so indirectly by observing its effects on surface movement. This task is further complicated by the presence of uncertainty in certain other characteristics linking the bottom and the surface (eg boundary conditions). The techniques currently used to adjust these settings are very basic and do not take into account these uncertainties, thereby increasing the error in this estimate.

Classical methods of parameter estimation usually imply the minimisation of an objective function, that measures the error between some observations and the results obtained by a numerical model. In the presence of uncertainties, the minimisation is not straightforward, as the output of the model depends on those uncontrolled inputs and on the control parameter as well. That is why we will aim at minimising the objective

function, to get an estimation of the control parameter that is robust to the uncertainties. In this work, a toy model of a coastal area has been modelled and implemented. The control parameter is the bottom friction, upon which classical methods of estimation are applied in a simulation-reestimation experiment. The model is then modified to include uncertainties on the boundary conditions in order to apply robust control methods. First, a sensitivity analysis of the objective function has been performed to assess the influence of each considered variable. Then, a study on the meaning of different concepts of robustness have been carried on. Typically, one then seeks an optimal parameter set that would minimise the variance or the mean of the original objective function. Various associated algorithms from the literature have been implemented. They all rely on surrogate models and black-box optimisation techniques to solve this estimation problem.

7.3.6. Sensitivity of a floating offshore wind turbine to uncertain parameters

Participants: Adrien Hirvoas, Elise Arnaud, Clémentine Prieur, Arthur Vidard.

In a fast-changing energy context, marine renewable energies in general and floating offshore wind energy in particular are a promising source of energy in France and aboard. The design of these structures is made in a specific regulated framework related to their environment. Floating offshore wind turbines are submitted to various continuous environmental loadings (wind, current, swell), which generate solicitations and fatigue in some components. Fatigue lifetime is estimated with a dedicated software that allows performing coupled multi-physics simulations of the system (hydrodynamics, aerodynamics, mechanics and controls). The inputs of these simulations necessarily include uncertainties regarding the environmental loadings and the physical parameters of the models as well. These uncertainties can have an influence on the simulated behaviour of the system. The core of this work consists in conducting a sensitivity analysis to assess, how the uncertainty on an output of a model can be attributed to sources of uncertainty among the inputs. The approach that is considered, is based on the calculation of Sobol indices with the FAST method, and a meta-model using Kriging. These indices are used to evaluate in what extend an input or group of inputs is responsible for the output variance. The perspectives of this study is to understand what king of measurements could be of interest to properly estimate the sensible parameters, and where these measurements should be monitored on the structure. Such an estimation will be performed with data assimilation approaches, which optimally combine numerical models and physical observations. This work in done in collaboration with IFPEN.

7.3.7. Uncertainty and robustess analysis for models with functional input/output.

Participants: Mohammed Reda El Amri, Clémentine Prieur.

Numerical models are commonly used to study physical phenomena. They imply many inputs parameters, and potentially provide a large number of quantities of interest as outputs. Practitioners are not only interested in the response of their model for a given set of inputs (forward problem) but also in recovering the set of inputs values leading to a prescribed value or range for the quantity of interest (inversion problem). In collaboration with IFP Energies nouvelles, we develop data-driven strategies for robust inversion under functional uncertainties. Reda El Amri's PhD thesis aim at developping such tools with application to pollutant emission control.

7.4. Assimilation of Images

Participants: Elise Arnaud, François-Xavier Le Dimet, Maëlle Nodet, Arthur Vidard, Nelson Feyeux.

7.4.1. 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.

Our current developments are targeted at the use of « 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. This approach is being applied to the tracking of oceanic oil spills [10]

A collaborative project started with C. Lauvernet (IRSTEA) in order to make use of our image assimilation strategies on the control of pesticide transfer.

7.4.2. Optimal transport for image assimilation

We investigate the use of optimal transport based distances for data assimilation, and in particular for assimilating dense data such as images. The PhD thesis of N. Feyeux studied the impact of using the Wasserstein distance in place of the classical Euclidean distance (pixel to pixel comparison). In a simplified one dimensional framework, we showed that the Wasserstein distance is indeed promising. Data assimilation experiments with the Shallow Water model have been performed and confirm the interest of the Wasserstein distance. Results have been presented at conferences and seminars and a paper is under minor revision at NPG [49].

7.5. 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. Laurent Gilquin defended his PhD in October 2016 [51] and Thomas Capelle defended his in April 2017 and published his latest results in [4].

ANGE Project-Team

7. New Results

7.1. Modelling of complex flows

7.1.1. Modelling and simulation of sediment transport

Participants: Emmanuel Audusse, Léa Boittin, Martin Parisot, Jacques Sainte-Marie.

Following previous works, a numerical scheme for the sediment layer is proposed and assessed. The influence of the viscosity on the behaviour of the sediment layer is studied. A numerical strategy for the resolution of the coupled model (water layer and sediment layer) is implemented. The behaviour of the coupled system is numerically assessed. Academic test cases are performed.

7.1.2. Modelling of photosynthesis through microalgae cultivation

Participants: Marie-Odile Bristeau, Jacques Sainte-Marie.

In collaboration with O. Bernard.

In the present multidisciplinary downscaling study, we reconstruct single cell trajectories in an open raceway and experimentally reproduce the according high frequency light pattern to observe its effect on the growth of Dunaliella salina. We show that the frequency of such a realistic signal plays a decisive role on the dynamics of photosynthesis, which reveal an unexpected photosynthetic response compared to that recorded under the on/off signals usually used in the literature. This study highlights the need for experiments with more realistic light stimuli in order to better understand microalgal growth at high cell density. We also propose an experimental protocol with simple piecewise constant, yet more realistic, light fluctuations.

7.1.3. Buoyancy modelling

Participants: Edwige Godlewski, Martin Parisot, Jacques Sainte-Marie, Fabien Wahl.

Firstly, the work of the previous year was completed and lead to the submission of an article [38]. More precisely the fixed point algorithm is rewritten using a new unknown. This allows to increase the numerical robustness and accuracy of the scheme. The proposed resolution is assessed on several stationary and non-stationary test cases with analytical solutions.

In the continuity of this work, the modelling of fluid-structure interaction resolution is added in the previous work in order to simulate floating structures for marine energy devices. In a first step only the vertical movement is studied, with no major scientific lock. In a second time the horizontal movement of the structure is considered and required a deeper analysis to ensure the entropy-stability at the discrete level.

7.1.4. A Free Interface Model for Static/Flowing Dynamics in Thin-Layer Flows of Granular Materials with Yield: Simple Shear Simulations and Comparison with Experiments

Participant: Anne Mangeney.

In collaboration with C. Lusso, F. Bouchut, A. Ern.

Flows of dense granular materials comprise regions where the material is flowing, and regions where it is static. In [15], we introduce two numerical methods to deal with the particular formulation of this model with a free interface. They are used to evaluate the respective role of yield and viscosity for the case of a constant source term, which corresponds to simple shear viscoplastic flows. Both the analytical solution of the inviscid model and the numerical solution of the viscous model (with a constant viscosity or the variable viscosity of the $\mu(I)$-rheology) are compared with experimental data.

7.1.5. Metamodelling of a road traffic assignment model

Participant: Vivien Mallet.

In collaboration with R. Chen, V. Aguiléra, F. Cohn, D. Poulet, F. Brocheton.

We proposed a metamodelling approach to design a close approximation to the traffic model, but with a very low computational cost. It consists in a dimensionality reduction of the model outputs by principal component analysis and a statistical emulation relying on regression and interpolation between training samples. A case study was carried out for the agglomeration of Clermont-Ferrand (France). Compared with traffic flow measurements, the performance of the metamodel is similar to that of the complete model during a one-month period, but the computational time decreases from 2 days on 110 cores to less than 1 minute on one core.

7.2. Assessments of models by means of experimental data and assimilation

7.2.1. Evaluation and calibration of mobile phones for noise monitoring application

Participants: Vivien Mallet, Raphaël Ventura.

In collaboration with V. Issarny, P-G. Raverdy, F. Rebhi.

The Ambiciti application was developed so as to acquire a larger control over the acquisition process by mobile phone sensors. Pink and narrowband noises were used to evaluate the phones' accuracy at levels ranging from background noise to 90 dB(A) inside the lab. Conclusions of this evaluation lead to the proposition of a calibration strategy that has been embedded in Ambiciti and applied to more than 50 devices during public events. In the perspective of citizens-driven noise sensing, in situ experiments were carried out, while additional tests helped to produce recommendations regarding the sensing context (grip, orientation, moving speed, mitigation, frictions, wind).

7.2.2. Assimilation of noise pollution data

Participants: Vivien Mallet, Raphaël Ventura.

In collaboration with P. Aumond, A. Can, V. Issarny.

We studied the generation of hourly noise maps in urban area at street resolution, based on temporally averaged simulation maps and mobile phone audio recordings. A data assimilation method produces an analysis noise map which is the so-called best linear unbiased estimator. We illustrated the method with a neighborhood-wide experiment.

Another work, lead by IFSTTAR, was dedicated to the spatial interpolation of point measurements collected at high density in Paris with a sound level meter. Compelling results were obtained with universal Kriging and a linear trend based on the distance to certain types of roads.

7.2.3. Granular and particle-laden flows: from laboratory experiments to field observations

Participant: Anne Mangeney.

In collaboration with R. Delannay, A. Valance, O. Roche and P. Richard.

A review article was written to provide an overview of dry granular flows and particle fluid mixtures, including experimental and numerical modelling at the laboratory scale, large scale hydrodynamics approaches and field observations. We also emphasize that the up-scaling from laboratory experiments to large scale geophysical flows still raises some theoretical physical challenges.

7.2.4. Continuum viscoplastic simulation of a granular column collapse on large slopes: $\mu(I)$ rheology and lateral wall effects

Participant: Anne Mangeney.

In collaboration with N. Martin, I. Ionescu, F. Bouchut and M. Farin.

We simulate here dry granular flows resulting from the collapse of granular columns on an inclined channel and compare precisely the results with laboratory experiments. The 2-D model is based on the so-called $\mu(I)$ rheology that induces a Drucker-Prager yield stress and a variable viscosity. We show that the use of a variable or a constant viscosity does not change significantly the results provided that these viscosities are of the same order. Finally, we observed that small-scale instabilities develop when refining the mesh.

7.3. Analysis of models in Fluid Mechanics

7.3.1. Analysis of the Riemann problem for a shallow water model with two velocities

Participants: Emmanuel Audusse, Edwige Godlewski, Martin Parisot.

In collaboration with N. Aguillon.

The question addressed in [24] is the hyperbolicity of a shallow water model with two velocities. The model is written in a nonconservative form and the analysis of its eigenstructure shows the possibility that two eigenvalues coincide. A definition of the nonconservative product is given which enables us to analyse the resonance and coalescence of waves. Eventually, we prove the well-posedness of the two dimensional Riemann problem with initial condition constant by half-plane.

7.3.2. Different formulations of an elliptic problem issued from geophysics

Participants: Cindy Guichard, Ani Miraçi, Yohan Penel, Jacques Sainte-Marie.

A simplified problem coming from [33] involving pressure and velocity unknowns is studied. Some weak formulations (conform or mixed) are derived and their well-posedness is analysed. These weak formulations are then discretised in a finite element framework with suitable discrete spaces.

7.4. Numerical methods for fluid flows

7.4.1. Kinetic entropy for the layer-averaged hydrostatic Navier-Stokes equations

Participants: Emmanuel Audusse, Marie-Odile Bristeau, Jacques Sainte-Marie.

In [26], the authors are interested in the numerical approximation of the hydrostatic free surface incompressible Navier-Stokes equations. By using a layer-averaged version of the equations, previous results obtained for shallow water system are extended. A vertically implicit / horizontally explicit finite volume kinetic scheme is designed that ensures the positivity of the approximated water depth, the well-balancing and a fully discrete energy inequality.

7.4.2. Numerical approximation of the 3d hydrostatic Navier-Stokes system with free surface

Participants: Marie-Odile Bristeau, Anne Mangeney, Jacques Sainte-Marie, Fabien Souillé.

In collaboration with S. Allgeyer, M. Vallée, R. Hamouda, D. Froger.

A stable and robust strategy is proposed to approximate incompressible hydrostatic Euler and Navier-Stokes systems with free surface. The idea is to use a Galerkin type approximation of the velocity field with piecewise constant basis functions in order to obtain an accurate description of the vertical profile of the horizontal velocity. We show that the model admits a kinetic interpretation, and we use this result to formulate a robust finite volume scheme for its numerical approximation.

7.4.3. Well balanced schemes for rotation dominated flows

Participants: Emmanuel Audusse, Do Minh Hieu, Yohan Penel.

In collaboration with P. Omnes.

In [27], we study the property of colocated Godunov type finite volume schemes applied to the linear wave equation with Coriolis source term. The purpose is to explain the bad behaviour of the classical scheme and to modify it in order to avoid accuracy issues around the geostophic equilibrium. We use tools from two communities: well-balanced schemes for the shallow water equation with topography and asymptotic preserving schemes for the low Mach model. CFL conditions that ensure the stability of fully discrete schemes are established. The extension to the nonlinear case is under study.

7.4.4. A two-dimensional method for a dispersive shallow water model

Participants: Nora Aïssiouene, Marie-Odile Bristeau, Anne Mangeney, Jacques Sainte-Marie.

In collaboration with C. Pares.

In [29], [6], we propose a numerical method for a two-dimensional dispersive shallow water system with topography [3]. A first approach in one dimension, based on a prediction-correction method initially introduced by Chorin-Temam has been presented in [33]. The prediction part leads to solving a shallow water system for which we use finite volume methods while the correction part leads to solving a mixed problem in velocity and pressure. From the variational formulation of the mixed problem proposed in [35], the idea is to apply a finite element method with compatible spaces to the two-dimensional problem on unstructured grids.

7.4.5. Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow

Participant: Martin Parisot.

Article [32] is devoted to the numerical resolution in multidimensional framework of a hierarchy of reduced models of the free surface Euler equations. An entropy-satisfying scheme is proposed for the monolayer dispersive models [40] and [3]. To illustrate the accuracy and the robustness of the strategy, several numerical experiments are performed. In particular, the strategy is able to deal with dry areas without particular treatment. A work in progress focuses on the adaptation of the entropy-satisfying scheme to the layerwise models proposed in [30].

7.4.6. A lateral coupling between river channel and flood plain with implicit resolution of shallow water equations

Participant: Martin Parisot.

In collaboration with S. Barthélémy, N. Goutal, M.H. Le, S. Ricci.

Multi-dimensional coupling in river hydrodynamics offers a convenient solution to properly model complex flow while limiting the computational cost and taking the advantage of most pre-existing models. The project aims to adapt the lateral interface coupling proposed in [39] to the implicit version and assess it with real data from the Garonne River.

7.4.7. The discontinuous Galerkin gradient discretisation

Participant: Cindy Guichard.

In collaboration with R. Eymard.

The Symmetric Interior Penalty Galerkin (SIPG) method, based on Discontinuous Galerkin approximations, is shown to be included in the Gradient Discretisation Method (GDM) framework. Therefore, it can take benefit from the general properties of the GDM, since we prove that it meets the main mathematical gradient discretisation properties on any kind of polytopal mesh. We illustrate this inheritance property on the case of the p-Laplace problem [13].

7.4.8. Gradient-based optimization of a rotating algal biofilm process

Participants: Pierre-Olivier Lamare, Jacques Sainte-Marie.

In collaboration with N. Aguillon, O. Bernard.

Here we focus on the optimal control of an innovative process where the microalgae are fixed on a support. They are thus successively exposed to light and dark conditions. The resulting growth can be represented by a dynamical system describing the denaturation of key proteins due to an excess of light. A PDE model of the Rotating Algal Biofilm is then proposed, representing local microalgal growth submitted to the time varying light. An adjoint-based gradient method is proposed to identify the optimal (constant) process folding and the (time varying) velocity of the biofilm.

7.4.9. Method of reflections

Participant: Julien Salomon.

In collaboration with G. Legendre, P. Laurent, G. Ciaramella, M. Gander, L. Halpern.

In [17], the authors carefully trace the historical development of the methods of reflections, give several precise mathematical formulations and an equivalence result with the alternating Schwarz method for two particles.

In [31], a general abstract formulation is proposed in a given Hilbert setting and the procedure is interpreted in terms of subspace corrections. The unconditional convergence of the sequential form is proven and a modification of the parallel one is proposed to make it unconditionally converging.

7.5. Modelling of environmental impacts and natural hazards

7.5.1. Numerical simulation of the 30–45 ka debris avalanche flow of Montagne Pelée volcano, Martinique: from volcano flank collapse to submarine emplacement

Participant: Anne Mangeney.

In collaboration with M. Brunet, L. Moretti, A. Le Friant, E.D. Fernandez Nieto, F. Bouchut.

We simulate here the emplacement of the debris avalanche generated by the last flank collapse event of Montagne Pelée volcano (30–45 ka), Martinique, Lesser Antilles. Our objective is to assess the maximum distance (i.e., runout) that can be reached by this type of debris avalanche as a function of the volume involved. This result provides new constraints on the emplacement processes of debris avalanches associated with these collapses which can drastically change the related hazard assessment such as the generated tsunami, in a region known for its seismic and volcanic risks.

7.5.2. Global sensitivity analysis and uncertainty quantification of on-road traffic emissions

Participant: Vivien Mallet.

In collaboration with R. Chen, V. Aguiléra, F. Cohn, D. Poulet, F. Brocheton.

Road traffic emissions of air pollutants depend on both traffic flow and vehicle emission factors. Global sensitivity analyses, especially the computation of Sobol' indices, were carried out for the traffic model and the air pollutant emissions. In the process, the traffic model was replaced by a metamodel, or surrogate model, in order to reduce the high computational burden. The results identified the most important input parameters. Furthermore, the uncertainties in traffic flow and pollutant emissions were quantified by propagating into the model the uncertainties in the input parameters.

7.5.3. Uncertainty quantification in atmospheric dispersion of radionuclides

Participants: Ngoc Bao Tran Le, Vivien Mallet.

In collaboration with I. Korsakissok, R. Périllat, A. Mathieu, D. Didier.

In collaboration with IRSN, we investigated the uncertainties of the atmospheric-dispersion forecasts that are used during an accidental release of radionuclides like the Fukushima disaster. In order to quantify the uncertainties, Monte Carlo simulations and calibrations were carried out and coupled with ensemble meteorological forecasts from the European Centre for Medium-Range Weather Forecasts.

7.5.4. Simulation of air and noise pollution at high resolution and large scale

Participant: Vivien Mallet.

In collaboration with C. Pesin, P. Béal.

We developed fast surrogates for urban pollution models that they can be applied at global scale while preserving the street resolution, the main physical constraints and the performance against observational data. The surrogate models are based on the original models, machine learning algorithms and observational data.

7.6. Software developments

7.6.1. Improvements in the FRESHKISS3D code

Participants: Marie-Odile Bristeau, Jacques Sainte-Marie, Fabien Souillé.

Several tasks have been achieved in the FRESHKISS3D software:

- Reworked unittests and basic continuous integration
- Optimized IO functions
- Added compatibility with new mesh format
- Added generic run script that only takes yaml data as input
- Added validation cases and new example scripts
- Added paraview integrated post-processing scripts
- Reworked API and online documentation with sphinx
- Simplified dependencies and upgraded python to 3.6
- Worked on new numerical schemes:
  - Added implicit scheme for vertical exchanges terms
  - Reworked vertical viscosity scheme
  - Added new fluxes computations
  - Fixed vrious bugs (second order, viscosity, water state law)
  - Added vertical settling scheme on tracer (suspension models)
- Added 3D interpolator
- Added lagrangian particle tracking with reflexions on boundaries
- C++ Non-hydrostatic code (Nora) converted in cython (80%)
- Developement of a « Vilaine » package designed for SAUR/IAV/ANGE project

7.6.2. Numerical simulation of Free Surface Navier Stokes equations with Telemac 3D

Participants: Emmanuel Audusse, Nicole Goutal.

In collaboration with P. Quemar, A. Decoene, O. Lafitte, A. Leroy, C.T. Phan.

This work takes place in a joint project with EDF-LNHE (Laboratoire national d'hydraulique et d'environnement). The aim of the project is to understand the limitation of the actual numerical solution of the free surface Navier Stokes equations with software TELEMAC 3D and to propose new ways to handle important points as the advective part, the divergence free constraint, the coupling between velocity and hydrostatic pressure or the boundary conditions. A study of the mild-slope equation is also performed in order to obtain comparison solutions.

CASTOR Project-Team

7. New Results

7.1. Mathematical theory of reduced MHD models

Participant: Hervé Guillard.

One of the fundamental model used for fusion plasma simulations is the magnetohydrodynamic (MHD) model. However, in practice, many theoretical and numerical works in this field use specific approximations of this model known as *reduced* MHD models. These models assume that in the presence of a strong magnetic field, the main dynamic reduces to incompressible motion in the plane perpendicular to the dominating magnetic field and to the propagation of Alfvén waves in the magnetic field direction. In the framework of the slab approximation for large aspect ratio tokamaks ($R/a >> 1$ where R and a are respectively the major and minor radius of the machine) we have studied last year the validity of this assumption using techniques coming from the asymptotic theory of hyperbolic equations with a large parameter. In particular, we have proved that the solutions of the full MHD system converge in a weak sense to the solutions of an appropriate reduced model even in the presence of ill-prepared initial data. This work continues with a tentative to relax the large aspect ratio assumption that is not verified in modern machines.

7.2. 2D C^1 triangular elements

Participants: Hervé Guillard, Ali Elarif.

In order to avoid some mesh singularities that arise when using quadrangular elements for complex geometries and flux aligned meshes, the use of triangular elements is a possible option that we have studied in the past years [30]. In particular, we have developped the geometric tools necessary for the construction of Powell-Sabin splines and have applied these methods for the approximation of some simple hyperbolic PDE systems (namely the Euler equation of fluid dynamics). The PhD thesis of Ali Elarif that has begun in october 2017 is devoted to the study of the applicability of these methods to more complex PDE models encountered in plasma physics and to an extension towards other triangular C^1 elements (Clough-Tocher elements).

7.3. Simulations of hydraulic jumps with a turbulent Shallow Water model

Participants: Hervé Guillard, Argiris Delis [Technical University of Crete, Greece], Yih-Chin Tai [National Cheng Kung University, Taiwan].

We have studied numerically an extension designed for turbulent flows of the shallow water model. The model is able to describe the oscillatory nature of turbulent hydraulic jumps and as such correct the deficiency of the classical shallow water equations. The model equations, originally developed for horizontal flow or flows occurring over small constant slopes, are straightforwardly extended here for modeling flows over non-constant slopes and numerically solved by a second-order well-balanced finite volume scheme. Further, a new set of exact solutions to the extended model equations are derived and several numerical tests are performed to validate the numerical scheme and its ability to predict the oscillatory nature of hydraulic jumps under different conditions. The comparisons with experiments performed at Tainan University are very satisfactory given the simplicity of the model [27].

7.4. Block-structured meshes

Participants: Hervé Guillard, Alexis Loyer, Adrien Loseille [Gamma3 team, Inria Saclay], Jalal Lakhlili [IPP Garching], Ahmed Ratnani [IPP Garching].

Due to the highly anisotropic character of strongly magnetized plasmas, a crucial point for numerical simulations is the construction of meshes that are aligned on the magnetic flux surfaces computed by Grad-Shafranov equilibrium solvers. In this work, we study an original method for the construction of flux aligned grids that respect the magnetic equilibrium topology and that can be applied to block-structured meshes using C^1 finite element methods (Hermite-Bézier/Cubic spline). This method relies on the analysis of the singularities of the magnetic flux function and the construction of the Reeb graph that allows the segmentation of the physical domain into sub-domains that can be mapped to a reference square domain. Once this domain decomposition has been done, the mapping of the sub-domain to reference patches can be done using integration along the streamlines of the flux function[23], [34]. This work is performed in the framework of the EoCoE European project (see section 8.2.1.2).

7.5. FEM-BEM coupling methods for Tokamak plasma axisymmetric free-boundary equilibrium computations in unbounded domains

Participants: Blaise Faugeras, Holger Heumann.

Incorporating boundary conditions at infinity into simulations on bounded computational domains is a repeatedly occurring problem in scientific computing. The combination of finite element methods (FEM) and boundary element methods (BEM) is the obvious instrument, and we adapt here for the first time the two standard FEM-BEM coupling approaches to the free-boundary equilibrium problem: the Johnson–Nédélec coupling and the Bielak–MacCamy coupling. We recall also the classical approach for fusion applications, dubbed according to its first appearance von-Hagenow–Lackner coupling and present the less used alternative introduced by Albanese, Blum and de Barbieri. We show that the von-Hagenow–Lackner coupling suffers from undesirable non-optimal convergence properties, that suggest that other coupling schemes, in particular Johnson–Nédélec or Albanese–Blum–de Barbieri are more appropriate for non-linear equilibrium problems. Moreover, we show that any of such coupling methods requires Newton-like iteration schemes for solving the corresponding non-linear discrete algebraic systems.

7.6. Optimal control of a coupled partial and ordinary differential equations system for the assimilation of polarimetry Stokes vector measurements in tokamak free-boundary equilibrium reconstruction with application to ITER

Participant: Blaise Faugeras.

The modelization of polarimetry Faraday rotation measurements commonly used in tokamak plasma equilibrium reconstruction codes is an approximation to the Stokes model. This approximation is not valid for the foreseen ITER scenarios where high current and electron density plasma regimes are expected. In this work a method enabling the consistent resolution of the inverse equilibrium reconstruction problem in the framework of non-linear free-boundary equilibrium coupled to the Stokes model equation for polarimetry is provided. Using optimal control theory we derive the optimality system for this inverse problem. A sequential quadratic programming (SQP) method is proposed for its numerical resolution. Numerical experiments with noisy synthetic measurements in the ITER tokamak configuration for two test cases, the second of which is an H-mode plasma, show that the method is efficient and that the accuracy of the identification of the unknown profile functions is improved compared to the use of classical Faraday measurements.
In the framework of JET Task T17-15, the method has been implemented in the new code NICE and equilibrium reconstruction studies using real JET measurements are currently being performed.

7.7. Equilibrium reconstruction with Equinox at JET

Participant: Blaise Faugeras.

Within the framework of JET Task T17-02 an update of the real-time equilibrium reconstruction code Equinox at JET in view of its coupling with the transport code Raptor has been performed. Mainly the computation of all the averaged geometric quantities which enter the transport equation have been added.

7.8. Equilibrium reconstruction within the framework of the European Integrated Tokamak Modelling WPCD project

Participants: Blaise Faugeras, Cédric Boulbe.

We have been involved in a benchmark study between the equilibrium reconstruction codes VACTH-EQUINOX, EQUAL and LIUQE on TCV equilibriums [EPS paper R. Coelho et al] The benchmark study lead us to include new functionalities to VACTH-EQUINOX such as the possibility to have an upper X-point. The adaptation of VACTH-EQUINOX to IMAS (Integrated Modelling & Analysis Suite), the ITER standard using IDS (Interface Data Structure) as data type, has been carried on. Equilibrium reconstructions using IMAS have been performed on real JET measurements and on the first recently available WEST measurements.

7.9. Coupling free boundary equilibrium code and magnetic controller on IMAS

Participants: Cedric Boulbe, Jakub Urban.

During the previous years, the free boundary equilibrium code CEDRES++ has been coupled to the transport solver ETS and a magnetic simulink controler using the Integrated Tokamak Modelling infrastructure (Project Eurofusion WPCD - Work Package Code Development). In 2017, we have started to port this tool on IMAS which is the integrated modelling infrastructure developped by ITER. CEDRES++ has been adapted to IMAS and has been coupled to a magnetic controller in the framework of the Eurofusion WPCD project. This activity has been a pilot project to test the C++ tools provided on IMAS.

7.10. An Automated Approach to Plasma Breakdown Design

Participants: Holger Heumann, Eric Nardon.

Plasma breakdown in a tokamak requires a large toroidal electric field E_ϕ and a low poloidal magnetic field B_p, i.e. a so-called *field null region*. The latter should remain as extended as possible for a sufficient duration (typically a few tens of ms), all the more if one operates at low E_ϕ (e.g. in ITER where $E_\phi = 0.3V/m$). Finding appropriate settings (i.e. premagnetization coils currents and voltage waveforms) to produce and maintain a good field null region is not a trivial task, in particular in the presence of highly conducting passive structures which make the problem dynamic. WEST is a good example of this situation, due to two toroidally continuous copper plates which have been added for vertical stabilization: indeed, the current in the plates ramps up fast when E_ϕ is applied, which tends to degrade the field null region.
Our automated approach to determining appropriate breakdown settings relies on a precise electromagnetic model of the machine (including the iron core) and solves a constrained optimization problem, where the objective function to be minimized quantifies the design goal: the averaged magnitude of B_p. After discretization we end up with *finite dimensional* convex constrained optimization problem, that can be solved efficiently with Sequential Quadratic Programming. The approach follows the lines of optimal control methods for plasma equilibria in [35] and [36].
The automated approach was already beneficial for obtaining first breakdowns in WEST during the initial launch in December 2016. The data collected during these breakdowns allowed for improving the electromagnetic model and the simulations reproduce now very well magnetic measurements and the shapes observed on the fast camera during the experiments.

7.11. A high order method for the approximation of integrals over simplicity defined hypersurfaces

Participants: Lukas Drescher, Holger Heumann, Kersten Schmidt.

We introduced a novel method to compute approximations of integrals over implicitly defined hypersurfaces. The new method is based on a weak formulation in L2(0,1) that uses the coarea formula to circumvent an explicit integration over the hypersurfaces. As such it is possible to use standard quadrature rules in the spirit of hp/spectral finite element methods, and the expensive computation of explicit hypersurface parametrizations is avoided. We derived error estimates showing that high order convergence can be achieved provided the integrand and the hypersurface defining function are sufficiently smooth.

7.12. FEMs on Composite Meshes for Tuning Plasma Equilibria in Tokamaks

Participants: Holger Heumann, Francesca Rapetti, Xiao Song.

We rely on a combination of different finite element methods on composite meshes, for the simulation of axisymmetric plasma equilibria in tokamaks. One mesh with Cartesian quadrilaterals covers the burning chamber and one mesh with triangles discretizes the region outside the chamber. The two meshes overlap in a narrow region around the chamber. This approach gives the flexibility to achieve easily and at low cost higher order regularity for the approximation of the flux function in the area that is covered by the plasma, while preserving accurate meshing of the geometric details in the exterior. The continuity of the numerical solution across the boundary of each subdomain is enforced by a mortar-like projection. Higher order regularity is very beneficial to improve computational tools for tokamak research. In [13], we showed that the numerical calculation of free boundary plasma equilibria highly benefits from approximating the poloidal flux through some higher regular FE functions in the interior of the limiter. In the present work we show how the composite meshes and higher regular finite element functions allow to single out snowflake configurations, that play an important role to mitigate heat load in divertors. Implementations and numerical test were carried out in FEEQS.M

7.13. Automating the design of tokamak experiment scenarios

Participants: Jacques Blum, Holger Heumann, Xiao Song.

The real-time control of plasma position, shape and current in a tokamak has to be ensured by the Poloidal Field (PF) system. A standard strategy is to feedback-control the currents in the PF coils in order to match reference currents, the latter being a combination of FeedForward (FF) and FeedBack (FB) terms. While the FB part allows a precise control, it can work only near the target and therefore the FF part is essential to "guide" the system, i.e. to approximately reach the target while remaining clear from hardware limits. An essential part of tokamak scenario design is therefore the construction of these FF waveforms. A tool for automatic FF waveforms optimisation (the inverse evolutive mode of the Free-Boundary Equilibrium [FBE] solver FEEQS.M) has been developed recently in the frame of a collaboration with IRFM, CEA. This tool reduces drastically the amount of human work needed to design optimized scenarios compatible with hardware limits. Xiao Song has performed first applications and validations of this tool on present and future machines such as WEST and ITER. This preliminary work focused on the choice of the cost-function and compared different choices for the same type of discharge. A second key aspect of this work was the treatment of inequality constraints via penalisation terms.

7.14. Higher order FEM for Free-Boundary Equilibrium in FEEQS

Participant: Holger Heumann.

We extended FEEQS.M to work with higher order finite element functions (polynomials of degree 1, ... 11, or Powell-Sabin spline finite element functions). This feature is currently available only for the static modes. The free-boundary aspect is addressed by a subdivision approach, that needs to be improved in the future, to increase the accuracy.

7.15. The two temperature MHD model

Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

The dynamics of plasma charged particles can be described by a two-fluid MHD model. This description considers a plasma as a mixture of ions fluid and electrons flow that are coupled by exchanged terms such as momentum transfer terms, ion and electron heating terms due to collisions, supplemented by the Maxwell's equations. This system is quite intricate so that it is usually reduced to more tractable models. We first derive the two-temperature model, the ideal and resistive MHD equations from the two-fluid MHD system, and show that they correspond to asymptotic regimes for weakly and strongly magnetized plasmas. We then propose a finite volume approximation to compute the solutions of these models in unstructured tessellations used to appropriate mesh the toroidal geometry of the tokamak, where flows the plasma. The formulation of the magnetic field as Euler potential ensures the divergence free constraint in cheap manner, while a relaxation scheme for the two temperature allows an accurate computation of the electron and ion temperatures.

7.16. Spectral element schemes for dispersive equations

Participants: Sebastian Minjeaud, Richard Pasquetti.

S. Minjeaud and R. Pasquetti have addressed the Korteweg-de Vries equation as an interesting model of high order PDE, in order to show that it is possible to develop reliable and effective schemes, in terms of accuracy, computational efficiency, simplicity of implementation and, if required, conservation of the lower invariants, on the basis of a (only) H^1-conformal Galerkin approximation, namely the Spectral Element Method (SEM). The proposed approach relies on the introduction of additional variables that can be trivially eliminated, because the SEM mass matrix is diagonal, thus allowing to define discrete high order differentiation operators. Highly accurate RK IMEX schemes are used in time, with implicit treatment of the third order term and explicit treatment of the convective one. While the conservation of the mass invariant is natural, the conservation of the energy invariant is enforced by interpolation between embedded IMEX schemes, with preservation of the time discretization accuracy. Applications to several test problems have shown the robustness and accuracy of the proposed method, that is *a priori* easily extensible to other PDEs and to multidimensional problems (See [38]).

7.17. Cubature nodes for spectral element methods on symplicial meshes

Participants: Richard Pasquetti, Francesca Rapetti.

In a recent JCP paper (see [37]), a higher order triangular spectral element method (TSEM) is proposed to address seismic wave field modeling. The main interest of this TSEM is that the mass matrix is diagonal, so that an explicit time marching becomes very cheap. In [16], R. Pasquetti and F. Rapetti have compared this cubature points based method to the Fekete-Gauss one, that makes use of Fekete points for interpolation and of Gauss points for quadrature. Moreover, they have proposed an extension of this cubature TSEM to address elliptic PDEs with non homogeneous Neumann or Robin boundary conditions. More recently, the cubature TSEM has been experimented with isoparametric mappings to consider the case of non polygonal computational domains. In any cases it turns out that the cubature TSEM compares well with the Fekete-Gauss one.

7.18. Validation of the Full MHD with Bohm Boundary conditions

Participants: Boniface Nkonga, Guido Huijsmans, Ashish Bhole.

We have implemented Bohm mach condition using penalty boundary integral over open field lines. We have added temperature dependent viscosity and parallel conductivity, Bohm condition on energy flux. The JET configuration has been considered with success to reach the equilibrium with Bohm conditions on the divertor. However, the solution needs to be improved at the separatrix close to the boundary by a proper numerical stabilization. These structures disappear with constant resistivity. First n=10 ballooning mode in JET has been considered and need to be compared to the reduced MHD results. In order to proceed to these comparisons, a model for the reduced model has been derived including modeling of the viscosity. Validations on circular geometries are on going.

7.19. Non-linear MHD simulations of QH-mode DIII-D plasmas : ITER high Q scenarios

Participants: Feng Liu, Guido Huijsmans, Alberto Loarte, Boniface Nkonga.

In nonlinear MHD simulations of DIII-D QH-mode plasmas it has been found that low n kink/peeling modes (KPMs) are unstable and grow to a saturated kink-peeling mode. The features of the dominant saturated KPMs, which are localized toroidally by non-linear coupling of harmonics, such as mode frequencies, density fluctuations and their effect on pedestal particle and energy transport, are in good agreement with the observations of the Edge Harmonic Oscillation (EHO) typically present in DIII-D QH-mode experiments. The non-linear evolution of MHD modes including both kink-peeling modes and ballooning modes, is investigated through MHD simulations by varying the pedestal current and pressure relative to the initial conditions of DIII-D QH-mode plasma. The edge current and pressure at the pedestal are key parameters for the plasma either saturating to a QH-mode regime or a ballooning mode dominant regime. The influence of E×B flow and its shear on the QH-mode plasma has been investigated. E×B flow shear has a strong stabilization effect on the medium to high-n modes but is destabilizing for the n=2 mode. The QH-mode extrapolation results of an ITER Q=10 plasma show that the pedestal currents are large enough to destabilize n=1-5 kink/peeling modes, leading to a stationary saturated kink-peeling mode.

7.20. Sharpening diffuse interfaces with compressible fluids on unstructured meshes

Participants: Alexandre Chiapolino, Richard Saurel, Boniface Nkonga.

Diffuse interface methods with compressible fluids, considered through hyperbolic multiphase flow models, have demonstrated their capability to solve a wide range of complex flow situations in severe conditions (both high and low speeds). These formulations can deal with the presence of shock waves, chemical and physical transformations, such as cavitation and detonation. Compared to existing approaches able to consider compressible materials and interfaces, these methods are conservative with respect to mixture mass, momentum, energy and are entropy preserving. Thanks to these properties they are very robust. However, in many situations, typically in low transient conditions, numerical diffusion at material interfaces is excessive. Several approaches have been developed to lower this weakness. In the present contribution, a specific flux limiter is proposed and inserted into conventional MUSCL type schemes, in the frame of the diffuse interface formulation of Saurel et al. (2009). With this limiter, interfaces are captured with almost two mesh points at any time, showing significant improvement in interface representation. The method works on both structured and unstructured meshes and its implementation in existing codes is simple. Computational examples showing method capabilities and accuracy are presented.

COFFEE Project-Team

6. New Results

6.1. A few words on the results of the year

- Analysis of models with constraints: existence of solutions for a multidimensional model [1].
- Analysis of the convergence of a particle approximation for a traffic flow model with constraints [2].
- Numerical study of a traffic flow model with constraints [3]
- Existence of martingale solutions for the stochastic isentropic Euler model [4]
- New approach to accelerate convergence of Newton's methods applied to highly nonlinear and heterogeneous porous media flow problems [5], [6]
- We provide the numerical analysis of Discrete Duality Finite Volume (DDFV) schemes on general meshes for the (linear) Stokes problem with Neumann boundary conditions (on a fraction of the boundary). We prove well-posedness for a stabilized version of the scheme and we derive some error estimates. Finally, our theoretical results are illustrated with numerical simulations and stabilized and unstabilized schemes are compared. [26]
- A new domain decomposition algorithm to couple a non-isothermal compositional liquid gas Darcy flow and a free gas flow occurring at the interface between the nuclear waste repository and the ventilation galleries. [16], [22]
- A new two-phase Darcy flow model in fractured porous medium with fractures represented as interfaces of co-dimension one coupled to the surrounding matrix. The model accounts accurately for highly discontinuous capillary pressures, for gravity in the fracture width and both for fractures acting as drains or barriers. [7], [18]
- Use of the code ComPASS at BRGM to study the hydrothermal system of Lamentin Bay in Martinique during the PhD thesis of Yannis Labeau from University of Martinique [14], [21]
- New explicit energy-momentum conserving scheme for Hamiltonian systems [28].
- L. Monasse obtained an ANR JCJC grant.
- Through the PhD of J. Llobell, progress have been made to set up new schemes on staggered grids for solving the Euler system of gas dynamics (full Euler equations, MUSCL version on MAC grids, Low Mach regimes).

FLUMINANCE Project-Team

7. New Results

7.1. Fluid motion estimation

7.1.1. Stochastic uncertainty models for motion estimation

Participants: Shengze Cai, Etienne Mémin, Musaab Khalid Osman Mohammed.

The objective consists here in relying on a stochastic transport formulation to propose a luminance conservation assumption dedicated to the measurement of large-scale fluid flows velocity. This formulation, relying on the modeling under location uncertainty principle developed in the team, has the great advantage to incorporate from the beginning an uncertainty on the unresolved (turbulent) motion measurement. This uncertainty modeled as a possibly inhomogeneous random field uncorrelated in time 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. It provides also a reinterpretation, in terms of uncertainty, of classical regularization functionals proposed in the context of motion estimation. Nevertheless, at variance to classical motion estimation methods, this approach enables to estimate the so-called regularization parameter, which is in our framework related to the variance of the unresolved component of motion component. The resulting parameter-free estimator has shown to outperform state-of-the-art results of the literature [14]. This kind of method is applied on turbulent flows and in the context of river hydrologic applications through a collaboration with the Irstea Lyon research group (HHLY). A method for the 3D reconstruction of the river plane has been also proposed in this context.This study is performed within the PhD thesis of Musaab Mohammed.

7.1.2. Surface Currents estimation from Shore-Based Videos

Participant: Pierre Derian.

A wavelet based motion estimator has been extended for the recovery of instantaneous fields of surface current from shore-based and unmanned aerial vehicle videos. This study published in [16] and [34] demonstrated clearly the high potential of this method in the nearshore, where the rapid development of webcams and drones offers a large amount of applications for swimming and surfing safety, engineering and naval security and research purpose, by providing quantitative information. This work has been conducted within a collaboration with the Legos laboratory.

7.1.3. Development of an image-based measurement method for large-scale characterization of indoor airflows

Participants: Dominique Heitz, Etienne Mémin, Romain Schuster.

The goal is to design a new image-based flow measurement method for large-scale industrial applications. From this point of view, providing *in situ* measurement technique requires: (i) the development of precise models relating the large-scale flow observations to the velocity; (ii) appropriate large-scale regularization strategies; and (iii) adapted seeding and lighting systems, like Hellium Filled Soap Bubles (HFSB) and led ramp lighting. This work conducted within the PhD of Romain Schuster in collaboration with the compagny ITGA has started in february 2016. The first step has been to evaluate the performances of a stochastic uncertainty motion estimator when using large scale scalar images, like those obtained when seeding a flow with smoke. The PIV characterization of flows on large fields of view requires an adaptation of the motion estimation method from image sequences. The backward shift of the camera coupled to a dense scalar seeding involves a large scale observation of the flow, thereby producing uncertainty about the observed phenomena. By introducing a stochastic term related to this uncertainty into the observation term, we obtained a significant improvement of the estimated velocity field accuracy [41].

7.1.4. 3D flows reconstruction from image data

Participants: Dominique Heitz, Cédric Herzet.

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.

This year, we continued our investigation on the problem of efficient volume reconstruction. During the last years, we have focussed our attention on several families of convex optimization algorithms allowing to accelerate the standard procedures encountered in the Tomo-PIV literature while accounting for the non-negativity and the sparsity of the sought solutions. So far, the assessment of the proposed algorithms were exclusively done on synthetic data. This year, we started the process of validating the proposed procedures on real experimental data.

We started through a collaboration with Irstea to study ensemble assimilation methods for the fast reconstruction of 3D tomo-PIV motion field. The approach relies on a simplified dynamics of the flow and is a generalization of one of the popular emerging flow reconstruction technique of the PIV community referred to as "Shake the box". The study will be developed within an Irstea post-doctoral funding.

7.1.5. Sparse-representation algorithms

Participant: Cédric Herzet.

The paradigm of sparse representations is a central concept 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. This year, we continued our investigations on “screening” methodologies, that is procedures allowing for the rapid identification of (some of) the zeros of the sought sparse vector. More specifically, we designed low-complexity procedures enabling to screen groups of atoms by only performing one single test. This work has been submitted to the IEEE international conference on acoustic, speech and signal processing (ICASSP).

7.2. Tracking, Data assimilation and model-data coupling

7.2.1. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

In this axis of work we are exploring 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 is experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we aim 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. A 4DVar assimilation technique based on the numerical code Incompact3D has been implemented for that purpose to control the inlet and initial conditions in order to reconstruct a turbulent wake flow behind an unknown obstacle. We are actually extending this first data assimilation technique to control the subgrid parameters. This study is performed in collaboration with Sylvain Laizet (Imperial College). In another axis of research, in collaboration with the CSTB Nantes centre and within the PhD of Yacine Ben Ali we will explore the definition of efficient data assimilation schemes for wind engineering. The goal will be here to couple Reynolds average model to pressure data at the surface of buildings. The final purpose will consist in proposing improved data-driven simulation models for architects.

7.2.2. Ensemble variational data assimilation of large-scale dynamics with uncertainty

Participant: Etienne Mémin.

This study is focused on the coupling of a large scale representation of the flow dynamics built from the location uncertainty principle 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 assessed the performance of an ensemble variational assimilation technique with direct image data observation. Good results have been obtained for simulation under location uncertainty of 1D and 2D shallow water models [26]. This open the way to the definition of efficient data assimilation schemes for the coupling of high resolution data with large scale dynamical system.

7.2.3. Reduced-order models for flows representation from image data

Participants: Mamadou Diallo, Dominique Heitz, Cédric Herzet, Etienne Mémin, Valentin Resseguier.

During the PhD thesis of Valentin Resseguier we proposed a new decomposition of the fluid velocity in terms of a large-scale continuous component with respect to time and a small-scale non continuous random component. Within this general framework, an uncertainty based representation of the Reynolds transport theorem and Navier-Stokes equations can be derived, based on physical conservation laws. This physically relevant stochastic model has been applied in the context of POD-Galerkin methods. This uncertainty modeling methodology provides a theoretically grounded technique to define an appropriate subgrid tensor as well drift correction terms. The pertinence of this stochastic reduced order model has been successfully assessed on several wake flows at different Reynold number. It has been shown to be much more stable than the usual reduced order model construction techniques. Beyond the definition of a stable reduced order model, the modeling under location uncertainty paradigm offers a unique way to analyse from the data of a turbulent flow the action of the small-scale velocity components on the large-scale flow [25]. Regions of prominent turbulent kinetic energy, direction of preferential diffusion, as well as the small-scale induced drift can be identified and analyzed to decipher key players involved in the flow. This study has been published in the journal of fluid mechanics. Note that these reduced order models can be extended to a full system of stochastic differential equation driving all the temporal modes of the reduced system (and not only the small-scale modes). This full stochastic system has been evaluated on wake flow at moderate Reynolds number. For this flow the system has shown to provide very good uncertainty quantification properties as well meaningful physical behavior with respect to the simulation of the neutral modes of the dynamics. This study described in the PhD of Valentin Resseguier will be soon submitted to a journal paper.

On the other hand, in the following of several approaches proposed by the team [49], [53], we continued our investigations on the estimation of deterministic reduced order model from partial observations. In this line of research, we proposed a Bayesian framework for the construction of reduced-order models from image data. Our framework combines observation and prior information on the system to reduce the model and takes into account the uncertainties on the parameters of the model. The proposed approach reduces to some well-known model-reduction techniques for complete observations (i.e., the observation operator can be inverted). A theoretical analysis of our methodology has been investigated in a simplified context (namely, the observations are supposed to be noiseless linear combinations of the state of the system). This result provides worst-case guarantees on the reconstruction performance which can be achieved by a reduced model built from the data. These contributions have led to publications in a journal [18] and a conference [33].

7.3. Analysis and modeling of turbulent flows and geophysical flows

7.3.1. Geophysical flows modeling under location uncertainty

Participants: Pierre Derian, Long Li, Etienne Mémin, Valentin Resseguier.

In this research axis we have devised a principle to derive representation of flow dynamics under uncertainty. Such an uncertainty is formalized through the introduction of a random term that enables 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 partially known forcing. Rigorously derived from a stochastic version of the Reynolds transport theorem [9], this framework, referred to as modeling under location uncertainty, encompasses several meaningful mechanisms for turbulence modeling. It indeed introduces without any supplementary assumption the following pertinent mechanisms for turbulence modeling: (i) a dissipative operator related to the mixing effect of the large-scale components by the small-scale velocity; (ii) a multiplicative noise representing small-scale energy backscattering; and (iii) a modified advection term related to the so-called *turbophoresis* phenomena, attached to the migration of inertial particles in regions of lower turbulent diffusivity.

In a series of papers we have shown how such modeling can be applied to provide stochastic representations of a variety of classical geophysical flows dynamics [24], [23], [22]. Numerical simulations and uncertainty quantification have been performed on Quasi Geostophic approximation (QG) of oceanic models. It has been shown that such models lead to remarkable estimation of the unresolved errors at variance to classical eddy viscosity based models. The noise brings also an additional degree of freedom in the modeling step and pertinent diagnostic relations and variations of the model can be obtained with different scaling assumptions of the turbulent kinetic energy (i.e. of the noise amplitude). The performances of such systems have been assessed also on an original stochastic representation of the Lorenz 63 derived from the modeling under location uncertainty [15]. In this study it has been shown that the stochastic version enabled to explore in a much faster way the region of the deterministic attractor. This effort has been undertaken within a fruitful collaboration with Bertrand Chapron (LOPS/IFREMER). In the PhD of Long Li, starting this year, we will continue this effort. The goal will be to propose relevant techniques to define or calibrate the noise term from data. In that prospect, we intend to explore statistical learning techniques.

7.3.2. Large eddies simulation models under location uncertainty

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

The models under location uncertainty recently introduced by Mémin (2014) [9] provide a new outlook on LES modeling for turbulence studies. These models are derived from a stochastic transport principle. The associated stochastic conservation equations are similar to the filtered Navier- Stokes equation wherein we observe a sub-grid scale dissipation term. However, in the stochastic version, an extra term appears, termed as "velocity bias", which can be treated as a biasing/modification of the large-scale advection by the small scales. This velocity bias, introduced artificially in the literature, appears here automatically through a decorrelation assumption of the small scales at the resolved scale. All sub-grid contributions for the stochastic models are defined by the small-scale velocity auto-correlation tensor. This large scale modeling has been assed and compared to several classical large-scale models on several flows, namely a flow over a circular cylinder at Re $\sim$ 3900 [32], a smooth channel flow at Re(tau) $\sim$ 395 [31] and Taylor-Green vortex flows at Reynolds 1600, 3000 and 5000 [20]. For all these flows the modeling under uncertainty has provided better results than classical large eddies simulation models. Within the PhD of Yacine Ben Ali we will explore with the CSTB Nantes centre the application of such models for the definition of Reynolds average simulation (RANS) models for wind engineering applications.

7.3.3. Singular and regular solutions to the Navier-Stokes equations (NSE) and relative turbulent models

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The common thread of this work is the problem set by J. Leray in 1934 : does a regular solution of the Navier-Stokes equations (NSE) with a smooth initial data develop a singularity in finite time, what is the precise structure of a global weak solution to the Navier-Stokes equations, and are we able to prove any uniqueness result of such a solution. This is a very hard problem for which there is for the moment no answer. Nevertheless, this question leads us to reconsider the theory of Leray for the study of the Navier-Stokes equations in the

whole space with an additional eddy viscosity term that models the Reynolds stress in the context of large-scale flow modelling. It appears that Leray's theory cannot be generalized turnkey for this problem, so that things must be reconsidered from the beginning. This problem is approached by a regularization process using mollifiers, and particular attention must be paid to the eddy viscosity term. For this regularized problem and when the eddy viscosity has enough regularity, we have been able to prove the existence of a global unique solution that is of class C^∞ in time and space and that satisfies the energy balance. Moreover, when the eddy viscosity is of compact support in space, uniformly in time, we recently shown that this solution converges to a turbulent solution to the corresponding Navier-Stokes equations when, the regularizing parameter goes to 0. These results are described in a paper that has been submitted to the journal Archive for Rational Mechanics and Analysis (ARMA).

In the same direction, we also finalized a paper in collaboration with L. Berselli (Univ. Pisa, Italy) about the well known Bardina's turbulent model. In this problem, we consider the Helmholtz filter usually used within the framework of Large Eddy Simulation. We carry out a similar analysis, by showing in particular that no singularity occurs for Bardina's model.

Another study in collaboration with B. Pinier, P. Chandramouli and E. Memin has been undertaken. This work takes place within the context of the PhD work of B. Pinier. We considered the standard turbulent models involving the Navier-Stokes equations with an eddy viscosity that depends on the Turbulent Kinetic Energy (TKE), coupled with a supplementary equation for the TKE. The problem holds in a 3D bounded domain, with the Manning law at the boundary for the velocity. We have modeled a flux condition at the boundary for the TKE. We prove that with these boundary conditions, the resulting problem has a distributional solution. Then a serie of numerical tests has been performed in a parallelepiped with a non trivial bottom, showing the accuracy of the model in comparison with a direct numerical simulation of the Navier-Stokes equations. This work will be submitted to a journal.

7.3.4. Turbulence similarity theory for the modeling of Ocean Atmosphere interface

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The Ocean Atmosphere interface plays a major role in climate dynamics. This interaction takes place in a thin turbulent layer. To date no sastifying universal models for the coupling of atmospheric and oceanic models exists. In practice this coupling is realized through empirically derived interaction bulks. In this study, corresponding to the PhD thesis of Benoit Pinier, we aim at exploring similarity theory to identify universal mean profile of velocity and temperature within the mixture layer. The goal of this work consists in exhibiting eddy viscosity models within the primitive equations. We will also explore the links between those eddy viscocity models and the subgrid tensor derived from the uncertainty framework studied in the Fluminance group. In that prospect, we have studied the impact of the introduction of a random modeling of the friction velocity on the classical wall law expression. This model derived within the modeling under location uncertainty principle enabled us to propose an improved model of the velocity profile with a clear formulation in particular in the buffer turbulent area between the viscous zone and the turbulent region. Preliminary results on chanel flows are very promising. We are actually assessing this model on turbulent boundary layer flow at high Reynold.

7.3.5. 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.

7.3.6. Numerical and experimental image and flow database

Participants: Pranav Chandramouli, 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 Pranav Chandramouli. The third challenge was to carried out a time resolved tomoPIV experiments in a turbulent wake flow. Then this data will be used to assess the performances of the 4DVar assimilation technique developed in the context of Pranav Chandramouli's PhD to reconstruct three-dimensional turbulent flows.

7.4. Visual servoing approach for fluid flow control

7.4.1. Closed-loop control of a spatially developing shear layer

Participants: Christophe Collewet, Johan Carlier.

This study aims at controlling one of the prototypical flow configurations encountered in fluid mechanics: the spatially developing turbulent shear layer occuring between two parallel incident streams with different velocities. Our goal is to maintain the shear-layer in a desired state and thus to reject upstream perturbations. In our conference IFAC paper (https://hal.inria.fr/hal-01514361) we focused on perturbations belonging to the same space that the actuators, concretely that means that we were only able to face perturbations of the actuator itself, like failures of the actuator. This year we enlarged this result to purely exogenous perturbations. An optimal control law has been derived to minimize the influence of the perturbation on the flow. To do that, an on-line estimation of the perturbation has been used. This work will be submitted to the upcoming IEEE Conference on Decision and Control. We have also generalized the works initiated during the post-doctoral stay of Tudor-Bogdan Airimitoaie (https://hal.archives-ouvertes.fr/hal-01101089) concerning the benefits of increasing the controlled degrees of freedom in the particular case of the heat equation. This work has been validated on the shear flow.

7.5. Coupled models in hydrogeology

7.5.1. Coupling of subsurface and seepage flows

Participants: Jocelyne Erhel, Jean-Raynald de Dreuzy.

Hillslope response to precipitations is characterized by sharp transitions from purely subsurface flow dynamics to simultaneous surface and subsurface flows. Locally, the transition between these two regimes is triggered by soil saturation. Here we develop an integrative approach to simultaneously solve the sub- surface flow, locate the potential fully saturated areas and deduce the generated saturation excess over- land flow. This approach combines the different dynamics and transitions in a single partition formulation using discontinuous functions. We propose to regularize the system of partial differential equations and to use classic spatial and temporal discretization schemes. We illustrate our methodology on the 1D hillslope storage Boussinesq equations. We first validate the numerical scheme on previous numerical experiments without saturation excess overland flow. Then we apply our model to a test case with dynamic transitions from purely subsurface flow dynamics to simultaneous surface and subsurface flows. Our results show that discretization respects mass balance both locally and globally, converges when the mesh or time step are refined. Moreover the regularization parameter can be taken small enough to ensure accuracy without suffering of numerical artefacts. Applied to some hundreds of realistic hillslope cases taken from Western side of France (Brittany), the developed method appears to be robust and efficient. This study performed within the H2MNO4 ANR project has been published in the journal Advances in Water Ressources [21].

7.5.2. Characterizations of Solutions in Geochemistry

Participant: Jocelyne Erhel.

We study the properties of a geochemical model involving aqueous and precipitation-dissolution reactions at a local equilibrium. By reformulating the model as an equivalent optimization problem, we prove existence and uniqueness of a solution. It is classical in thermodynamic to compute diagrams representing the phases of the system. We introduce here the new precipitation diagram that describes the mineral speciation in function of the parameters of the system. Using the polynomial structure of the problem, we provide characterizations and an algorithm to compute the precipitation diagram. Numerical computations on some examples illustrate this approach. This work, is part of the H2MNO4 initiative. It has been recently submitted to a journal for publication [45].

7.5.3. Reactive transport in fractured-porous media

Participants: Yvan Crenner, Jean-Raynald de Dreuzy, Jocelyne Erhel.

Fractures must be carefully considered for the geological disposal of radioactive wastes. They critically enhance diffusivity, speed up solute transport, extend mixing fronts, and in turn, modify the physico-chemical conditions of reactivity in the Excavation Damaged Zone (EDZ) of the galleries. On the other hand, the pyrite oxidation could be considered like the main reaction due to the diffusion of oxygen through the gallery. Moreover, we assume that this reaction is complete in these geological conditions. First, we propose a numerical explicit reactive transport model in a fractured medium for an overall reaction. Afterwards, we present simulations outputs of the pyrite-oxygen reaction in EDZ zone. This study supported by ANDRA has been published in a conference [27].

7.5.4. Reactive transfers for multi-phasic flows

Participants: Jocelyne Erhel, Bastien Hamlat.

This study focuses on the mathematical modeling of reactive transfers for multi-phasic flows in porous medium. This study supported by IFPEN has been presented in a conference paper [37].

7.6. Linear solvers

7.6.1. Variable s-step GMRES

Participants: Jocelyne Erhel, David Imberti.

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. 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. This work supported by the EoCoE grant has been published in a conference proceeding [38] and in the journal [28].

7.6.2. Krylov method applied to reactive transport

Participant: Jocelyne Erhel.

Reactive transport models couple advection dispersion equations with chemistry equations. If the reactions are at thermodynamic equilibrium, then the system is a set of partial differential and algebraic equations. After space and implicit time discretizations, a nonlinear system of equations must be solved at each time step. The Jacobian matrix of the nonlinear system can be written with a Kronecker product coupling transport and chemistry. Krylov methods are well-suited to solve such linear systems because the matrix vector product can be done efficiently. The main challenge is to design a preconditioning matrix. We propose here to use the special structure of the matrix. Preliminary experiments show that Krylov methods are much more efficient than a direct method which does not use the coupled structure. This work supported by ANDRA has been published at the occasion of an invited conference [28].

LEMON Team

6. New Results

6.1. Boundary conditions and Schwarz waveform relaxation method for linear viscous Shallow Water equations in hydrodynamics

In [1] we propose in the present work an extension of the Schwarz waveform relaxation method to the case of viscous shallow water system with advection term. We first show the difficulties that arise when approximating the Dirichlet to Neumann operators if we consider an asymptotic analysis based on large Reynolds number regime and a small domain aspect ratio. Therefore we focus on the design of a Schwarz algorithm with Robin like boundary conditions. We prove the well-posedness and the convergence of the algorithm.

6.2. Modeling and control of in-situ decontamination of large water resources

In [3] we address the problem of the optimal control of in situ decontamination of water resources. We review several modeling, simulation and optimization techniques for this problem and their results. We show the benefit of combining tools from finite dimensional optimal control theory and numerical simulations of hydrodynamics equations, for providing simple and efficient feedback strategies.

6.3. On nontraditional quasi-geostrophic equations

In [7] we work on nontraditional models where the so-called traditional approximation on the Coriolis force is removed. In the derivation of the quasi-geostrophic equations, we carefully consider terms in δ/ε, where δ (aspect ratio) and ε (Rossby number) are both small numbers. We provide here some rigorous crossed-asymptotics with regards to these parameters, prove some mathematical results and compare QHQG and QG models.

6.4. Source term closures in shallow water models with porosity

In [4] the validity of flux and source term formulae used in shallow water models with porosity for urban flood simulations is assessed by solving the two-dimensional shallow water equations over computational domains representing periodic building layouts. The models under assessment are the Single Porosity (SP), the Integral Porosity (IP) and the Dual Integral Porosity (DIP) models. 9 different geometries are considered. 18 two-dimensional initial value problems and 6 two-dimensional boundary value problems are defined. This results in a set of 96 fine grid simulations. Analysing the simulation results leads to the following conclusions: (i) the DIP flux and source term models outperform those of the SP and IP models when the Riemann problem is aligned with the main street directions, (ii) all models give erroneous flux closures when is the Riemann problem is not aligned with one of the main street directions or when the main street directions are not orthogonal, (iii) the solution of the Riemann problem is self-similar in space-time when the street directions are orthogonal and the Riemann problem is aligned with one of them, (iv) a momentum balance confirms the existence of the transient momentum dissipation model presented in the DIP model, (v) none of the source term models presented so far in the literature allows all flow configurations to be accounted for(vi) future laboratory experiments aiming at the validation of flux and source term closures should focus on the high-resolution, two-dimensional monitoring of both water depth and flow velocity fields.

6.5. Consistency and bicharacteristic analysis of integral porosity shallow water models

The Integral Porosity and Dual Integral Porosity two-dimensional shallow water models have been proposed recently as upscaled models for urban floods. Very little is known so far about their consistency and wave propagation properties. Simple numerical experiments show that both models are unusually sensitive to the computational grid. In the present paper, a two-dimensional consistency and characteristic analysis is carried out for these two models. In [5] the following results are obtained: (i) the models are almost insensitive to grid design when the porosity is isotropic, (ii) anisotropic porosity fields induce an artificial polarization of the mass/momentum fluxes along preferential directions when triangular meshes are used and (iii) extra first-order derivatives appear in the governing equations when regular, quadrangular cells are used. The hyperbolic system is thus mesh-dependent, and with it the wave propagation properties of the model solutions. Criteria are derived to make the solution less mesh-dependent, but it is not certain that these criteria can be satisfied at all computational points when real-world situations are dealt with.

6.6. Dual integral porosity shallow water model for urban flood modelling

With CPU times 2 to 3 orders of magnitude smaller than classical shallow water-based models , the shallow water equations with porosity are a promising tool for large-scale modelling of urban floods. In [6], a new model formulation called the Dual Integral Porosity (DIP) model is presented and examined analytically and computationally with a series of benchmark tests. The DIP model is established from an integral mass and momentum balance whereby both porosity and flow variables are defined separately for control volumes and boundaries, and a closure scheme is introduced to link control volume-and boundary-based flow variables. Previously developed Integral Porosity (IP) models were limited to a single set of flow variables. A new transient momentum dissipation model is also introduced to account for the effects of sub-grid scale wave action on porosity model solutions, effects which are validated by fine-grid solutions of the classical shallow-water equations and shown to be important for achieving self-similarity in dam-break solutions. One-dimensional numerical test cases show that the proposed DIP model outperforms the IP model, with signicantly improved wave propagation speeds, water depths and discharge calculations. A two-dimensional field scale test case shows that the DIP model performs better than the IP model in mapping the floods extent and is slightly better in reproducing the anisotropy of the flow field when momentum dissipation parameters are calibrated.

6.7. DG method for dispersive Green-Naghdi equations

Concerning the development of the WaveBox code, we have introduced in [2] the first available numerical code allowing to solve some fully nonlinear and weakly dispersive asymptotic shallow water models on unstructured meshes. More precisely, we introduce a discontinuous Finite Element formulation (discontinuous-Galerkin) on simplicial unstructured meshes for the study of free surface flows based on the fully nonlinear and weakly dispersive Green-Naghdi equations. Working with a new class of asymptotically equivalent equations, which have a simplified analytical structure, we consider a decoupling strategy: we approximate the solutions of the classical shallow water equations sup- plemented with a source term globally accounting for the non-hydrostatic effects and we show that this source term can be computed through the resolution of scalar elliptic second-order sub-problems, with a use of a L-DG method. The assets of the proposed discrete formulation are: (i) the handling of arbitrary unstructured simplicial meshes, (ii) an arbitrary order of approximation in space, (iii) the exact preservation of the motionless steady states, (iv) the preservation of the water height positivity, (v) a simple way to enhance any numerical code based on the nonlinear shallow water equations. To improve the efficiency of the resolution of the elliptic part of the formulation, we also investigate the use of very recent skeleton Hybrid-High-Order (HHO) methods. These methods allow to dramatically reduce the number of degrees of freedom (DOF), using only the DOF located on the mesh skeleton. To initiate the development of such methods for nonlinear and un-stationnary problems, a new discrete formulation was developed for the advective Cahn-Hilliard equations in [17]. Such an approach will be extended to more complex asymptotic shallow water models in a near future.

MAGIQUE-3D Project-Team

6. New Results

6.1. Seismic Imaging and Inverse Problems

6.1.1. Mathematical determination of the Fréchet derivative with respect to the domain for a fluid-structure scattering problem. Case of polygonal-shaped domains.

Participants: Hélène Barucq, Elodie Estecahandy.

The characterization of the Fréchet derivative of the elasto-acoustic scattered field with respect to Lipschitz continuous polygonal domains is established. The considered class of do- mains is of practical interest since two-dimensional scatterers are always transformed into polygonal- shaped domains when employing finite element methods for solving direct and inverse scattering problems. The obtained result indicates that the Fr´echet derivative with respect to the scatterer of the scattered field is the solution of the same elasto-acoustic scattering problem but with addi- tional right-hand side terms in the transmission conditions across the fluid-structure interface. This characterization has the potential to advance the state-of-the-art of the solution of inverse obstacle problems.
This work has been done in collaboration with Prof. Rabia Djellouli (California State University at Northridge) and has been accepted for publication in Siam Journal of Applied Mathematics [16].

6.1.2. Shape-reconstruction and parameter identification of an elastic object immersed in a fluid

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Julien Diaz, Kevin Lagnoux.

We have developed a procedure to reconstruct the shape and material parameters of an elastic obstacle immersed in a fluid medium from some external measurements given by the so called far-field pattern. It is a nonlinear and ill-posed problem which is solved by applying a Newton-like iterative method involving the Fréchet derivatives of the scattered field. These derivatives express the sensitivity of the scattered field with respect to the parameters of interest. They are defined as the solution of boundary value problems which differ from the direct one only at the right-hand sides level. We have been able to establish the well-posedness of each problem in the case of a regular obstacle and it would be interesting in the near future to extend those results to the case of scatterers with polygonal boundaries. It requires to work with less regular Sobolev spaces for which the definition of traces is not obvious. We have also provided an analytical representation of the Fréchet derivatives in the case of a circle. This provide a way of validating the numerical experiments and it would be interesting to obtain their expression in the case of elliptical scatterers or spherical ones. It is worth mentioning that this work has been done only in the case of isotropic media and it would be interesting to extend it to anisotropic media as well. It requires to establish analytic representations of the scattered field in anisotropic media which is more difficult because it involves more parameters.

We have studied the response of the data to the different parameters. It turns out that the sensitivity of the far field pattern is very different regarding the shape or the material parameters. We have delivered a sensitivity analysis which has been essential for understanding that the reconstruction of the material parameters is conditioned by the recovering of the shape parameters. This makes the full reconstruction very difficult and sometimes unstable. In particular, in the case of a disk-shaped obstacle, when addressing the role of the frequency in the reconstruction, we have been faced to the issue of the existence of Jones modes which had been already observed by Elodie Estecahandy in her PhD thesis. Next, we have introduced a series of numerical experiments that have been performed by applying two algorithms which propose two strategies of full reconstruction regarding the material parameters are retrieved simultaneously with the shape or not. It turns out that both work similarly delivering the same level of accuracy but the simultaneous reconstruction requires less iterations. We have thus opted for retrieving all the parameters simultaneously. Since realistic

configurations include noisy data, we have performed some simulations for the reconstruction of the shape along with the Lamé coefficients for different noise levels. Other interesting experiments have been carried out using a multistage procedure where the parameters of interest are the density of the solid interior, the shape of the obstacle and its position. We have considered the case of Limited Aperture Data in back-scattering configurations, using multiple incident plane waves, mimicing a physical disposal of non-destructive testing. This is an encouraging ongoing work which deserves to be completed by considering a wide range of examples including more general geometries of the scatterer. It should also be extended by dealing with limited aperture data using only one incident wave (which will probably require multiple frequency data).

These results have been obtained in collaboration with Rabia Djellouli (California State University at Northridge, USA) and were presented to the Waves 2017 conference.

6.1.3. Shape-reconstruction and parameter identification of an anisotropic elastic object immersed in a fluid

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Julien Diaz.

We extended the solution methodology for reconstructing the shape and material parameters of an elastic obstacle (see 6.1.2) to the case of anisotropic media. This is a very challenging case which still deserves further works. We have obtained some results but since the impact of some of the anisotropic parameters on the FFP is even weaker than the Lamé coefficients, the reconstruction of these parameters together with the shape parameters requires several frequencies and carefully adapted regularization parameters. It is in particular difficult to retrieve the Thomsen parameters ϵ and δ because their reconstruction requires to have an accurate adjustment on the rest of material and shape parameters. The recovery process is thus computationally intensive and some efforts should be done in the near future to decrease the computational costs. We were able to recover all the anisotropic parameters when the shape were assumed to be known. However, when trying to recover both shape and material parameters, we could only recover the shape and some of the physical parameters (namely the three most important ones : the density and the two velocities V_p and V_s). We should now find a way to determine all the Thomsen parameters together with the shape. Then, we will have to deal with more complex media such as TTI media (this will add the angle of anisotropy as additional parameter). The last step will be to consider general anisotropy, which could be done by recovering each element of the elastic stiffness tensor. This is simple to implement, since the derivative of the stiffness tensor with respect to one of its component is easily computable (it is a tensor composed of zeroes and ones). However, the stability of the reconstruction is not guaranteed, since we will strongly increase the number of components to be retrieved.

These results have been obtained in collaboration with Rabia Djellouli (California State University at Northridge, USA).

6.1.4. Mathematical analysis and solution methodology for a class of inverse spectral problems arising in the design of optical waveguides

Participant: Hélène Barucq.

We analyze mathematically the problem of determining refractive index profiles from some desired/measured guided waves propagating in optical fibers. We establish the uniqueness of the solution of this inverse spectral problem assuming that only one guided mode is known. Then, we propose an iterative computational procedure for solving numerically the considered inverse spectral problem. Numerical results are presented to illustrate the potential of the proposed regularized Newton algorithm to efficiently and accurately retrieve the refractive index profiles even when the guided mode measurements are highly noisy.

This work has been submitted for publication in a peer-reviewed journal. It has been done in collaboration with Rabia Djellouli (California State University at Northridge, USA) and Chokri Bekkey (University of Monastir, Tunisia)

6.1.5. Time-harmonic seismic inverse problem with Cauchy data

Participant: Florian Faucher.

This work is a collaboration with Giovanni Alessandrini (Università di Trieste), Maarten V. de Hoop (Rice University), Romina Gaburro (University of Limerick) and Eva Sincich (Università di Trieste).
We study the performance of Full Waveform Inversion (FWI) from time-harmonic Cauchy data via conditional well-posedness driven iterative regularization. The Cauchy data can be obtained with dual sensors measuring the pressure and the normal velocity. We define a novel misfit functional which, adapted to the Cauchy data, allows the independent location of experimental and computational sources. The conditional well-posedness is obtained for a hierarchy of subspaces in which the inverse problem with partial data is Lipschitz stable. Here, these subspaces yield piecewise linear representations of the wave speed on given domain partitions. Domain partitions can be adaptively obtained through segmentation of the gradient. The domain partitions can be taken as a coarsening of an unstructured tetrahedral mesh associated with a finite element discretization of the Helmholtz equation. We illustrate the effectiveness of the iterative regularization through computational experiments with data in dimension three. In comparison with earlier work, the Cauchy data do not suffer from eigenfrequencies in the configurations.
The resulting paper is [47] and is also connected to the following conference presentations, [36], [27].

6.1.6. Quantitative Convergence of Full Waveform Inversion in the Frequency Domain

Participants: Hélène Barucq, Florian Faucher.

This work is a collaboration with Guy Chavent (Inria Rocquencourt).
We study the convergence of the inverse problem associated with the frequency domain wave equations for the recovery of subsurface parameters. The numerical method selected for the resolution is the Full Waveform Inversion (FWI), which designs an iterative minimization algorithm. We study the convergence of the scheme in the context of least squares minimization. We establish numerical estimates based on the Fréchet derivatives for the radius of curvature and the deflection. We quantify the (complex) frequency progression to select to foster the convergence, and illustrate the effect of the subsurface geometry. From the curvature estimates, we also provide an insight of the robustness with noise depending on the situation. We supplement the numerical analysis with numerical experiments to demonstrate the results.
The results have been presented in the following conference, [36], [27], [26], [25].

6.1.7. Contributions to seismic full waveform inversion for time harmonic wave equations: stability estimates, convergence analysis, numerical experiments involving large scale optimization algorithms

Participants: Hélène Barucq, Florian Faucher.

In this project, we investigate the recovery of subsurface Earth parameters. We consider the seismic imaging as a large scale iterative minimization problem, and deploy the Full Waveform Inversion (FWI) method. The reconstruction is based on the wave equations because the characteristics of the measurements indicate the nature of the medium in which the waves propagate. First, the natural heterogeneity and anisotropy of the Earth require numerical methods that are adapted and efficient to solve the wave propagation problem. In this study, we have decided to work with the harmonic formulation, i.e., in the frequency domain.

The inverse problem is then established in order to frame the seismic imaging. It is a nonlinear and ill-posed inverse problem by nature, due to the limited available data, and the complexity of the subsurface characterization. However, we obtain a conditional Lipschitz-type stability in the case of piecewise constant model representation. We derive the lower and upper bound for the underlying stability constant, which allows us to quantify the stability with frequency and scale. It is of great use for the underlying optimization algorithm involved to solve the seismic problem. We review the foundations of iterative optimization techniques and provide the different methods that we have used in this project. The Newton method, due to the numerical cost of inverting the Hessian, may not always be accessible. We propose some comparisons to identify the benefits of using the Hessian, in order to study what would be an appropriate procedure regarding the accuracy and time. We study the convergence of the iterative minimization method, depending on different aspects such as the geometry of the subsurface, the frequency, and the parametrization. In particular, we quantify the frequency progression, from the point of view of optimization, by showing how the size of the basin of attraction evolves with frequency.

Following the convergence and stability analysis of the problem, the iterative minimization algorithm is conducted via a multi-level scheme where frequency and scale progress simultaneously. We perform a collection of experiments, including acoustic and elastic media, in two and three dimensions. The perspectives of attenuation and anisotropic reconstructions are also introduced.

6.1.8. Quantitative localization of small obstacles with single-layer potential fast solvers

Participants: Hélène Barucq, Florian Faucher, Ha Pham.

In this work, we numerically study the inverse problem of locating small circular obstacles in a homogeneous medium using noisy backscattered data collected at several frequencies. The main novelty of our work is the implementation of a single-layer potential based fast solver (called FSSL) in a Full-Waveform inversion procedure, to give high quality reconstruction with low-time cost. The efficiency of FSSL was studied in our previous works. We show reconstruction results with up to 12 obstacles in structured or random configurations with several initial guesses, all allowed to be far and different in nature from the target. This last assumption is not expected in results using nonlinear optimization schemes in general. For results with 6 obstacles, we also investigate several optimization methods, comparing between nonlinear gradient descent and quasi-Newton, as well as their convergence with different line search algorithms.
The resulting research report is [45].

6.2. Mathematical modeling of multi-physics involving wave equations

6.2.1. Atmospheric Radiation Boundary Conditions for the Helmholtz Equation

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé.

An article is to be published in M2AN, see [14]. This work offers some contributions to the numerical study of acoustic waves propagating in the Sun and its atmosphere. The main goal is to provide boundary conditions for outgoing waves in the solar atmosphere where it is assumed that the sound speed is constant and the density decays exponentially with radius. Outgoing waves are governed by a Dirichlet-to-Neumann map which is obtained from the factorization of the Helmholtz equation expressed in spherical coordinates. For the purpose of extending the outgoing wave equation to axisymmetric or 3D cases, different approximations are implemented by using the frequency and/or the angle of incidence as parameters of interest. This results in boundary conditions called Atmospheric Radiation Boundary Conditions (ARBC) which are tested in ideal and realistic configurations. These ARBCs deliver accurate results and reduce the computational burden by a factor of two in helioseismology applications. This work has been done in collaboration with Laurent Gizon and Michael Leguèbe (Max-Planck-Institut für Sonnensystemforschung, Gottingen, Germany).

6.2.2. Atmospheric radiation boundary conditions for high frequency waves in time-distance helioseismology

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé.

An article has been published in Astronomy and Astrophysics [22]. The temporal covariance between seismic waves measured at two locations on the solar surface is the fundamental observable in time-distance helioseismology. Above the acoustic cutoff frequency (5.3 mHz), waves are not trapped in the solar interior and the covariance function can be used to probe the upper atmosphere. We wish to implement appropriate radiative boundary conditions for computing the propagation of high-frequency waves in the solar atmosphere. We consider the radiative boundary conditions recently developed by Barucq et al. (2017) for atmospheres in which sound-speed is constant and density decreases exponentially with radius. We compute the cross-covariance function using a finite element method in spherical geometry and in the frequency domain. The ratio between first-and second-skip amplitudes in the time-distance diagram is used as a diagnostic to compare boundary conditions and to compare with observations. We find that a boundary condition applied 500 km above the photosphere and derived under the approximation of small angles of incidence accurately reproduces the 'infinite atmosphere' solution for high-frequency waves. When the radiative boundary condition is applied 2 Mm above the photosphere, we find that the choice of atmospheric model affects the time-distance diagram.

In particular, the time-distance diagram exhibits double-ridge structure when using a VAL atmospheric model. This is a collaboration with Damien Fournier, Laurent Gizon, Chris Hanson and Michael Leguèbe (Max-Planck-Institut für Sonnensystemforschung, Gottingen, Germany).

6.2.3. Computational helioseismology in the frequency domain: acoustic waves in axisymmetric solar models with flows

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé.

An article has been published in Astronomy and Astrophysics [23]. Context. Local helioseismology has so far relied on semi-analytical methods to compute the spatial sensitivity of wave travel times to perturbations in the solar interior. These methods are cumbersome and lack flexibility. Aims. Here we propose a convenient framework for numerically solving the forward problem of time-distance helioseismology in the frequency domain. The fundamental quantity to be computed is the cross-covariance of the seismic wavefield. Methods. We choose sources of wave excitation that enable us to relate the cross-covariance of the oscillations to the Green's function in a straightforward manner. We illustrate the method by considering the 3D acoustic wave equation in an axisymmetric reference solar model, ignoring the effects of gravity on the waves. The symmetry of the background model around the rotation axis implies that the Green's function can be written as a sum of longitudinal Fourier modes, leading to a set of independent 2D problems. We use a high-order finite-element method to solve the 2D wave equation in frequency space. The computation is embarrassingly parallel, with each frequency and each azimuthal order solved independently on a computer cluster. Results. We compute travel-time sensitivity kernels in spherical geometry for flows, sound speed, and density perturbations under the first Born approximation. Convergence tests show that travel times can be computed with a numerical precision better than one millisecond, as required by the most precise travel-time measurements. Conclusions. The method presented here is computationally efficient and will be used to interpret travel-time measurements in order to infer, e.g., the large-scale meridional flow in the solar convection zone. It allows the implementation of (full-waveform) iterative inversions, whereby the axisymmetric background model is updated at each iteration. This work is a collaboration with Aaron Birch, Damien Fournier, Laurent Gizon, Chris Hanson, Michael Leguèbe and Emanuele Papini(Max-Planck-Institut für Sonnensystemforschung, Gottingen, Germany) and with Thorsten Hohage (Gôttingen University, Germany).

6.2.4. The virtual workshop : towards versatile optimal design of musical wind instruments for the makers

Participants: Juliette Chabassier, Robin Tournemenne.

Our project aims at proposing optimization solutions for wind instrument making. Our approach is based on a strong interaction with makers and players, aiming at defining interesting criteria to optimize from their point of view. After having quantified those criteria under the form of a cost function and a design parameters space, we wish to implement state-of-the-art numerical methods (finite elements, full waveform inversion, neuronal networks, diverse optimization techniques...) that are versatile (in terms of models, formulations, couplings...) in order to solve the optimization problem. More precisely, we wish to take advantage of the fact that sound waves in musical instruments satisfy the laws of acoustics in pipes (PDE), which gives us access to the full waveform inversion technique, usable in harmonic or temporal regime. The methods that we want to use are attractive because the weekly depend on the chosen criterion, and they are easily adaptable to various physical situations (multimodal decomposition in the pipe, coupling with the embouchure, ...), which can therefore be modified a posteriori. The goal is to proceed iteratively between instrument making and optimal design (the virtual workshop) in order to get close to tone quality related and playability criteria.

6.2.5. Energy based model and simulation in the time domain of linear acoustic waves in a radiating pipe

Participants: Juliette Chabassier, Robin Tournemenne.

We model in the time domain linear acoustic waves in a radiating pipe without daming. The acoustic equations system in formulated in flow and pressure, which leads to a first order space time equations system. The radiation condition is also written as a first order in time equation, and is parametrized by two real coefficients. Moreover, an auxiliary variable is introduced at the radiating boundary. The choice of this variable is adapted to the considered source type in order to ensure the model stability by energy techniques, under some conditions on the radiating condition. We then propose a stable space time explicit discretization, which ensures the dissipation of a discrete energy. The novelty of the discretization lies, on the one hand, in the variational nature of the space approximation (which leads to artibrary order finite elements with no required matrix inversion), and on the other hand, on the definition of the auxiliary variable for any acoustic source type (which leads to the decay of a well defined energy). Finaly, we quantify the frequential domain of validity of the used radiation condition by comparison with theoretical and experimental models of the litterature. This is a collaboration with Morgane Bergot (Université Claude Bernard, Lyon 1).

6.2.6. Computation of the entry impedance of a dissipative radiating pipe

Participants: Juliette Chabassier, Robin Tournemenne.

Modeling the entry impedance of wind instruments pipes is essential for sound synthesis or instrument qualification. We study this modeling with the finite elements method in one dimension (FEM1D) and with the more classically used transfer matrix method (TMM). The TMM gives an analytical formula of the entry impedance depending on the bore (intern geometry of the instrument) defined as a concatenation of simple elements (cylinders, cones, etc). The FEM1D gives the entry impedance for any instrument geometry. The main goals of this work are to assess the viability of the FEM1D and to study the approximations necessary for the TMM in dissipative pipes. First, lossless Weber's equation in one dimension is studied with arbitrary radiation conditions. In this context and for cylinders or cones, the TMM is exact. We verify that the error made with FEM1D for fine enough elements is as small as desired. When we consider viscothermal losses, the TMM does not solve the classical Kirchhoff model because two terms are supposed constant. In order to overcome this model approximation, simple elements, on which are based the TMM, are decomposed into much smaller elements. The FEM1D does not necessitate any model approximation, and it is possible to show that it solves the dissipative equation with any arbitrarily small error. With this in hand, we can quantify the TMM model approximation error.

6.2.7. Hybrid discontinuous finite element approximation for the elasto-acoustics.

Participants: Hélène Barucq, Julien Diaz, Elvira Shishenina.

Discontinuous Finite Element Methods (DG FEM) have proven their numerical accuracy and flexibility. However, numerically speaking, the high number of degrees of freedom required for computation makes them more expensive, compared to the standard techniques with continuous approximation. Among the different variational approaches to solve boundary value problems there exists a distinct family of methods, based on the use of trial functions in the form of exact solutions of the governing equations. The idea was first proposed by Trefftz in 1926, and since then it has been largely developed and generalized. By its definition, Trefftz-DG methods reduce numerical cost, since the vari- ational formulation contains the surface integrals only. Thus, it makes possible exploration of the meshes with different geometry, in order to create more realistic application. Trefftz-type approaches have been widely used for time-harmonic problems, while their implementation is still limited in time domain. The particularity of Trefftz-DG methods applied to the time-dependent formulations consists in the use of space-time meshes. Even though it creates another computational difficulty, due to a dense form of the matrix, which represents the global linear system, the inversion of the full "space-time" matrix can be reduced to the inversion of one block-diagonal matrix, which corresponds to the interactions in time. In the present work, we develop a theory for solving the coupled elasto-acoustic wave propagation system. We study well-posedness of the problem, based on the error estimates in mesh-dependent norms. We consider a space-time polynomial basis for numerical discretization. The obtained numerical results are validated with analytical solutions. Regarding the advantages of the method, following properties have been proven by the numerical tests: high flexibility in the choice of basis functions, better order of convergence, low dispersion. These results have been obtained in collaboration with Henri Calandra (TOTAL) and have been published in a research report [43]. A paper has been submitted and a second one is being prepared.

6.2.8. Construction of stabilized high-order hybrid Galerkin schemes.

Participants: Hélène Barucq, Aurélien Citrain, Julien Diaz.

We have compared the perfomance of Discontinuous Galerkin Methods and Spectral Element Methods on academic benchmark and on realistic geophysical model in two dimensions. We have shown that, for a given accuracy, SEm on quadrilateral meshes could be 10 times faster than DGm, which justifies our strategy to consider SEm wherever it is possible to use quadrilateral/hexahedral cells. These first results have been presented in Matthias conference. Then, we have considered the SEM/DG coupling proposed for electromagnetics in [78] and we have implemented it in our acoustics code. We are now analyzing the performance of this strategy and we are extending it to deal with elastodynamic and elasto-acoustic coupling. The following steps will be the extension of the analysis to 3D dimensional problems and the application to realistic test case. The main bottleneck is obviously to the definition of an efficient strategy to couple tetrahedra and hexahedra. Indeed, if in the 2D case, the edges of both triangles and quadrilaterals are all segments the faces of tetrahedra are triangle while the faces of hexahedra are quadrilaterals. Hence, in 2D it sufficed to define integration on segment, while in 3D it will be necessary to consider integration of various polygon resulting of the intersection of triangle and quadrilaterals. Once this strategy is defined and implemented, we expect to be able to reduce the computational of the plateform that we develop jointly with Total by a factor between 5 and 10. These results have been obtained in collaboration with Henri Calandra (TOTAL) and Christian Gout (INSA Rouen).

6.2.9. Modeling of dissipative porous media.

Participants: Juliette Chabassier, Julien Diaz, Fatima Jabiri.

In this work we have considered the modeling of 1D acoustic wave propagation coupled with visco-thermical losses that occur in porous media. We have proposed a family of dissipative models from which we have been able to obtain a quasi-constant quality factor (which is an indicator of the dissipation as a function of the frequency). We have derived stability conditions on the parameters of the model thanks to an energy analysis and we have rewritten the problem of designing a quasi-constant quality factor as a contrained least-square optimization problem. The parameters to optimize are the parameters of the family of dissipative models and the constraints are the stability of the final model. We are now considering the extension of the family to more general formulations and to heterogeneous media, before tackling multidimensional problems. These results have been obtained during the Master internship of Fatima Jabiri, in collaboration with Sébastien Imperiale (Inria Project-Team M3DISIM)

6.2.10. Asymptotic models for the electric potential across a highly conductive casing. Application to the field of resitivity measurements.

Participants: Hélène Barucq, Aralar Erdozain, Victor Péron.

A configuration that involves a steel-cased borehole is analyzed, where the casing that covers the borehole is considered as a highly conductive thin layer. Asymptotic techniques are presented as the suitable tool for deriving reduced problems capable of dealing with the numerical issues caused by the casing when applying the traditional numerical methods. The derivation of several reduced models is detailed by employing two different approaches, each of them leading to different classes of models. The stability and convergence of these models is studied and uniform estimates are proved. The theoretical orders of convergence are supported by numerical results obtained with the finite element method. We develop an application to the field of resistivity measurements. The second derivative of the potential which solves a reduced model has been employed to recover the resistivity of rock formations. These results are in accordance with an experiment of Kaufmann for the reference solution and have been obtained in collaboration with David Pardo (UPV/EHU).

6.2.11. Semi-Analytical Solutions for the Electric Potential across a Highly Conductive Casing.

Participants: Hélène Barucq, Aralar Erdozain, Victor Péron.

A transmission problem for the electric potential is considered, where one part of the domain is a high-conductive casing. Semi-analytical solutions are derived for several asymptotic models. These asymptotic models are designed to replace the casing by appropriate impedance conditions in order to avoid numerical instabilities. A decomposition in Fourier series of the solution to these asymptotic models is characterized. As an application we reproduced successfully the experiment of Kaufmann, using his same parameters, but computing with a fourth order asymptotic model. This experiment allows to recover the resitivity of rock formations employing a second derivative of the potential along the vertical direction. These results have been obtained in collaboration with Ignacio Muga (Pontificia Universidad Catolica of Valparaiso).

6.2.12. Asymptotic modeling of the electromagnetic wave scattering problem by a small sphere perfectly conducting.

Participants: Justine Labat, Victor Péron, Sébastien Tordeux.

In the context of non-destructive testing in medical imaging or civil engineering, the detection of small heterogeneities can be a difficult task in three dimensional domains. The complexity for solving numerically the direct problem both in terms of computation time and memory cost is due to the small size of obstacles in comparison with the incident wavelength and the large size of the domain of interest. Then the fine mesh size makes unsuitable or too expensive the use of classical numerical methods type continuous and discontinuous finite element methods or boundary element methods. The use of reduced models allows to get an approximation of the exact solution at a certain accuracy with a lower cost. We develop a Matched Asymptotic Expansions method to solve a time-harmonic electromagnetic scattering problem by a small sphere. This method allows to replace the scatterer by an equivalent asymptotic point source. In practice, it consists in defining an approximate solution using multi-scale expansions over far and near fields, related in a matching area. When the scatterer is a sphere, we make explicit the asymptotic expansions until the second order of approximation, relatively to the sphere radius. Numerical results make evident the convergence rate with respect to the sphere radius. Reference solutions are analytical solutions computed thanks to Montjoie Code. This work has been presented in the *Caleta Numerica* seminar, Pontificia Universidad Catolica of Valparaiso, Chili [48].

6.2.13. Asymptotic Models and Impedance Conditions for Highly Conductive Sheets in the Time-Harmonic Eddy Current Model.

Participant: Victor Péron.

This work is concerned with the time-harmonic eddy current problem for a medium with a highly conductive thin sheet. We present asymptotic models and impedance conditions up to the second order of approximation for the electromagnetic field. The conditions are derived asymptotically for vanishing sheet thickness ϵ where the skin depth is scaled like ϵ. The first order condition is the perfect electric conductor boundary condition. The second order condition turns out to be a Poincaré-Steklov map between tangential components of the magnetic field and the electric field [49]. Numerical experiments have been performed to assess the accuracy of the second order model. Complementary simulations will be conducted to study the robustness with respect to the sheet conductivity and the convergence of the modelling error. These results have been obtained in collaboration with Mohammad Issa and Ronan Perrussel (LAPLACE, CNRS/IMPT/UPS, Univ. de Toulouse) and this work has been presented in the international conference ACOMEN 2017 [33].

6.2.14. Numerical robustness of single-layer method with Fourier basis for multiple obstacle acoustic scattering in homogeneous media

Participants: Hélène Barucq, Juliette Chabassier, Ha Pham, Sébastien Tordeux.

We investigate efficient methods to simulate the multiple scattering of obstacles in homogeneous media. With a large number of small obstacles on a large domain, optimized pieces of software based on spatial discretization such as Finite Element Method (FEM) or Finite Difference lose their robustness. As an alternative, we work with an integral equation method, which uses single-layer potentials and truncation of Fourier series to describe the approximate scattered field. In the theoretical part of the paper, we describe in detail the linear

systems generated by the method for impenetrable obstacles, accompanied by a well-posedness study. For the numerical performance study, we limit ourselves to the case of circular obstacles. We first compare and validate our codes with the highly optimized FEM-based software Montjoie. Secondly, we investigate the efficiency of different solver types (direct and iterative of type GMRES) in solving the dense linear system generated by the method. We observe the robustness of direct solvers over iterative ones for closely-spaced obstacles, and that of GMRES with Lower–Upper Symmetric Gauss–Seidel and Symmetric Gauss–Seidel preconditioners for far-apart obstacles.
This work has been published in the journal Wave Motion, [15] and is also connected to the following conference presentations, [31], [41].

6.2.15. A study of the Numerical Dispersion for the Continuous Galerkin discretization of the one-dimensional Helmholtz equation

Participants: Hélène Barucq, Ha Pham, Sébastien Tordeux.

This work is a collaboration with Henri Calandra (TOTAL).
Although true solutions of Helmholtz equation are non-dispersive, their discretizations suffer from a phenomenon called numerical dispersion. While the true phase velocity is constant, the numerical one changes with the discretization scheme, order and mesh size. In our work, we study the dispersion associated with classical finite element. For arbitrary order of discretization, without using an Ansatz, we construct the numerical solution on the whole $\mathbb{R}$, and obtain an asymptotic expansion for the phase difference between the exact wavenumber and the numerical one. We follow an approach analogous to that employed in the construction of true solutions at positive wavenumbers, which involves Z-transform, contour deformation and limiting absorption principle. This perspective allows us to identify the numerical wavenumber with the angle of analytic poles. Such an identification is useful since the latter (analytic poles) can be numerically evaluated by an algorithm, which then yields the value of numerical wavenumber.
This work is detailed in the research report [44].

6.3. Supercomputing for Helmholtz problems

6.3.1. Numerical libraries for hybrid meshes in a discontinuous Galerkin context

Participants: Hélène Barucq, Lionel Boillot, Aurélien Citrain, Julien Diaz.

Elasticus team code has been designed for triangles and tetrahedra mesh cell types. The first part of this work was dedicated to add quadrangle libraries and then to extend them to hybrid triangles-quadrangles (so in 2D). This implied to work on polynomials to form functions basis for the (discontinuous) finite element method, to finally be able to construct reference matrices (mass, stiffness, ...).

A complementary work has been done on mesh generation. The goal was to encircle an unstructured triangle mesh, obtained by third-party softwares, with a quadrangle mesh layer. At first, we built scripts to generate structured triangle meshes, quadrangle meshes and hybrid meshes (triangles surrounded by quadrangles). We are now able to couple unstructured triangle mesh with structured quadrangle mesh, and we are now working on the implementation of the coupling between Discontinuous Galerkin methods (for the triangles) and Spectral Element methods (for the quadrangles).

6.3.2. Hybridizable Discontinuous Galerkin methods for solving the elastic Helmholtz equations

Participants: Marie Bonnasse-Gahot, Julien Diaz.

The advantage of performing seismic imaging in frequency domain is that it is not necessary to store the solution at each time step of the forward simulation. Unfortunately, the drawback of the Helmholtz equations, when considering 3D realistic elastic cases, lies in solving large linear systems. This represents today a challenging task even with the use of High Performance Computing (HPC). To reduce the size of the global linear system, we developed a Hybridizable Discontinuous Galerkin method (HDGm). It consists in expressing the unknowns of the initial problem in function of the trace of the numerical solution on each face of the mesh cells. In this way the size of the matrix to be inverted only depends on the number of degrees of freedom on each face and on the number of the faces of the mesh, instead of the number of degrees of freedom on each cell and on the number of the cells of the mesh as we have for the classical Discontinuous Galerkin methods (DGm). The solution to the initial problem is then recovered thanks to independent elementwise calculation. These results have been published in [18]. This is a collaboration with Henri Calandra (Total) and Stéphane Lanteri (Inria Project Team Nachos)

6.3.3. Scalability of linear solvers for Hybridizable Discontinuous Galerkin methods

Participants: Marie Bonnasse-Gahot, Julien Diaz.

We coupled our HDG code with tested two linear solvers: a parallel sparse direct solver MUMPS (MUltifrontal Massively Parallel sparse direct Solver) and a hybrid solver MaPHyS (Massively Parallel Hybrid Solver) which combines direct and iterative methods. In the framework of the european project HPC4E, we analyzed the scalability of the two solvers on the plateform Plafrim We compared the performances of the two solvers when solving 3D elastic waves propagation over HDGm. These comparisons were presented at the 2017 EAGE Workshop on High Performance Computing for Upstream and at MATHIAS 2017 conferences. This is a collaboration with Henri Calandra (Total), Luc Giraud, Mathieu Kuhn (Inria Project-Team Hiepacs) and Stéphane Lanteri (Inria Project Team Nachos).

6.4. Hybrid time discretizations of high-order

6.4.1. Construction and analysis of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations.

Participants: Juliette Chabassier, Julien Diaz, Anh-Tuan Ha.

We submitted a paper to M2AN. This paper deals with the construction of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations. This scheme is obtained by replacing the inversion of a matrix, that comes naturally after using the technique of the Modified Equation on the second order Leap Frog scheme applied to dissipative linear wave equations, by an explicit approximation of its inverse. The stability of the scheme is studied first using an energy analysis, then an eigenvalue analysis. Numerical results in 1D illustrate the good behavior regarding space/time convergence and the efficiency of the newly derived scheme compared to more classical time discretizations. A loss of accuracy is observed for non smooth profiles of dissipation, and we propose an extension of the method that fixes this issue. Finally, we assess the good performance of the scheme for a realistic dissipation phenomenon in Lorentz's materials. This work has been done in collaboration with Sébastien Imperiale (Inria Project-Team M3DISIM) and Alain Anh-Tuan Ha (Internship at Magique 3D in 2016).

6.4.2. Higher-order optimized explicit Runge-Kutta schemes for linear ODEs

Participants: Hélène Barucq, Marc Duruflé, Mamadou N'Diaye.

In this work, we have constructed optimized explicit Runge-Kutta schemes for linear ODEs that we called Linear-ERK. Theses schemes can be applied to the following ODE

$$M_h \frac{dU}{dt} = K_h U + F(t)$$

where M_h is the mass matrix, K_h the stiffness matrix and $F(t)$ a source term. Linear-ERK schemes are constructed using polynomial stability functions which are obtained by maximizing the CFL number. We have considered a polynomial stability function based on the Taylor series expansion of an exponential function. Then, we have added extra terms beyond the terms of the Taylor expansion without changing the order of accuracy. The coefficients of those extra terms have been computed by optimizing the CFL number such that the stability region of the developed scheme includes a typical spectrum. This spectrum has been obtained by computing eigenvalues of the matrix $M_h^{-1}K_h$ for the wave equation solved on a square with Hybrid Discontinuous Formulation (HDG). The optimization is performed by using the algorithm developed by D. Ketcheson and coworkers. By proceeding this way, we have obtained optimized explicit schemes up to order 8. We have also determined the CFL number and the efficiency on the typical spectrum for each explicit scheme. We have provided algorithms to implement these schemes and numerical results to compare them.

This work is a chapter of the thesis defended by Mamadou N'diaye on December 8, 2017, under the joint supervision of Hélène Barucq and Marc Duruflé.

6.4.3. High-order locally implicit time schemes for linear ODEs

Participants: Hélène Barucq, Marc Duruflé, Mamadou N'Diaye.

In this work we have proposed a method that combines optimized explicit schemes and implicit schemes to form locally implicit schemes for linear ODEs, including in particular ODEs coming from the space discretization of wave propagation phenomena. This method can be applied to the following ODE

$$M_h \frac{dU}{dt} = K_h U + F(t)$$

Like in the local time-stepping developed by Grote and co-workers, the computational domain is split into a fine region and a coarse region. The matrix A_h is given as

$$A_h = M_h^{-1}K_h = A_h P + A_h(I - P)$$

where P is the projector on the fine region of the computational domain. Then the proposed locally implicit method is obtained from the combination of the A-stable implicit schemes we have developed in 2016 (Padé schemes or Linear-SDIRK schemes detailed in [17]) on the fine region and explicit schemes with optimal CFL number in the coarse region. The developed method has been used to solve the acoustic wave equation and we have checked the convergence in time of these schemes for order 4, 6 and 8.

This work has been presented at the Mathias annual Total seminar and is a chapter of the thesis defended by Mamadou N'diaye on December 8, 2017, under the joint supervision of Hélène Barucq and Marc Duruflé.

SERENA Project-Team

7. New Results

7.1. A posteriori stopping criteria for domain decomposition methods

Participants: Sarah Ali Hassan, Michel Kern, Martin Vohralík.

Publication: [45]

In [45] we propose a new method for stopping iterations in a domain decomposition (DD) algorithm. The approach is based on a posteriori error estimates, and builds estimators that distinguish between the (space and time) discretization errors and that caused by the DD iterations. This enables stopping the iterations as soon as the DD error is smaller than the discretization error. In practice, numerous unnecessary iterations can be avoided, as illustrated in Figure 1 (here we stop at iteration 17 in place of the usual 61, economizing 72 % iterations). The method has been extended to global-in-time domain decomposition and to nonlinear problems. This was the topic of the Ph.D. thesis of Sarah Ali Hassan.

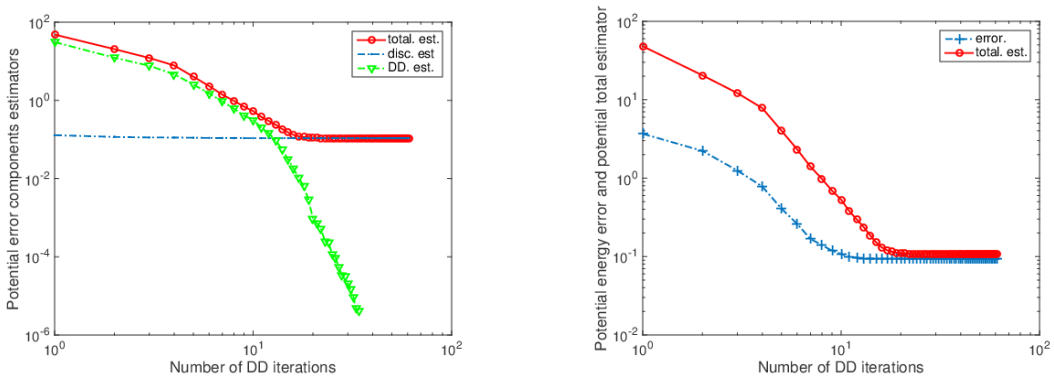


Figure 1. Error component estimates (left) and total energy error and its estimate (right), DD with GMRES solver

7.2. Finite element quasi-interpolation and best-approximation

Participant: Alexandre Ern.

Publication: [21]

In [21], we introduce a quasi-interpolation operator for scalar- and vector-valued finite element spaces constructed on affine, shape-regular meshes with some continuity across mesh interfaces. This operator gives optimal estimates of the best approximation error in any L^p-norm assuming regularity in the fractional Sobolev spaces $W^{r,p}$, where $p \in [1, \infty]$ and the smoothness index r can be arbitrarily close to zero. The operator is stable in L^1, leaves the corresponding finite element space point-wise invariant, and can be modified to handle homogeneous boundary conditions. The theory is illustrated on H^1-, $\mathbb{H}(\mathrm{curl})$-, and $\mathbb{H}(\mathrm{div})$-conforming spaces.

7.3. Hybrid High-Order methods for hyperelasticity

Participants: Alexandre Ern, Nicolas Pignet.

Publication: [13]

In [13], we devise and evaluate numerically Hybrid High-Order (HHO) methods for hyperelastic materials undergoing finite deformations. The HHO methods use as discrete unknowns piecewise polynomials of order $k \geq 1$ on the mesh skeleton, together with cell-based polynomials that can be eliminated locally by static condensation. The discrete problem is written as the minimization of a broken nonlinear elastic energy where a local reconstruction of the displacement gradient is used. Two HHO methods are considered: a stabilized method where the gradient is reconstructed as a tensor-valued polynomial of order k and a stabilization is added to the discrete energy functional, and an unstabilized method which reconstructs a stable higher-order gradient and circumvents the need for stabilization. Both methods satisfy the principle of virtual work locally with equilibrated tractions. We present a numerical study of the two HHO methods on test cases with known solution and on more challenging three-dimensional test cases including finite deformations with strong shear layers and cavitating voids. We assess the computational efficiency of both methods, and we compare our results to those obtained with an industrial software using conforming finite elements and to results from the literature. The two HHO methods exhibit robust behavior in the quasi-incompressible regime. In Figure 2 , we present some results for a hollow cylinder under shear and compression.

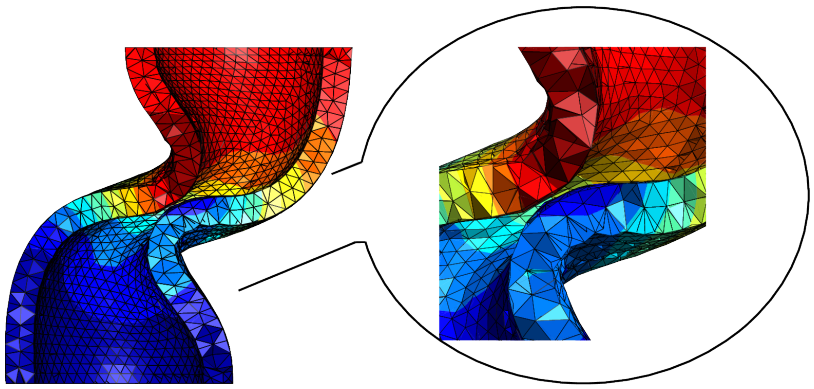


Figure 2. Euclidean displacement norm on the deformed configuration for the shear and compressed cylinder, and a zoom where the deformations are the most important. The color scale goes from 0.0 *(blue) to* 1.8 *(red).*

7.4. A nonlinear consistent penalty method for positivity preservation

Participant: Alexandre Ern.

Publication: [16]

In [16], we devise and analyze a new stabilized finite element method to solve the first-order transport (or advection-reaction) equation. The method combines the usual Galerkin/Least-Squares approach to achieve stability with a nonlinear consistent penalty term inspired by recent discretizations of contact problems to weakly enforce a positivity condition on the discrete solution. We prove the existence and the uniqueness of the discrete solution. Then we establish quasioptimal error estimates for smooth solutions bounding the usual error terms in the Galerkin/Least-Squares error analysis together with the violation of the maximum principle by the discrete solution. A numerical example is presented in Figure 3 .

7.5. A simple a posteriori estimate on general polytopal meshes

Participant: Martin Vohralík.

Figure 3. Elevations of solutions using piecewise quadratic elements. Left: standard method, the nodal discrete maximum principle violation is 21%. *Right: consistent penalty method, violation is less than* $4 \cdot 10^{-3}\%$.

Publication: [30]

The recent publication [30] develops an a posteriori error estimate for lowest-order locally conservative methods on meshes consisting of general polytopal elements. We focus here on the ease of implementation and evaluation cost of the methodology based on H^1-conforming potential reconstructions and $\mathbf{H}(\mathrm{div})$-conforming flux reconstructions that we develop in the SERENA project-team. In particular, the evaluation of our estimates for steady linear diffusion equations merely consists in some local matrix-vector multiplications, where, on each mesh element, the matrices are either directly inherited from the given numerical method, or easily constructed from the element geometry, while the vectors are the flux and potential values on the given element. This is probably the smallest computational price that one can imagine. We next extend our approach to steady nonlinear problems. We obtain a guaranteed upper bound on the total error in the fluxes that is still obtained by local matrix-vector multiplications, with the same element matrices as above. Moreover, the estimate holds true on any linearization and algebraic solver step and allows to distinguish the different error components. Finally, we apply this methodology to unsteady nonlinear coupled degenerate problems describing complex multiphase flows in porous media. It leads to an easy-to-implement and fast-to-run adaptive algorithm with guaranteed overall precision, adaptive stopping criteria, and adaptive space and time mesh refinements. An example of its application to a complex porous media flow (three-phases/three-components black-oil problem) can be found in Figure 4 .

7.6. Sharp algebraic and total a posteriori error bounds

Participant: Martin Vohralík.

Publication: [66]

In [66], we derive guaranteed, fully computable, constant-free, and sharp upper and lower a posteriori estimates on the algebraic, total, and discretization errors of finite element approximations of the Poisson equation obtained by an arbitrary iterative solver. Though guaranteed bounds on the discretization error, when the associated algebraic system is solved exactly, are now well-known and available, this is definitely not the case for the error from the linear algebraic solver (algebraic error), and a beautiful problem arises when these two error components interact. We try to analyze it here while identifying a decomposition of the algebraic error over a hierarchy of meshes, with a global residual solve on the coarsest mesh. Mathematically, we prove equivalence of our computable total estimate with the unknown total error, up to a generic polynomial-degree-independent constant. Numerical experiments illustrate sharp control of all error components and accurate prediction of their spatial distribution in several test problems, as we illustrate it in Figure 5 for the higher-order conforming finite element method and the conjugate gradient algebraic solver.

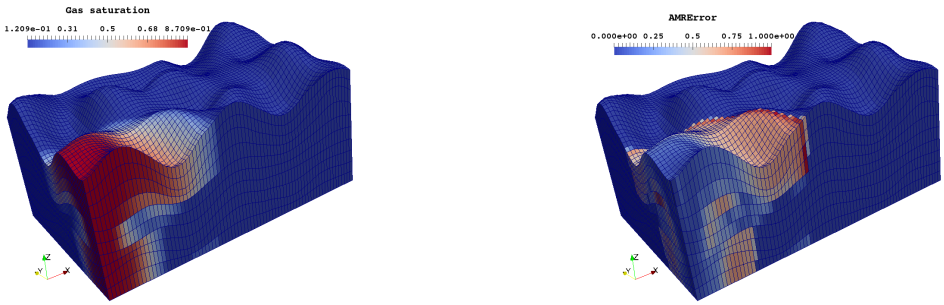


Figure 4. Simulated gas saturation after 1000 days (left) and corresponding a posteriori error estimate (right)

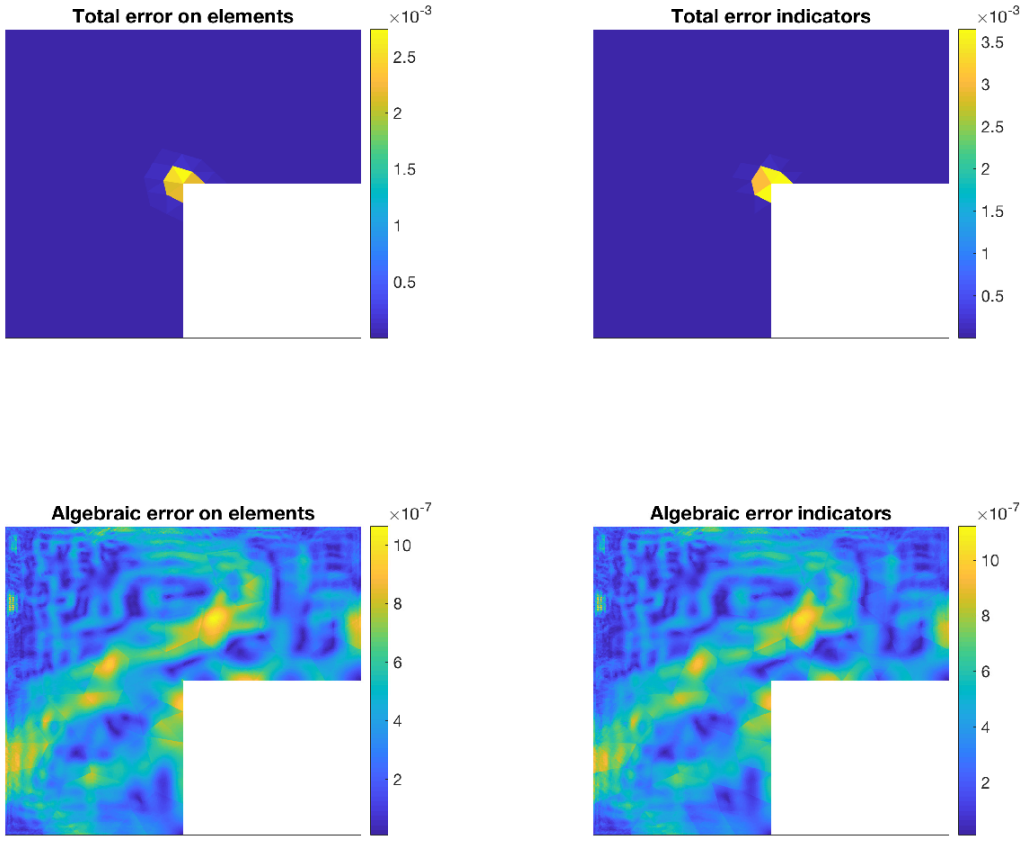


Figure 5. Actual total error (top left) and its a posteriori error estimate (top right). Actual algebraic error (bottom left) and its a posteriori error estimate (bottom right).

7.7. Analytic expressions of the solutions of advection-diffusion problems in 1D with discontinuous coefficients

Participant: Géraldine Pichot.

Publication: [64]

Grants: H2MN04 3

In [64], we provide a general methodology to compute the resolvent kernel as well as the density when available for a one-dimensional second-order differential operators with discontinuous coefficients. In a sequel, the computed resolvent kernel will be used to set-up an efficient and accurate simulation scheme.

STEEP Project-Team

7. New Results

7.1. Calculating spatial urban sprawl indices using open data

Urban sprawl has been related to numerous negative environmental and socioeconomic impacts. Meanwhile, urban areas have been growing at alarming rates, urging for assessing sprawl towards sustainable development. However, sprawl is an elusive term and different approaches to measure it have lead to heterogeneous results. Moreover, most studies rely on private/commercial data-sets and their software is rarely made public, impeding research reproducibility and comparability. Furthermore, many works give as result a unique value for a region of analysis, dismissing local spatial diversity that is vital for urban planners and policy makers.

Based on our last year's initial work [19], we have developed an extended open source framework for assessing urban sprawl using open data. Locations of residential and activity units are used to measure mixed use development and built-up dispersion, whereas the street network is used to measure accessibility between different land uses. Sprawl patterns are identified, and the resulting spatial information allows focusing on particular neighborhoods for a fine-grained analysis, as well as visualizing each sprawl dimension separately.

This work has been published in [10], [14], [15] and the associated implementation is available as open source software (see previous section).

7.2. A method for downscaling open population data

To extend our ongoing work on urban sprawl indicators (see above), we have developed a method to perform dissagregated population estimations at building level using open data. Our goal is to estimate the number of people living at the fine level of individual households by using open urban data and coarse-scaled population data. First, a fine scale description of residential land use per building is built using OpenStreetMap. Then, using coarse-scale gridded population data, we perform the down-scaling for each household given their containing area for residential usage. We rely solely on open data in order to ensure replicability, and to be able to apply our method to any city in the world, as long as sufficient data exists. The evaluation is carried out using fine-grained census block data for cities in France as ground-truth.

This work is published in [11] and the associated software implementation is made available as open source code at https://github.com/lgervasoni/urbansprawl.

7.3. Ecological accounting

The most important result obtained on this front bears on the quantification of the errors associated with the national road freight transport database (SITRAM). This database is informed year by year through a dedicated sampling campaign, but the errors associated with the various types of material goods transported have never been quantified. This was achieved by our team through the use of appropriate error estimators. This result is eagerly awaited by a number of scientific teams and public territorial agencies. Furthermore, the methodology that we have developed can easily be transposed to other countries. This result constitutes an important piece in the overall effort that the team has devoted to the question of the quantification of uncertainties in material flow analyses. This work is in the press in the leading journal in this field, the *Journal of Industrial Ecology*.

7.4. A computer framework for measuring urban land-use mix

The number of people living in cities has been increasing considerably since 1950, from 746 million to 3.9 billion in 2014, and more than 66% of the world's population are projected to live in urban areas by 2050. As this continuing population growth and urbanization are projected to add 2.5 billion people to the world's urban population in 30 years, this situation brings new challenges on how to conceive cities that host such amounts of population in a sustainable way. This sustainability question should address several aspects, ranging from economical to social and environmental matters among others. In this work, we focus on the formalization of a measure of mixed use development or land use mix in a city, i.e. how the structure of the city can help to provide a car-free sustainable living. Such type of land use mix has been largely proven to contain beneficial outcomes in terms of sustainability and to positively contribute to societal outcome, health, and public transportation among others. We developped a framework to compute mixed uses development index. A main caracteristic of our approach is to use only crowd-sourcing data (from OpenStreetMap) to extract the geo-localized land uses. Due to the universality of this data source, we are able to process any geographical area in the world, as long as sufficient data are available in OSM. A Kernel Density Estimation is performed for each of the land uses, outputing the spatial distribution of the different land uses. Based on this representation, a measure of land use mix is then calculated using the Entropy Index. The resulting GIS output shows enriched information for urban planners, supporting and aiding the decision-making procedure.

The framework, still in the phase of validation, was applied on the cities of London and Grenoble [19]. Future work includes integrating the LUM output for measuring the urban sprawl phenomenon and performing numerical interpretations of desirable mixed use values. We will also study the potential integration to transportation models, where land use mix correlation with the activities and residential uses can help to improve demand estimation. In addition, further investigation can be done by means of analyzing in detail the different types of activities. Finally, the estimation of LUM can be refined by taking into account, besides their location, the accessibility between different land uses, which is partly conditioned by the transportation infrastructure.

7.5. Calibration and sensitivity analysis for LUTI models

This year, we have consolidated our previous works on calibration of LUTI models, in particular of the Tranus model. The developed approaches are currently applied to instantiate a complete Tranus model for the Grenoble catchment area, in collaboration with AURG (Urban Planning Agency of the Grenoble area) and Brian Morton (U North Carolina) ; see [12], [13].

TONUS Team

7. New Results

7.1. Palindromic methods

7.1.1. Palindromic discontinuous Galerkin method in 2D and 3D

Participants: David Coulette, Florence Drui, Emmanuel Franck, Philippe Helluy, Laurent Navoret.

In the previous year (see [7]) we have proposed a method to solve hyperbolic systems like the Euler equations with an unconditionally stable high-order method. This method is based on a kinetic representation of the hyperbolic system. The kinetic equations are solved with an upwind DG method. It requires no matrix storage. High order is obtained through palindromic composition methods. The concept has been test in 1D. During this year we extend the method to 2D and 3D and applied it to fluid mechanics. Currently we are working on improving this method on realistic cases for MHD instabilities. The objective is to compare the results with the European code JOREK.

We are also working on methods for applying boundary conditions in a stable way with the palindromic method (postdoc of Florence Drui).

7.1.2. Kinetic model for palindromic methods

Participants: David Coulette, Emmanuel Franck, Laurent Navoret.

One of the most important drawbacks of the Palindromic method is the numerical dispersion associated to the high-order time scheme. To limit this problem we propose to replace the DG method by a semi-Lagrangian method and design new kinetic representations which are more accurate. We also studied the stability of these news models. The first results were good and currently we are working on the 2D extension and the coupling with limiter technics.

7.1.3. Finite element relaxation methods for fluid models

Participants: David Coulette, Emmanuel Franck.

In parallel to our work on the Palindromic method based on a kinetic relaxation model, we studied in [17] a variant based on the Xin-Jin relaxation model. Coupled with a finite element method we obtain an implicit solver for Euler equations where we invert only Laplacians and mass matrices. The first results show that the method is more efficient in CPU costs and memory. The finite elements used are the same as in JOREK.

7.2. MHD problems

Participant: Emmanuel Franck.

7.2.1. Compatible Implicit finite element for linear MHD

In this work we consider a linear MHD problem. The aim is to design an implicit method able to preserve the energy equation and the divergence free constraints in realistic Tokamak geometry. The first idea is to use a splitting scheme between the wave and convection parts coupled with an implicit scheme for each subsystem. In order to discretize each sub-system we use compatible B-Splines FE method wich allows us to preserve the invariants and to use a reduction of the implicit problem to be inverted. The idea was improved on simple geometries. We are currently extending the method on realistic geometries.

7.2.2. Splitting and relaxation for JOREK code

The Jorek code is the main European code for the simulation of Tokamak instabilities. The inversion of the full matrix is based on Block Jacobi preconditioning which is not efficient in some cases and very greedy in memory. We are investigating a new splitting scheme similar to the one used in works on compatible Finite Elements. We have also just begun to investigate the relaxation method used in the Palindromic scheme to solve the reduced MHD model of JOREK.

7.3. Finite Volume approximations of the Euler system with variable congestion

Participants: Pierre Degond, Piotr Minakowski, Laurent Navoret, Ewelina Zatorska.

We are interested in the numerical simulations of the Euler system with variable congestion encoded by a singular pressure. This model describes for instance the macroscopic motion of a crowd with individual congestion preferences. In [3] we propose an asymptotic preserving (AP) scheme based on a conservative formulation of the system in terms of density, momentum and density fraction. A second order accuracy version of the scheme is also presented. We validate the scheme on one-dimensional test cases and compare it with a scheme developed in a previous work and extended here to higher order accuracy. We finally carry out two-dimensional numerical simulations and show that the model exhibits typical crowd dynamics.

7.4. Numerical scheme for sheath equilibria

Participants: Mehdi Badsi, Michel Mehrenberger, Laurent Navoret.

We are interested in developing a numerical method for capturing stationary sheaths that a plasma forms in contact with a metallic wall. This work is based on a bi-species (ion/electron) Vlasov-Ampère model proposed in [19]. The main question addressed in this work is to know if classical numerical schemes can preserve stationary solutions with boundary conditions, since these solutions are not a priori conserved at the discrete level. In the context of high-order semi-Lagrangian method, due to their large stencil, interpolation near the boundary of the domain also requires a specific treatment. Moreover, for preventing instabilities from developing in large time, the proposed method guaranties that the discrete Gauss equation is satisfied in time.

7.5. Recurrence phenomenon for finite element grid based Vlasov solver

Participants: Michel Mehrenberger, Laurent Navoret, Thi Nhung Pham.

When using a grid based solver (finite element/DG scheme, discontinuous Galerkin semi-Lagrangian scheme) and spatial periodic boundary conditions, the simulations of the Vlasov-Poisson system exhibit numerical reappearance of initial perturbations at some time called recurrence time. This time depends on the numerical parameters (degree and mesh size of the finite element mesh). With a given number of degrees of freedom, considering a large degree approximation makes the phenomenon appear earlier in the simulation and thus makes this choice less attractive. In our work [9], we highlight that the time and the intensity of the recurrence are related to the quadrature rules used for computing the charge density. In particular, quadratures that are exact on trigonometric polynomials weaken the recurrence effect.

7.6. PICSL

Participants: Yann Barsamian, Joackim Bernier, Sever Hirstoaga, Michel Mehrenberger.

7.6.1. Particle in Cell and Semi-Lagrangian schemes for two species plasma simulations

Thanks to a classical first order dispersion analysis, we are able to check the validity of 1D$\times$1D two species Vlasov-Poisson simulations; the extension to second order is performed and shown to be relevant for explaining further details. In order to validate multidimensional effects, we propose in [14] a 2D $\times$ 2D single species test problem that has true 2D effects coming from the sole second order dispersion analysis. Finally, we perform, in the same code, full 2D×2D nonlinear two species simulations with mass ratio and consider the mixing of semi-Lagrangian and Particle-in-Cell methods. This work has been initiated at CEMRACS 2016.

7.7. TARGET

Participants: Nicolas Bouzat, Guillaume Latu, Camilla Bressan, Michel Mehrenberger, Virginie Grandgirard.

7.7.1. TArgeting Realistic GEometry in Tokamak code gysela

The framework of the work in [16] is the Semi-Lagrangian setting for solving the gyrokinetic Vlasov equation and the Gysela code. A new variant for the interpolation method is proposed that can handle the mesh singularity in the poloidal plane at $r = 0$ (a polar system is used for the moment in Gysela). A non-uniform meshing of the poloidal plane is proposed, instead of a uniform one, in order to save memory and computations. The interpolation method, the gyroaverage operator, and the Poisson solver are revised in order to cope with non-uniform meshes. A mapping that establishes a bijection from polar coordinates to more realistic plasma shapes is used to improve the realism. Convergence studies are provided to establish the validity and robustness of our new approach. This work has been initiated at CEMRACS 2016.

7.8. Field-aligned interpolation for gyrokinetics

Participants: Yaman Güclü, Philippe Helluy, Guillaume Latu, Michel Mehrenberger, Laura Mendoza, Eric Sonnendrücker, Maurizio Ottaviani.

This work is devoted to the study of field-aligned interpolation in semi-Lagrangian codes. This work has been initiated in 2013; this year the article has been accepted [5]. In the context of numerical simulations of magnetic fusion devices, this approach is motivated by the observation that gradients of the solution along the magnetic field lines are typically much smaller than along a perpendicular direction. In toroidal geometry, field-aligned interpolation consists of a 1D interpolation along the field line, combined with 2D interpolations on the poloidal planes (at the intersections with the field line). A theoretical justification of the method is provided in the simplified context of constant advection on a 2D periodic domain: unconditional stability is proven, and error estimates are given which highlight the advantages of field-aligned interpolation. The same methodology is successfully applied to the solution of the gyrokinetic Vlasov equation, for which we present the ion temperature gradient (ITG) instability as a classical test case: first we solve this in cylindrical geometry (screw-pinch), and next in toroidal geometry (circular Tokamak). In the first case, the algorithm is implemented in Selalib (semi-Lagrangian library), and the numerical simulations provide linear growth rates that are in accordance with the linear dispersion analysis. In the second case, the algorithm is implemented in the Gysela code, and the numerical simulations are benchmarked with those employing the standard (not aligned) scheme. Numerical experiments show that field-aligned interpolation leads to considerable memory savings for the same level of accuracy; substantial savings are also expected in reactor-scale simulations.

We are also currently implementing into SCHNAPS a general transport solver for addressing non-conforming patches in complex geometries. The objective is to be able to design meshes that are able to deal with magnetic aligned geometries. The resulting scheme will be used for solving kinetic equations, of course. But it can also be the building block of a palindromic method applied on curved and non-conforming meshes.

7.9. InKS

Participants: Olivier Aumage, Julien Bigot, Ksander Ejjaaouani, Michel Mehrenberger.

7.9.1. A programming model to decouple performance from semantics in simulation codes

Existing programming models lead to a tight interleaving of semantics and computer optimization concerns in high-performance simulation codes. With the increasing complexity and heterogeneity of supercomputers this requires scientists to become experts in both the simulated domain and the optimization process and makes the code difficult to maintain and port to new architectures. The report in [12] proposes InKS, a programming model that aims to improve the situation by decoupling semantics and optimizations in code so as to ease the collaboration between domain scientists and experts in high-performance optimizations. We define the InKS language that enables developers to describe the semantics of a simulation code with no concern for performance. We describe the implementation of a compiler able to automatically execute this code without making any explicit execution choice. We also describe a method to manually specify these choices to reach high-performance. Our preliminary evaluation on a 3D heat equation solver demonstrates the feasibility of the automatic approach as well as the ability to specify complex optimizations while not altering the semantic part. It shows promising performance where two distinct specifications of optimization choices in InKS offer similar performance as existing hand-tailored versions of the solver.

7.10. Performance of Particle-in-Cell methods

Participants: Yann Barsamian, Sever Hirstoaga, Eric Violard.

In a two-dimensional framework, in [6] we optimized a Particle-in-Cell (PIC) code by analyzing different data structures for the particles and for the grid fields with the aim of improving the cache reuse and by using the vectorization from the compiler. We also parallelized the code with OpenMP/MPI and satisfactory strong and weak scaling up to 8192 cores were obtained on the supercomputer CURIE.
Currently [15] we are extending and improving this work to a three-dimensional electrostatic PIC code.

7.11. Comparison of multiscale PIC methods

Participants: Nicolas Crouseilles, Sever Hirstoaga, Xiaofei Zhao.

In [2] we study different types of multiscale methods to numerically study the long-time Vlasov–Poisson equation with a strong magnetic field. The multiscale methods are an asymptotic preserving Runge–Kutta scheme, an exponential time differencing scheme, the stroboscopic averaging method and a uniformly accurate two-scale formulation. Extensive numerical experiments are conducted to investigate and compare the accuracy, efficiency, and long-time behavior of all the methods. The methods with the best performance under different parameter regimes are identified.

BIOCORE Project-Team

7. New Results

7.1. Mathematical methods and methodological approach to biology

7.1.1. Mathematical analysis of biological models

7.1.1.1. Mathematical study of semi-discrete models

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, 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 [67]. We used such models and analyzed their properties in several practical situations that are developed in Section 7.2.3 , 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, through the introduction of artificial habitats for the predatores or through the introduction of biological control agents in deterministic [13] or stochastic fashion. This was the main topic of Nicolas Bajeux's PhD thesis [11].

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 the study of the durability of plant resistance to root-knot nematodes [28], [41].

7.1.1.2. Model reduction and sensitivity analysis

Participants: Suzanne Touzeau, Jean-Luc Gouzé, Stefano Casagranda, Valentina Baldazzi.

Analysis and reduction of biochemical models. Dynamic models representing complex biological systems with numerous interactions can reach high dimensions and include complex nonlinearities. A model reduction method based on process weighing and pruning was developed [57] and implemented on various models [39]. A global sensitivity analysis was performed to check the method robustness against parameter uncertainty and variability. A more general method robust to initial conditions has been elaborated. This work was part of Stefano Casagranda's PhD thesis [12] and is also a collaboration with Bayer (Sophia-Antipolis).

7.1.1.3. Estimation and control

Participants: Suzanne Touzeau, Natacha Go, Jean-Luc Gouzé.

Parameter identification in complex systems. In complex biological systems, especially when data are scarce, identifying the model parameters is a challenge and raises identifiability issues. So we developed a specific ABC-like method, less computationally expensive than standard Bayesian fitting procedures such as ABC [6]. We used this method to fit a within-host immunological model to a large data set of individual viremia profiles. Our aim was not to reproduce individual profiles, but to identify several parameter sets compatible with the data and reflecting the variability among individuals. So we based our fitting criterion on viral indicators rather than the whole viremia dynamics [44]. This work was part of Natacha Go's post-doctorate, supported by the MIHMES project, in collaboration with the Roslin Institute, Edinburgh, UK. It benefited from the resources and support of NEF computation cluster.

Parameter identification in compartmental systems. In collaboration with F. Dayan (Exactcure), we work on practical problems of identifiability of parameters in linear pharmacokinetic models. This was the subject of the internship of Laurent Dragoni.

7.1.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Olivier Bernard, Valentina Baldazzi, Stefano Casagranda, Francis Mairet, Ivan Egorov, Sofia Almeida, Claudia Lopez Zazueta, Lucie Chambon, Luis Gomes Pereira, Eleni Firippi, Ignacio Lopez Munoz.

7.1.2.1. Hybrid models analysis

Applying differential dynamic logic to biological networks. In [26] we have explored the framework of differential dynamic logic for the analysis of hybrid systems and, in particular, piecewise linear models of biological networks (collaboration with D. Figueiredo and M.A. Martins from the University of Aveiro, Portugal).

Attractor computation using interconnected Boolean networks. Following the work in [10] and [58], we have generalized the method for computation of the asymptotic graph. In addition, we have extended this methodology for the case of Boolean networks with synchronous updates (collaboration with D. Figueiredo and M.A. Martins from the University of Aveiro, Portugal).

Periodic orbits in non monotonic negative feedback circuits. We study the occurrence of periodic solutions in an n-dimensional class of negative feedback systems defined by smooth vector fields with a window of not necessarily monotonic activity. By circumscribing the smooth system by two piecewise linear ones, we show there exists an invariant toroidal region which contains a periodic orbit of the original smooth system [9]. We prove that this orbit is unique under some conditions on the parameters.

7.1.2.2. Continuous models analysis

Reduced models for the mammalian cell cycle and clock. In the context of project ANR ICycle, we have focused on identifying and analysing the main mechanisms underlying the cell cycle and the circadian clock in mammalian cells. A reduced two-dimensional model of the cell cycle is described [38]; the model faithfully predicts the period of the cell cycle in response to an external growth factor input (experimental data on the periods is from F. Delaunay's lab). This work is in collaboration with F. Delaunay and part of the PhD thesis of Sofia Almeida.

Interconnection of reduced models of the mammalian cell cycle and clock. Also in the context of project ANR ICycle, we have studied several possibilities for the interconnection between these two mammalian oscillators, using the reduced model already described in [38] and two different possible oscillatory circuits of low dimension. This work is part of the Master's thesis of Eleni Firippi.

Modeling the apoptotic signaling pathway. The goal is to study the origins of cell-to-cell variability in response to anticancer drugs and provide a link between complex cell signatures and cell response phenotype [45]. To do this, we have been analysing models of the apoptosis pathway to compare the effects of different sources of variability at the transcriptional, translational and receptor levels (collaboration with J. Roux, for the PhD thesis of Luis Pereira).

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 [17]. We also study other models of the global cellular machinery, and growth models ([22]). This is part of the PhD thesis of Stefano Casagranda, and done in collaboration with Inria IBIS project-team, in particular with D. Ropers.

Analysis and reduction of a model of sugar metabolism in peach fruit. Predicting genotype-to-phenotype relationships under contrasting environments is a big challenge for plant biology and breeding. A model of sugar metabolism in peach fruit has been recently developed and applied to 10 peach varieties [25]. The aim of this ongoing work is to reduce model's size and complexity to allow for calibration on a whole progeny of 106 genotypes and for further application to virtual breeding (collaboration with B. Quilot-Turion and Mohamed Memmah (INRA Avignon) and part of the PhD thesis of Hussein Kanso).

Analysis of an integrated cell division-endoreduplication and expansion model. The development of a new organ depends on cell-cyle progression and cell expansion, but the interaction and coordination between these processes is still unclear. An integrated model of fruit development has been developed and used to investigate the regulation of cell expansion capabilities. To this aim, different control schemes are tested by means of specific model variants and simulation results compared to observed data in tomato [14].

7.1.2.3. 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, in collaboration with Inria

IBIS project-team. We showed that a good suboptimal control solution could be implemented in the cell by ppGpp (a small molecule involved in the regulation of ribosomes) [5]. We developed different versions of the problem [43], [36], and consider a new problem where the aim is to optimize the production of a product (ANR projects Reset and Maximic).

Control of 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. A nonlinear control is built [32].

7.1.2.4. Large scale metabolic modeling

Metabolic modeling generally assumes balanced growth, *i.e.* that there is no accumulation of intermediate compound, and that the metabolism is rapidly at quasi steady state. We have proposed a new approach called DRUM where this hypothesis is relaxed by splitting the metabolic network into subnetworks and assuming that some compounds can accumulate between the subnetworks [2], [49]. This approach was successfully applied to several cases where the strong variations in light or nutrient resources induce a strong accumulation in the microalgal cells which could not be represented by the state of the art approaches [48]. More recently we have expended this approach to the modeling of diauxic growth for heterotrophic or mixotrophic microalgae [15].

7.1.2.5. Slow-Fast analysis of metabolic models

Metabolic modeling generally assumes balanced growth, *i.e.* that there is no accumulation of intermediate compound, and that the metabolism is rapidly at quasi steady state. We go beyond this hypothesis by considering that some metabolic reactions are slow, while other are fast. Then we analyse the differential system using Tikhonov's Theorem. We compare the results obtained using the Drum approach [2], and show that Drum is a reasonable approximation, provided that growth rate stays low. This is part of the PhD thesis of Claudia Lopez Zazueta.

7.2. Fields of application

7.2.1. Bioenergy

7.2.1.1. Modeling microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Walid Djema, Ignacio Lopez Munoz, David Demory, Ouassim Bara, Jean-Philippe Steyer.

Experimental developments

Running experiments in controlled dynamical environments. The experimental platform made of continuous photobioreactors driven by a set of automaton controlled by the ODIN software is a powerful and unique tool which gave rise to a quantity of very original experiments. Such platform improved knowledge of several biological processes such as lipid accumulation or cell cycle under light fluctuation, etc. [55],[19].

This experimental platform was used to control the long term stress applied to a population of microalgae. This Darwinian selection procedure generated two new strains after more than 6 months in the so called selectiostats. A strain with $+92\%$ lipids was obtained, another more transparent resulting in $+92\%$ enhancement in productivity [18].

Other experiments were carried out to reproduce the light signal percept by a cell in a raceway pond [60], derived from hydrodynamical studies [64]. 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 [59]. Experiments with coloured light demonstrated that the growth rate results from the absorbed light, whatever its wavelength.

On top of this, we carried out outdoor pilot experiments with solar light. We tested the impact of various temperatures, resulting from different shadowing configurations on microalgal growth rate [56],[40]. This is the topic of Bruno Assis Pessi's master thesis.

These works have been carried out in collaboration with A. Talec and E. Pruvost (CNRS/UPMC - Oceanographic Laboratory of Villefranche-sur-Mer LOV).

Metabolism of carbon storage and lipid production. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*, on the basis of the DRUM framework , in order to simulate autotophic, heterotropic and mixotrophic growth, and to determine how to reduce substrate inhibition [15]. The model was extended for other substrates such as glucose or glycerol. A simplified model was developped by I. Lopez to represent the dynamics of polar lipids, especially when faced to a high oxygen concentration.

Modeling the coupling between hydrodynamics and biology. In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on a new multilayer discretisation of Navier-Stokes equations) with microalgae growth was developed [51]. This model is supported by the work of ANGE aiming at improving the discretization scheme to more finely represent the hydrodynamics of the raceway and more accurately reconstruct Lagrangian trajectories. The statistical analysis of both theoretical properties of probability densities for perfectly mixed systems and output of Lagrangian simulations demonstrate the accurate reconstruction of the trajectories. As a consequence, more relevant experimental protocols have been proposed to more realistically design simplified light signal for experiments.

Modeling photosynthetic biofilms. Several models have been developed to represent the growth of microalgae within a biofilm. A first structured physiological model uses mixture theory to represent the microalgae growth, based on the consideration of intracellular reserves triggering the processes of growth, respiration and excretion. We consider separately the intracellular storage carbon (lipids and carbohydrates) and the functional part of microalgae [29]. Another approach accounts for the dynamics of the light harvesting systems when cells are submitted to rapid successions of light and dark phases. A simpler model was developed and used to identify the optimal working mode of a process based on photosynthetic biofilm growing on a conveyor belt.

Modeling microalgae production processes. The integration of different models developed within BIOCORE [52] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems.

Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [72]. A simplified version of this model, describing microalgal growth under varying light and temperature conditions predicts microalgal productivity in the perspective of large scale biofuel production [23].

In the framework of the ANR project Purple Sun, we developed a thermal model of a raceway pond within a greenhouse in order to estimate the culture temperature. We also included in the microalgae model the effect of light wavelength. This model has been calibrated on experimental data from LOV and has been used to support lighting strategy in order to optimize microalgal productivity (a patent on this process has been submitted). We have shown in [40] that a control strategy based on shadowing with solar panel can significantly improve productivity, especially during the early growth stage of the culture.

A procedure for rapid outdoor model calibration, from lab data, has been proposed and applied to the microalgae *Dunaliella salina* [56].

Modeling thermal adaptation in microalgae. We have studied several models of microalgae growth to different temperatures [27]. In particular, we have detailed the impact of higher temperatures on cell mortality [20]. Experiments have been carried out in collaboration with A.-C. Baudoux (Biological Station of Roscoff) in order to study growth of various species of the microalgae genus *Micromonas* at different temperatures. After calibration of our models, we have shown that the pattern of temperature response is strongly related to the site where cells were isolated. We derived a relationship to extrapolate the growth response from isolation location. With this approach, we proved that the oceanwide diversity of *Micromonas* species is very similar to the oceanwide diversity of the phytoplankton. We have used Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. We could then predict the evolution of this biodiversity in a warming ocean and show that phytoplankton must be able to adapt within 1000 generation to avoid a drastic reduction in biodiversity.

Modeling viral infection in microalgae. Experiments have been carried out in collaboration with A.-C. Baudoux (Biological Station of Roscoff) in order to study the impact of viral infections on the development

of populations of *Micromonas* at different temperatures. This work revealed a qualitative change in viral infection when temperature increases. A model was developed to account for the infection of a *Micromonas* population, with population of susceptible, infected and also free viruses. The model turned out to accurately reproduce the infection experiments at various temperatures, and the reduction of virus production above a certain temperature [24].

7.2.1.2. Control and Optimization of microalgae production

Optimization of the bioenergy production systems. A model predictive control approach was run based on simple microalgae models coupled with thermal physical models. Optimal operation in continuous mode for outdoor cultivation was determined when allowing variable culture depth. Assuming known weather forecasts considerably improved the control efficiency [23].

Interactions between species. We had formerly proposed an adaptive controller which regulates the light at the bottom of the reactor [70]. 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. We have also proposed a strategy based on light stresses in order to penalize the strains with a high pigment content and finally select microalgae with a low Chlorophyll content [69]. This characteristic is of particular interest for maximizing biomass production in dense culture. The strategy has been carried out at the LOV and eventually the productivity of *Tisochrysis lutea* was improved by 75%. A patent on this strategy has been submitted.

Strategies to improve the temperature response have also been studied. We modelled the adaptive dynamics for a population submitted to a variable temperature [62]. This was used at the LOV to design experiments with periodic temperature stresses during 200 days aiming at enlarging the thermal niche of *Tisochrysis lutea*. It resulted in an increase by 2 degrees of the thermal niche [18].

Finaly, optimal strategies when selecting the strain of interest within a set of n species competing for the same substrate has been proposed [16].

7.2.2. Biological depollution

7.2.2.1. Control and optimization of bioprocesses for depollution

Participants: Olivier Bernard, Carlos Martinez Von Dossow, Jean-Luc Gouzé.

Although bioprocesses involve an important biodiversity, the design of bioprocess control laws are generally based on single-species models. In [68], 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 monospecific equilibrium point, coexistence, or even a failure point. Finally, we have shown that, for this case, the robustness can be improved by a saturation of the control.

7.2.2.2. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Carlos Martinez Von Dossow.

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 [74].

We have proposed several models to account for the biodiversity in the microalgal pond and for the interaction between the various species. These models were validated with data from the Saur company. More specifically, we have included in the miroalgae model the impact of the strong turbidity, and derived a theory to better understand the photolimitation dynamics especially when accounting for the photo-inhibition in the illuminated periphery of the reactor. Optimal control strategies playing with the dilution rate, shadowing or modifying depth were then studied [40].

7.2.2.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer, Marjorie Alejandra Morales Arancibia.

Environmental impact assessment. In the sequel of the pioneering life cycle assessment (LCA) work of [65], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable.
We studied a new paradigm to improve the energy balance by combining biofuel production with photovoltaic electricity. This motivated the design of the purple sun ANR-project where electricity is produced by semi transparent photovoltaic panels [50] under which photosynthetic microalgae are growing. The LCA of a greenhouse with, at the same time, photovoltaic panels and low emissivity glasses is studied. Depending on the period of the year, changing the species can both improve productivity and reduce environmental footprint.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE (Laboratory of Environmental Biotechnology, Narbonne) and Pierre Collet (IFPEN).

7.2.3. Design of ecologically friendly plant production systems

7.2.3.1. Controlling plant arthropod 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 optimization of natural enemies introductions has been investigated for several years [66], [73], unveiling the crucial influence of within-predator density dependent processes. Since some natural enemies may be more prone to exhibit positive density dependent dynamics rather than negative ones, we studied the impact of positive predator-predator interactions on the optimal biological control introduction strategies [13]. Extension of this result have been performed to take into account stochasticity by develeoping a master equation for the combined continuous-stochastic process and a purely stocastic model. This last part of N. Bajeux's PhD thesis mycitePhD:bajeux was performed in collaboration with Vincent Calcagno (ISA).

Characteristics of space and the behavior and population dynamics of parasitoids. We studied the influence of space on the spread of biological control agents through computer simulations and laboratory experiments on *Trichogramma*. This is the topic of Marjorie Haond's PhD thesis (ISA, 2015-). In particular, we showed both theoretically and experimentally how habitat richness [63] shape the spatio-temporal dynamics of populations in spatially structured environments. This work is being performed in collaboration with Elodie Vercken (ISA) and Lionel Roques (BioSP, Avignon).

Model of coffee berry borer dynamics. We built a first model describing the coffee berry borer dynamics, in order to design efficient and sustainable control strategies, including alternative methods to pesticides (cropping practices, trapping, biological control). This single-season model is based on the insect life-cycle and includes the berry availability during a cropping season. Local and global stability results, the latter using Lyapunov functions, were obtained for both the pest-free and the endemic equilibria. Furthermore, this model was extended to integrate the berry maturation age. The well-posedness of the resulting PDE model was shown. This research pertains to Yves Fotso Fotso's PhD thesis, who visited BIOCORE during 4 months in 2017 in the framework of the EPITAG associate team.

7.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Julien Guégan, Yves Fotso-Fotso, Israel Tankam-Chedjou.

Sustainable management of plant resistance. We studied other plant protection methods 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.

Experiments were conducted in INRA Avignon, followed by high-throughput sequencing (HTS) to identify the dynamics of virus strains competing 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 [31]. Also, the Quantitative Trait Loci (QTL) controlling viruses effective population sizes (linked to genetic drift) have been identified for two different viruses, showing the genetic origin of these parameters and the presence of general and virus specific QTLs [34]. This was done in collaboration with Frédéric Fabre (INRA Bordeaux) and Benoît Moury (INRA Avignon).

We also developed an epidemiological model describing the dynamics of root-knot nematodes in a protected vegetable cropping system, to design optimal management strategies of crop resistance. The model was fitted to experimental and field data. Preliminary results show that alternating susceptible and resistant crops not only increased the resistance durability, but reduced the disease intensity over time [28], [41]. This research pertains to Samuel Nilusmas' PhD thesis.

We developed and partly calibrated a (spatio-)temporal epidemiological model of the phoma stem canker of oilseed rape, to design sustainable resistance deployment strategies. Ongoing work includes the completion of this study and the development of a user-friendly simulation tool. It will be achieved through the MoGeR project, in collaboration with BIOGER (INRA Grignon) and partners from technical institutes and cooperatives. It benefits from the resources and support of NEF computation cluster.

Model of nematodes-plantain roots dynamics. We developed and analysed a seasonal model describing the interactions between nematodes and plantain roots, to design efficient and sustainable control strategies, including alternative methods to pesticides (cropping practices, resistant or tolerant banana cultivars, biological control). It is a doubly hybrid system, so as to take into account the plantain root growth. A slow-fast dynamics approximation was used to obtain local stability results for the pest-free equilibrium and exact solutions around this equilibrium. Conditions were derived for nematode extinction, depending in particular on the delay between cropping seasons. This research pertains to Israël Tankam Chedjou's PhD thesis, who visited BIOCORE during 4 months in 2017 in the framework of the EPITAG associate team.

Mate limitation and cyclic epidemics. We studied the effect of mate limitation in parasites which perform both sexual and asexual reproduction in the same host. Since mate limitation implies positive density dependence at low population density, we modeled the dynamics of such species with both density-dependent (sexual) and density-independent (asexual) transmission rates. A first simple SIR model incorporating these two types of transmission from the infected compartment, suggested that combining sexual and asexual spore production can generate persistently cyclic epidemics [30].

7.2.3.3. Optimality/games in population dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Ivan Egorov, Pierre-Olivier Lamare.

Optimal resource allocation. Mycelium growth and sporulation are considered for phytopathogenic fungi. For biotrophic fungi, a flow of resource is uptaken by the fungus without killing its host; in that case, life history traits (latency-sporulation strategy) have been computed based on a simple model considering a single spore initiating the mycelium, several spores in competition and applying optimal resource allocation [42], and several spores in competition through a dynamic game. The solution of this dynamic game has been shown to be the equilibrium of two-trait adaptive dynamics in Julien Guégan's internship. Also, the obtained sporulation strategy has been put in a PDE model to evaluate how the characteristics of the fungus evolve along a colonization gradient. This work, in the framework of the ANR Funfit project, is done with Fabien Halkett of INRA Nancy.

Dynamic games as a model of animal foraging. P. Bernhard has continued his investigations of dynamic games with randomly arriving players as a model of animal foraging and of competition in open markets. He has written the chapter "Robust Control and Dynamic Games"in the Handbook of Dynamic Games Theory [54].

7.3. Patents

Two patents were proposed for improving growth of microalgae with green light [53] and for enhancing the thermal niche after long term thermal stress in a continuous reactor [47].

CARMEN Project-Team

7. New Results

7.1. A parameter optimization method to solve the ECG inverse problem

Existing electrocardiographic inverse models express their results either in terms of potentials on the heart surface or in terms of activation times in the heart. G. Ravon developed a new method which gives a potentially more useful answer in terms of three parameters of the underlying action potentials in the heart [27]. Since there are more parameters, care had to be taken to avoid overfitting. Tests on in-silico and ex-vivo data showed good results: the method gave better activation maps than the method of fundamental solutions to which it was compared, and fitted the repolarization phase of the ECG accurately. Figure shows an example of an inversely estimated repolarization map.

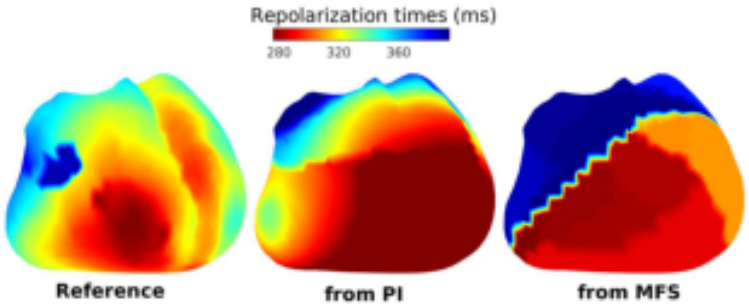


Figure 3. Reference repolarization map and inversely estimated maps, using the newly developed method (middle) and the method of fundamental solutions to which it was compared (right). The new method results in more realistic patterns.

7.2. Optimal control to Bidomain-Bath model

This project is concerned with the study of the convergence analysis for an optimal control of a bidomain-bath model. The bidomain-bath model equations describe the cardiac bioelectric activity at the tissue bath volumes where the control acts at the boundary of the tissue domain. In recent work [13] [44], we established the well-posedeness of the direct bidomain-bath model by a discrete Galerkin approach. The convergence proof is based on deriving a series of a priori estimates and using a general $L2$-compactness criterion. Moreover, the well-posedeness of the adjoint problem and the first order necessary optimality conditions are shown. Comparing to the direct problem, the convergence proof of the adjoint problem is based on using a general L^1-compactness criterion. The numerical tests are demonstrated which achieve the successful cardiac defibrillation by utilizing less total current. Finally, the robustness of the Newton optimization algorithm is presented for different finer mesh geometries.

7.3. Bidomain Calcium Dynamics in Cardiac Cell

In our project [35], we are interested in modeling the interaction of Calcium dynamics in a bidomain medium including Sarcolemma and Sarcoplasmic reticulum. The governing equations consist of a nonlinear reaction-diffusion system representing the various calcium fuxes and theirs buffers in the two medias. A priori stability

bounds and the solvability of the system is analyzed using a fixed-point approach. We introduce a finite element method to numerically solve our model equations. Moreover, we establish existence of discrete solutions and show convergence to a weak solution of the original problem. Finally, we report several 2D and 3D numerical experiments illustrating the behavior of the proposed scheme.

7.4. Cardiac electromechanics with physiological ionic model

This project [37] is concerned with the mathematical analysis of a coupled elliptic-parabolic system modeling the interaction between the propagation of electric potential coupled with general physiological ionic models and subsequent deformation of the cardiac tissue. A prototype system belonging to this class is provided by the electromechanical bidomain model, which is frequently used to study and simulate electrophysiological waves in cardiac tissue. The coupling between muscle contraction, biochemical reactions and electric activity is introduced with a so-called active strain decomposition framework, where the material gradient of deformation is split into an active (electrophysiology-dependent) part and an elastic (passive) one. We prove existence of weak solutions to the underlying coupled electromechanical bidomain model under the assumption of linearized elastic behavior of the updated nonlinear diffusivities. The proof of the existence result is proved by means of a non-degenerate approximation system, the Faedo-Galerkin method, and the compactness method.

7.5. Electrocardiographic lead fields

Currently a monodomain reaction-diffusion model is a well-established method to simulate the electrical activity of the heart [58], [59], even more so because it can be adapted to approximate a bidomain model very closely [46], [49]. Computing the electrocardiogram (ECG) from the results of such models is harder because it requires large linear systems to be solved, and does not scale well to large numbers of processors. A possible solution is to use so-called lead fields, the electrocardiographic term for a linear combination of Green's functions that express the ECG potential as an integral over a field of electric current dipoles. M. Potse has implemented and tested methods to compute and use lead fields for ECG simulation with the Propag code. It turned out that this classical method is practical and sufficiently accurate, and gives a huge scaling advantage on modern highly parallel computers. This result is of practical importance for our applied work, and a journal manuscript on this topic will be submitted in January 2018.

7.6. Rapid localization of arrhythmia

Our pilot studies using model data [25] have shown promising results for a proposed simple and rapid localization method to be used in the catheterization laboratory. We have found that even with only a few ECG electrodes accuracies in the order of millimeters can be achieved for the position of an arrhythmia origin with respect to a catheter position. A journal manuscript on this topic is expected to be complete in February 2018.

7.7. Bilayer model

The rigorous proof that mathematically found the bilayer model from S. Labarthe PhD thesis [56] was actually published in SIAM Journal of Applied Math [15]. Based on sophisticated energy estimates, it proves that the bilayer model is the rigorous limit of the underlying three-dimensional model.

7.8. multi-electrode array measurement

In the context of the CardioXComp project, in collaboration with the REO team and the company Notocord, we proposed a strategy to analyze the signals acquired by multi-electrode array (MEA) on cultures of hiPSC-CMs cells with drug compounds, and to automatically deduce the channels affected by the drug. First, in [10], we study how MEA meausrement can be modeled in such a way that the produced in-silico signals are comparable to real ones. The main problems concern the heterogenities of cell cultures. Then, in [20] a method based on parameter identification, by comparing the in-silico and real signals, was used on signals acquired with commercial systems on cell culture with various drugs. The IC-50 and dose-response of several drugs could be assessed. This kind of techniques could contribute to promote the technology based on MEA and hiPSC-CMs.

7.9. High-order integration methods for ion channel models

On November 15, 2017, C. Douanla Lontsi defended his PhD thesis [9] on the numerical analysis of time-stepping methods for the cardiac monodomain equations. A huge amount of work was carried out in the thesis. The thesis builds on the seminal Rush-Larsen technique [62], [57], and the recent novel computational interest in exponential integrator methods. Two new exponential methods of arbitrarily high order are proposed (EABk and RLk). Most notably, Rush-Larsen techniques of order $k = 2, 3, 4$ were entirely explicited. The theory was adapted to analyse these methods and convergence proofs were derived. The complete Dahlquist stability region of these methods was documented. Finally, the methods were integrated into an IMEX strategy to solve the monodomain equation in 1D to 3D problems, with two ionic models (BR and TNNP). The results essentially show that order at least 3 is required to leas to reasonably accurate simulations. Three journal papers were submitted in 2017, [39], [41], [40].

7.10. High-order finite-volume discretizations

Y. Coudière and R. Turpault proposed new simple and efficient high-order finite volume discretizations to be applied to the monodomain equation. They showed how the method can be easily implemented up to the order 6, with very good results, also for simulation of complex propagation patterns. The results were published in [16].

7.11. Identification of multiple space dependent ionic parameters in cardiac electrophysiology modelling

In this paper, we consider the inverse problem of space dependent multiple ionic parameters identification in cardiac electrophysiology modelling from a set of observations. We use the monodomain system known as a state-of-the-art model in cardiac electrophysiology and we consider a general Hodgkin-Huxley formalism to describe the ionic exchanges at the microscopic level. This formalism covers many physiological trans-membrane potential models including those in cardiac electrophysiology. Our main result is the proof of the uniqueness and a Lipschitz stability estimate of ion channels conductance parameters based on some observations on an arbitrary subdomain.

DRACULA Project-Team

5. New Results

5.1. Cancer

5.1.1. Long-term treatment effects in chronic myeloid leukemia

We propose and analyze in [12] a simplified version of a partial differential equation (PDE) model for chronic myeloid leukemia (CML) derived from an agent-based model proposed by Roeder *et al.* in [38]. This model describes the proliferation and differentiation of leukemic stem cells in the bone marrow and the effect of the drug Imatinib on these cells. We first simplify the PDE model by noting that most of the dynamics occurs in a subspace of the original 2D state space. Then we determine the dominant eigenvalue of the corresponding linearized system that controls the long-term behavior of solutions. We mathematically show a non-monotonous dependence of the dominant eigenvalue with respect to treatment dose, with the existence of a unique minimal negative eigen-value. In terms of CML treatment, this shows that there is a unique dose that maximizes the decay rate of the CML tumor load over long time scales. Moreover this unique dose is lower than the dose that maximizes the initial tumor load decay. Numerical simulations of the full model confirm that this phenomenon is not an artifact of the simplification. Therefore, while optimal asymptotic dosage might not be the best one at short time scales, our results raise interesting perspectives in terms of strategies for achieving and improving long-term deep response.

5.1.2. A model of interaction between the immune system and cancer cells in chronic myelogenous leukemia

We describe in [11] a simple model for the interaction between leukemic cells and the autologous immune response in chronic phase chronic myelogenous leukemia (CML). This model is a simplified version of the model we proposed in 2015 ([34]). Our simplification is based on the observation that certain key characteristics of the dynamics of CML can be captured with a three compartments model: two for the leukemic cells (stem cells and mature cells) and one for the immune response. We characterize the existence of steady states and their stability for generic forms of immunosuppres-sive effects of leukemic cells. We provide a complete co-dimension one bifurcation analysis. Our results show how clinical response to tyrosine kinase inhibitors treatment is compatible with the existence of a stable low-disease, treatment-free steady state.

5.1.3. A hybrid computation model to describe the progression of multiple myeloma and its intra-clonal heterogeneity

Multiple myeloma (MM) is a genetically complex hematological cancer that is characterized by proliferation of malignant plasma cells in the bone marrow. MM evolves from the clonal premalignant disorder monoclonal gammopathy of unknown significance (MGUS) by sequential genetic changes involving many different genes, resulting in dysregulated growth of multiple clones of plasma cells. The migration, survival, and proliferation of these clones require the direct and indirect interactions with the non-hematopoietic cells of the bone marrow. We develop in [14], a hybrid discrete-continuous model of MM development from the MGUS stage. The discrete aspect of the model is observed at the cellular level: cells are represented as individual objects which move, interact, divide, and die by apoptosis. Each of these actions is regulated by intracellular and extracellular processes as described by continuous models. The hybrid model consists of the following submodels that have been simplified from the much more complex state of evolving MM: cell motion due to chemotaxis, intracellular regulation of plasma cells, extracellular regulation in the bone marrow, and acquisition of mutations upon cell division. By extending a previous, simpler model in which the extracellular matrix was considered to be uniformly distributed, the new hybrid model provides a more accurate description in which cytokines are produced by the marrow microenvironment and consumed by the

myeloma cells. The complex multiple genetic changes in MM cells and the numerous cell-cell and cytokine-mediated interactions between myeloma cells and their marrow microenviroment are simplified in the model such that four related but evolving MM clones can be studied as they compete for dominance in the setting of intraclonal heterogeneity.

5.1.4. Fast and binary assay for predicting radiosensitivity based on the nucleoshuttling of ATM protein: development, validation and performances

The societal and clinical impact of post-radiotherapy adverse tissue events (AE) has highlighted the need of molecular parameters to predict individual radiosensitivity. Recent studies have stressed the role of the phosphorylated forms of the ATM protein (pATM) and its nucleoshuttling in response to radiation. The statistical performance of the pATM immunofluorescence assay to predict AE is promising. However, immunofluorescence requires a time-consuming amplification of cells. The purpose of the study in [24] was to develop a predictive assay based on the ELISA technique that renders faster the previous approach. Materials and methods This study was performed on 30 skin fibroblasts from 9 radioresistant and 21 AE patients. Patients were divided in 2 groups, radioresistant (toxicity grade<2) and radiosensitive (toxicity grade ≥ 2). The quantity of nuclear pATM molecules was assessed by ELISA method at 10 min and 1 h after 2 Gy and compared to pATM immunofluorescence data. Results The pATM ELISA data were found in quantitative agreement with the immunofluorescence ones. A ROC analysis was applied first to two data sets (a training ($n = 14$) and a validating ($n = 16$) one) and thereafter to the whole data with a 2-fold cross-validation method. The assay showed an AUC value higher than 0.8, a sensitivity of 0.8 and a specificity ranging from 0.75 and 1, which strongly document the predictive power of the pATM ELISA assay. Conclusion This study showed that the assessment of nuclear pATM quantity after 2 Gy via ELISA technique can be the basis of a predictive assay with the highest statistical performance among the available predictive approaches.

5.2. Immune response

5.2.1. Identification of nascent memory CD8 T cells and modeling of their ontogeny

Primary immune responses generate short-term effectors and long-term protective memory cells. The delineation of the genealogy linking naive, effector and memory cells has been complicated by the lack of phenotypes discriminating effector from memory differentiation stages. Using transcriptomics and phenotypic analyses, we identify (see [17]) a novel marker combination that allows us to track nascent memory cells within the effector phase. We then use a formal approach based on mathematical models describing the dynamics of population-size evolutions to test potential progeny links and demonstrate that most cells follow a linear naive-early effector-late effector-memory pathway. Moreover, our mathematical model allows long-term prediction of memory cell numbers from a few early experimental measurements. Our work thus provides a phenotypic means to identify effector and memory cells, as well as a mathematical framework to investigate the ontology of their generation and to predict the outcome of immunization regimens in terms of memory cell numbers generated.

5.2.2. Modelling the dynamics of virus infection and immune response in space and time

Spreading of viral infection in the tissues such as lymph nodes or spleen depends on virus multiplication in the host cells, their transport and on the immune response. Reaction–diffusion systems of equations with delays in proliferation and death terms of the immune cells represent an appropriate model to study this process. The properties of the immune response and the initial viral load determine the regimes of infection spreading. In the proposed model [13], the proliferation rate of the immune cells is represented by a bell-shaped function of the virus concentration which increases for small concentrations and decreases if the concentration is sufficiently high. We use such a model system to show that an infection can be completely eliminated or it can remain present together with a decreased concentration of immune cells. Finally, immune cells can be completely exhausted leading to a high virus concentration in the tissue. In addition, we predicted two novel regimes of infection dynamics not observed before. Infection propagation in the tissue can occur as a superposition of two travelling waves: first wave propagates as a low level infection front followed by a high level infection

front with a smaller speed of propagation. Both of the travelling waves can have a positive or a negative speed corresponding to infection advancement or retreat. These regimes can be accompanied by instabilities and the emergence of complex spatiotemporal patterns.

5.2.3. Estimates and impact of lymphocyte division parameters from CFSE data using mathematical modeling

Carboxyfluorescein diacetate succinimidyl ester (CFSE) labelling has been widely used to track and study cell proliferation. In [23], we use mathematical modeling to describe the kinetics of immune cell proliferation after an in vitro polyclonal stimulation tracked with CFSE. This approach allows us to estimate a set of key parameters, including ones related to cell death and proliferation. We develop a three-phase model that distinguishes a latency phase, accounting for non-divided cell behaviour, a resting phase and the active phase of the division process. Parameter estimates are derived from model results, and numerical simulations are then compared to the dynamics of in vitro experiments, with different biological assumptions tested. Our model allows us to compare the dynamics of CD4+ and CD8+ cells, and to highlight their kinetic differences. Finally we perform a sensitivity analysis to quantify the impact of each parameter on proliferation kinetics. Interestingly, we find that parameter sensitivity varies with time and with cell generation. Our approach can help biologists to understand cell proliferation mechanisms and to identify potential pathological division processes.

5.3. Erythropoiesis

5.3.1. Investigating the role of the experimental protocol in phenylhydrazine-induced anemia on mice recovery

Production of red blood cells involves growth-factor mediated regulation of erythroid progenitor apoptosis and self-renewal. During severe anemia, characterized by a strong fall of the hematocrit followed by a recovery phase, these controls allow a fast recovery of the hematocrit and survival of the organism. Using a mathematical model of stress erythropoiesis and an ad hoc numerical method, we investigate in [9] the respective roles of anemia-inducing phenylhydrazine injections and physiological regulation on the organism's recovery. By explicitly modeling the experimental protocol, we show that it mostly characterizes the fall of the hematocrit following the anemia and its severeness, while physiological process regulation mainly controls the recovery. We confront our model and our conclusions to similar experiments inducing anemia and show the model's ability to reproduce several protocols of phenylhydrazine-induced anemia. In particular, we establish a link between phenylhydrazine effect and the severeness of the anemia.

5.3.2. Numerical integration of an erythropoiesis model with explicit growth factor dynamics

Erythropoiesis, the red blood cell production process, involves interactions between cell populations with different differentiation states, mainly immature progenitor cells and mature erythrocytes, and growth factors such as erythropoietin and glucocorticoids, known to respectively inhibit cell apoptosis, stimulate proliferation and differentiation, and stimulate self-renewal. The feedback regulation of this process allows a very fast and efficient recovery in the case of a severe anemia. We consider in [8] an age-structured model of red blood cell production accounting for these feedback regulations and the dynamics of growth factors. We theoretically show the existence of a unique positive steady state for the model and we propose a numerical method to obtain an approximation to its solution. Experiments are reported to show numerically, on one hand, the optimal convergence order of the numerical scheme and, on the other hand, a fine approximation to real experimental data, with a suitable selection of the parameters involved.

5.4. Methodological developments

5.4.1. Traveling waves for a model of hematopoiesis

The formation and development of blood cells (hematopoiesis) is a very complex process. This process involves a small population of cells called hematopoietic stem cells (HSCs). The HSCs are undifferentiated

cells, located in the bone marrow before they become mature blood cells and enter the blood stream. They have a unique ability to produce either similar cells (self-renewal), or cells engaged in one of different lineages of blood cells: red blood cells, white cells and platelets (differentiation). The HSCs can be either in a proliferating or in a quiescent phase. In [6], we distinguish between dividing cells that enter directly to the quiescent phase and dividing cells that return to the proliferating phase to divide again. We propose a mathematical model describing the dynamics of HSC population, taking into account their spatial distribution. The resulting model is an age-structured reaction-diffusion system. The method of characteristics reduces this model to a coupled reaction-diffusion equation and difference equation with delay. We study the existence of traveling wave fronts connecting the zero steady state with the unique positive uniform one. We use a monotone iteration technique coupled with the upper and lower solutions method.

5.4.2. A hybrid finite volume method for advection equations and its applications in population dynamics

We present in [30] a very adapted finite volume numerical scheme for transport type-equation. The scheme is an hybrid one combining an anti-dissipative method with down-winding approach for the flux ([35]; [36]) and an high accurate method as the WENO5 one ([37]). The main goal is to construct a scheme able to capture in exact way the numerical solution of transport type-equation without artifact like numerical diffusion or without "stairs" like oscillations and this for any regular or discontinuous initial distribution. This kind of numerical hybrid scheme is very suitable when properties on the long term asymptotic behavior of the solution are of central importance in the modeling what is often the case in context of population dynamics where the final distribution of the considered population and its mass preservation relation are required for prediction.

5.4.3. Inferring gene regulatory networks from single-cell data: a mechanistic approach

The recent development of single-cell transcriptomics has enabled gene expression to be measured in individual cells instead of being population-averaged. Despite this considerable precision improvement, inferring regulatory networks remains challenging because stochasticity now proves to play a fundamental role in gene expression. In particular, mRNA synthesis is now acknowledged to occur in a highly bursty manner. We propose in [21] to view the inference problem as a fitting procedure for a mechanistic gene network model that is inherently stochastic and takes not only protein, but also mRNA levels into account. We first explain how to build and simulate this network model based upon the coupling of genes that are described as piecewise-deterministic Markov processes. Our model is modular and can be used to implement various biochemical hypotheses including causal interactions between genes. However, a naive fitting procedure would be intractable. By performing a relevant approximation of the stationary distribution, we derive a tractable procedure that corresponds to a statistical hidden Markov model with interpretable parameters. This approximation turns out to be extremely close to the theoretical distribution in the case of a simple toggle-switch, and we show that it can indeed fit real single-cell data. As a first step toward inference, our approach was applied to a number of simple two-gene networks simulated in silico from the mechanistic model and satisfactorily recovered the original networks. Our results demonstrate that functional interactions between genes can be inferred from the distribution of a mechanistic, dynamical stochastic model that is able to describe gene expression in individual cells. This approach seems promising in relation to the current explosion of single-cell expression data.

5.5. Physiology

5.5.1. A multiscale modeling approach for the regulation of the cell cycle by the circadian clock

We present in [18] a multiscale mathematical model for the regulation of the cell cycle by the circadian clock. Biologically, the model describes the proliferation of a population of heterogeneous cells connected to each other. The model consists of a high dimensional transport equation structured by molecular contents of the cell cycle-circadian clock coupled oscillator. We propose a computational method for resolution adapted from the concept of particle methods. We study the impact of molecular dynamics on cell proliferation and show an example where discordance of division rhythms between population and single cell levels is observed.

This highlights the importance of multiscale modeling where such results cannot be inferred from considering solely one biological level.

5.5.2. The lifespan and turnover of microglia in the human brain

The hematopoietic system seeds the CNS with microglial progenitor cells during the fetal period, but the subsequent cell generation dynamics and maintenance of this population have been poorly understood. We report in [25] that microglia, unlike most other hematopoietic lineages, renew slowly at a median rate of 28% per year, and some microglia last for more than two decades. Furthermore, we find no evidence for the existence of a substantial population of quiescent long-lived cells, meaning that the microglia population in the human brain is sustained by continuous slow turnover throughout adult life.

5.5.3. Impact of fat mass and distribution on lipid turnover in human adipose tissue

Differences in white adipose tissue (WAT) lipid turnover between the visceral (vWAT) and subcutaneous (sWAT) depots may cause metabolic complications in obesity. In [26], we compare triglyceride age and, thereby, triglyceride turnover in vWAT and sWAT biopsies from 346 individuals and find that subcutaneous triglyceride age and storage capacity are increased in overweight or obese individuals. Visceral triglyceride age is only increased in excessively obese individuals and associated with a lower lipid removal capacity. Thus, although triglyceride storage capacity in sWAT is higher than in vWAT, the former plateaus at substantially lower levels of excess WAT mass than vWAT. In individuals with central or visceral obesity, lipid turnover is selectively increased in vWAT. Obese individuals classified as 'metabolically unhealthy' (according to ATPIII criteria) who have small subcutaneous adipocytes exhibit reduced triglyceride turnover. We conclude that excess WAT results in depot-specific differences in lipid turnover and increased turnover in vWAT and/or decreased turnover in sWAT may result in metabolic complications of overweight or obesity.

M3DISIM Project-Team

7. New Results

7.1. Mathematical and Mechanical Modeling

7.1.1. Modelling of collagen fibers elastic properties

Participants: Peter Baumgartner, Florent Wijanto, Jean-Marc Allain [correspondant], Matthieu Caruel [Univ. Paris-Est].

Our studies on collagen tissues have shown that the collagen fibers are able to elongate inelastically under stretch. In tendons, this effect has been attributed to the non-permanent cross-bridges that connect the different collagen fibrils (to assemble a fiber). This sliding effect appears experimentally to be reversible (at least partially) if the tissue is left long enough at its initial resting length. However, this sliding is classically included as an irreversible plastic response, or as a damage of the tissue. We are building a model based on a stochastic description of the binding and unbinding of the cross-bridges. This approach will enable us to have a microscopically based picture of the sliding, which will be able to explain some alterations in case of aging or pathologies of the tissue. At the moment, we have shown the importance of the density of cross-bridges in the cooperative response of the system. A publication is in preparation on the topic.

7.1.2. Multi-scale modeling of cardiac contraction

Participants: François Kimmig, Matthieu Caruel [Univ. Paris-Est], Dominique Chapelle [correspondant], Philippe Moireau.

This work aims at proposing a set of models of the muscular contraction targeting different scales in time and space and that can be used in the context of heart simulation. To this end, we developed so far two models using different approaches for the modeling of the force generating process at the molecular level called power stroke. First, we revised the standard chemo-mechanical models, which see the power stroke as a series of chemical states. Following the idea introduced by Truskinovsky and collaborators describing the power stroke as a continuum of mechanical states with the dynamics of the myosin head in the prescribed energy landscape governed by Langevin equations, we incorporated the attachment and detachment dynamics in the form of jump processes. In a second step, noting that the power stroke time scale is much shorter than that of heart contraction, we eliminated the power strokes dynamics and derived a two state – attached and detached – simplified model, each state being in fact associated with a statistical distribution of myosin head configurations. Both models have been integrated into our simulation framework CardiacLab, in order to benefit from the other modeling compartments available in the code, such as the geometrically reduced model of the heart left ventricle also developed in the team. These modeling elements will be confronted with experiments performed on cardiac muscle cells by collaborators in the team of Professor Lombardi at the University of Florence.

7.1.3. Mathematical and numerical modeling of shear waves propagation in the heart

Participants: Federica Caforio, Sébastien Imperiale [correspondant], Dominique Chapelle.

Shear acoustic waves remotely induced by the acoustic radiation force (ARF) of a focused ultrasound beam generated by piezoelectric sensors have been recently used in biomedical applications, e.g. in transient elastography techniques. By measuring the propagation velocity of generated shear waves in biological tissues, it is possible to locally assess biomechanical properties highly sensitive to structural changes corresponding to physiological and pathological processes. Recent experimental studies show the feasibility of applying transient elastography to the cardiac setting. In this context, the wave propagation induced by the ARF is superposed with the nonlinear mechanics associated with the heart deformation during the cardiac cycle. The aim of this work is to mathematically justify an original expression of the excitation induced by the ARF in nonlinear solids, based on energy considerations and asymptotic analysis. In soft media the propagation

velocity of shear waves ($1 - -10\text{m.s}^{-1}$) is much smaller than the velocity of pressure waves (1500m.s^{-1}). The approach we propose consists in considering a family of problems parametrised by a small parameter ε related to the velocity ratio between the two wave propagation phenomena, the high frequency of the piezoelectric source term and the viscosity. In order to derive a simplified model for the expression of ARF, we investigate the limit behavior of the solution for $\varepsilon \to 0$. We show that the leading term of the expansion is related to the underlying nonlinear mechanics of the heart deformation, and the first two correction terms correspond to a fast-oscillating pressure wave excited by the probe, and an elastic field having as source term a nonlinear function of the first corrector. This field corresponds to the shear acoustic wave induced by the ARF.

7.1.4. Analysis and 2-scale convergence of a heterogeneous microscopic bidomain model

Participants: Sébastien Imperiale [correspondant], Annabelle Collin [Monc].

The aim of this work is to provide a complete mathematical analysis of the periodic homogenization procedure that leads to the macroscopic bidomain model in cardiac electrophysiology. We consider space-dependent and tensorial electric conductivities as well as space-dependent physiological and phenomenological non-linear ionic models. We provide the nondimensionalization of the bidomain equations and derive uniform estimates of the solutions. The homogenization procedure is done using 2-scale convergence theory which enables us to study the behavior of the non-linear ionic models in the homogenization process.

7.1.5. A reduced thoracic model for inverse problem solving in seismocardiography

Participants: Alexandre Laurin, Sébastien Imperiale [correspondant], Dominique Chapelle, Philippe Moireau.

Seismocardiography (SCG) is the study of low-frequency (< 60 Hz) vibrations of the thorax caused by the beating heart. Although it is assumed that SCG signals are caused by forces applied on the interior of the thorax by the heart, no comprehensive model exists to describe the parameters and relationships that govern the system. The main goal of this study is to describe in some detail the filter applied by the thorax on cardiac forces, taking into account its zone of contact with the heart as well as the zone of measurement, i.e. the location of the accelerometer on a participant's chest. A secondary goal is to identify the smallest set of parameters capable of reproducing the filter, reducing the model while retaining its capacity to lend itself to physiological interpretation. Finally, we described a method to use the reduced model to estimate cardiac forces from measured thoracic accelerations. The overall aim of the study is to develop numerical methods that can augment the existing SCG interpretations to include mechanical indices of the heart muscle, and do so in real time.

7.2. Numerical Methods

7.2.1. Numerical methods for computing cyclic steady states

Participants: Ustim Khristenko, Patrick Le Tallec [correspondant].

This work is focused on two techniques for fast computing of the steady cyclic states of evolution problems in non-linear mechanics with space-time periodicity conditions. This kind of problems can be faced in various applications, for instance in the rolling of a tyre with periodic sculptures as well as in a beating heart. Direct solvers for such problems are not very convenient, since they require the inversion of very large matrices. In order to avoid this, a cyclic solution is usually computed as an asymptotic limit of the associated initial value problem with prescribed initial data. However, when the relaxation time is high, convergence to the limit cycle can be very slow. The first technique considered is the Newton-Krylov method, looking for the unknown initial state that provides the space-time periodic solution. This initial state is defined by the space-time periodicity condition, solved with the Newton-Raphson technique. Since the associated Jacobian cannot be expressed explicitly, the method uses one of the matrix-free Krylov iterative solvers. Using information stored while computing the residual to solve the linear system makes its calculation time negligible with respect to the residual calculation time. The second method is the delayed feedback control: an observer-controller type modification of the standard evolution to the limit cycle by introducing a feedback control term, based on

the periodicity error. The main result is the optimal form of the control term for a very general class of linear evolution problems, providing the fastest convergence to the cyclic solution. This control has also been adapted and tested for nonlinear problems. The methods discussed have been assessed using academic applications and they have also been implemented into the Michelin industrial code – applied to the rolling tyre model – as well as into the M3DISIM code for the cardiac contraction problem.

7.2.2. Solving isotropic elastodynamics using potentials

Participants: Sébastien Imperiale [correspondant], Jorge Albella.

This work has the potential to provide an original efficient method for the computations of elastic waves propagation in soft media (such as biological tissues), based on the property that pressure and shear waves decouple in isotropic media. Towards this direction, we considered the numerical solution of 2D elastodynamics isotropic equations using the decomposition of the displacement fields into potentials. This appears as a challenge for finite element methods, and we have addressed here the particular question of free boundary conditions. A stable (mixed) variational formulation of the evolution problem is proposed.

7.2.3. The Arlequin method for transient wave scattering by small obstacles

Participants: Sébastien Imperiale [correspondant], Jorge Albella.

In this work we extend the Arlequin method to overlapping domain decomposition technique for transient wave equation scattering by small obstacles. The main contribution of this work is to construct and analyze some variants of the Arlequin method from the continuous level to the fully discrete level. The constructed discretizations allow to solve wave propagation problems while using non-conforming and overlapping meshes for the background propagating medium and the surrounding of the obstacle, respectively. Hence we obtain a flexible and stable method in terms of the space discretization – an inf-sup condition is proven – while the stability of the time discretization is ensured by energy identities.

7.2.4. Construction of a fourth-order time scheme for dissipative wave equations

Participants: Sébastien Imperiale [correspondant], Juliette Chabassier [Magique-3d], Julien Diaz [Magique-3d].

This works deals with the construction of a fourth-order, energy preserving, explicit time discretization for dissipative linear wave equations. This discretization is obtained by replacing the inversion of a matrix – that comes naturally after using the technique of the Modified Equation on the second order Leap Frog scheme applied to dissipative linear wave equations – by an explicit approximation of its inverse. The stability of the scheme is studied first using an energy analysis, then an eigenvalue analysis. Numerical results in 1D illustrate the good behavior regarding space/time convergence and the efficiency of the newly derived scheme compared to more classical time discretizations. A loss of accuracy is observed for non-smooth profiles of dissipation, and we propose an extension of the method that fixes this issue. Finally, we assess the good performance of the scheme for a realistic dissipation phenomenon in Lorentz materials.

7.2.5. Coupled variational formulations of linear elasticity and the DPG methodology

Participant: Patrick Le Tallec [correspondant].

In this work, we develop a general approach akin to domain-decomposition methods to solve a single linear PDE, but where each subdomain of a partitioned domain is associated with a distinct variational formulation coming from a mutually well-posed family of so-called broken variational formulations of the original PDE. It can be exploited to solve challenging problems in a variety of physical scenarios where stability or a particular mode of convergence is desired in some part of the domain. The linear elasticity equations are solved in this work, but the approach can be applied to other equations as well. The broken variational formulations, which are essentially extensions of more standard formulations, are characterized by the presence of mesh-dependent broken test spaces and interface trial variables at the boundaries of the elements of the mesh. This allows necessary information to be naturally transmitted between adjacent subdomains, resulting in coupled variational formulations which are then proved to be globally well-posed. They are solved numerically using the DPG methodology, which is especially crafted to produce stable discretizations of broken formulations. Finally, expected convergence rates are verified in two different illustrative examples. This work has resulted in the publication [19].

7.2.6. A discontinuous Galerkin approach for cardiac electrophysiology

Participant: Radomir Chabiniok [correspondant].

Cardiac electrophysiology simulations are numerically challenging due to the propagation of a steep electrochemical wave front, and thus require discretizations with small mesh sizes to obtain accurate results. In this work – in collaboration with the Institute for Computational Mechanics, Technical University Munich and published in [21] – we present an approach based on the Hybridizable Discontinuous Galerkin method (HDG), which allows an efficient implementation of high-order discretizations into a computational framework. In particular using the advantage of the discontinuous function space, we present an efficient p-adaptive strategy for accurately tracking the wave front. HDG allows to reduce the overall degrees of freedom in the final linear system to those only on the element interfaces. Additionally, we propose a rule for a suitable integration accuracy for the ionic current term depending on the polynomial order and the cell model to handle high-order polynomials. Our results show that for the same number of degrees of freedom coarse high-order elements provide more accurate results than fine low-order elements. Introducing p-adaptivity further reduces computational costs while maintaining accuracy by restricting the use of high-order elements to resolve the wave front. For a patient-specific simulation of a cardiac cycle, p-adaptivity reduces the average number of degrees of freedom by 95% compared to the non-adaptive model. In addition to reducing computational costs, using coarse meshes with our p-adaptive high-order HDG method also simplifies practical aspects of mesh generation and postprocessing.

7.3. Inverse Problems

7.3.1. Discrete-time optimal filtering or Mortensen observer discretization

Participant: Philippe Moireau [correspondant].

In this work, we seek exact formulations of the optimal estimator and filter for a non-linear framework, as the Kalman filter is for a linear framework. The solution is well established with the Mortensen filter in a continuous-time setting, but we seek here its counterpart in a discrete-time context. We demonstrate that it is possible to pursue at the discrete-time level an exact dynamic programming strategy and we find an optimal estimator combining a prediction step using the model and a correction step using the data. This optimal estimator reduces to the discrete-time Kalman estimator when the operators are in fact linear. Furthermore, the strategy that consists of discretizing the least square criterion and then finding the exact estimator at the discrete level allows to determine a new time-scheme for the Mortensen filter which is proven to be consistent and unconditionally stable, with also a consistent and stable discretization of the underlying Hamilton-Jacobi-Bellman equation. This work has resulted in the publication [30].

7.3.2. An iterative method for identifying a stress-free state in image-based biomechanics

Participant: Martin Genet [correspondant].

Continued advances in computational power and methods have enabled image-based biomechanical modeling to become an important tool in basic science, diagnostic and therapeutic medicine, and medical device design. One of the many challenges of this approach, however, is identification of a stress-free reference configuration based on in vivo images of loaded and often prestrained or residually stressed soft tissues and organs. Fortunately, iterative methods have been proposed to solve this inverse problem, among them Sellier's method. This method is particularly appealing because it is easy to implement, converges reasonably fast, and can be coupled to nearly any finite element package. By means of several practical examples, however, we demonstrate that in its original formulation Sellier's method is not optimally fast and may not converge for problems with large deformations. Nevertheless, we can also show that a simple, inexpensive augmentation of Sellier's method based on Aitken's delta-squared process can not only ensure convergence but also significantly accelerate the method. This work has resulted in the publication [31].

7.3.3. A continuum finite strain formulation for finite element image correlation

Participant: Martin Genet [correspondant].

We propose a novel continuum finite strain formulation of the equilibrium gap principle – originally introduced in [Claire, Hild and Roux, 2004, Int. J. Num. Meth. Eng.] at the discrete level for linearized elasticity – used as a regularizer for finite element-based image correlation problems. Consistent linearization and finite element discretization is provided. The method is implemented using FEniCS & VTK, in a freely available Python library. The equilibrium gap constraint regularizes the image correlation problem, even in the presence of noise, and without affecting strain measurements.

7.3.4. Front shape similarity measure for Eikonal PDE data assimilation

Participants: Annabelle Collin [Monc], Philippe Moireau [correspondant].

We present a shape-oriented data assimilation strategy suitable for front-tracking problems through the example of wildfire. The concept of "front" is used to model, at regional scales, the burning area delimitation that moves and undergoes shape and topological changes under heterogeneous orography, biomass fuel and micrometeorology. The simulation-observation discrepancies are represented using a front shape similarity measure inspired from image processing and based on the Chan-Vese contour fitting functional. We show that consistent corrections of the front location and uncertain physical parameters can be obtained using this measure applied on a level-set fire growth model solving for an eikonal equation. This study involves a Luenberger observer for state estimation, including a topological gradient term to track multiple fronts, and a reduced-order Kalman filter for joint parameter estimation. We also highlight the need – prior to parameter estimation – for sensitivity analysis based on the same discrepancy measure, and for instance using polynomial chaos metamodels, to ensure that a meaningful inverse solution is achieved. The performance of the shape-oriented data assimilation strategy is assessed on a synthetic configuration subject to uncertainties in front initial position, near-surface wind magnitude and direction. The use of a robust front shape similarity measure paves the way toward the direct assimilation of infrared images and is a valuable asset in the perspective of data-driven wildfire modeling. This work has resulted in the publication [32].

7.3.5. The mechanism of monomer transfer between two distinct PrP oligomers

Participants: Aurora Armiento, Marie Doumic [Mamba], Philippe Moireau [correspondant].

In mammals, Prion pathology refers to a class of infectious neuropathologies whose mechanism is based on the self-perpetuation of structural information stored in the pathological conformer. The characterisation of the PrP folding landscape has revealed the existence of a plethora of pathways conducing to the formation of structurally different assemblies with different biological properties. However, the biochemical interconnection between these diverse assemblies remains unclear. The PrP oligomerisation process leads to the formation of neurotoxic and soluble assemblies called O1 oligomers with a high size heterodispersity. By combining the measurements in time of size distribution and average size with kinetic models and data assimilation, we revealed the existence of at least two structurally distinct sets of assemblies, termed Oa and Ob, forming O1 assemblies. We propose a kinetic model representing the main processes in prion aggregation pathway: polymerisation, depolymerisation, and disintegration. The two groups interact by exchanging monomers through a disintegration process that increases the size of Oa. Our observations suggest that PrP oligomers constitute a highly dynamic population. This work has resulted in the publication [14].

7.3.6. Joint-state and parameters estimation using ROUKF for HIV mechanistic models

Participants: Annabelle Collin [Monc], Philippe Moireau [correspondant], Mélanie Prague [Sism].

Various methods have been used in the statistical field to estimate parameters in mechanistic models. In particular, an approach based on penalised likelihood for the estimation of parameters in ordinary differential equations with non linear models on parameters (ODE-NLME) has proven successful. For instance, we consider the NIMROD program as a benchmark for estimation in these models. However, such an approach is time consuming. To circumvent this problem, we consider data assimilation approaches that historically arose in the context of geophysics. Here, we propose a Luenberger (also called nudging) state observer coupled with a parameter Kalman-based observer (RoUKF filter, also called SEIK filter) to perform a joint state and parameter estimation on a dataset composed of longitudinal observations of biomarkers for multiple patients. We compare these methods in terms of performances and computation time. We discuss how the concept of

random effect can be modeled using Kalman-based filter and its limitations. We illustrate both methods in simulation and on two datasets (the ALBI ANRS 070 trial and the Aquitaine cohort observational data) using an HIV mechanistic model.

7.4. Experimental Assessments

7.4.1. Microstructural interpretation of mouse skin mechanics from multiscale characterization

Participant: Jean-Marc Allain [correspondant].

Skin is a complex, multi-layered organ, with important functions in the protection of the body. The dermis provides structural support to the epidermal barrier, and thus has attracted a large number of mechanical studies. As the dermis is made of a mixture of stiff fibres embedded in a soft non-fibrillar matrix, it is classically considered that its mechanical response is based on an initial alignment of the fibres, followed by the stretching of the aligned fibres. Using a recently developed set-up combining multiphoton microscopy with mechanical assay, we imaged the fibres network evolution during dermis stretching. These observations, combined with a wide set of mechanical tests, allowed us to challenge the classical microstructural interpretation of the mechanical properties of the dermis: we observed a continuous alignment of the collagen fibres along the stretching. All our results can be explained if each fibre contributes by a given stress to the global response. This plastic response is likely due to inner sliding inside each fibre. The non-linear mechanical response is due to structural effects of the fibres network in interaction with the surrounding non-linear matrix. This multiscale interpretation explains our results on genetically-modified mice with a simple alteration of the dermis microstructure. This work has resulted in the publication [27].

7.4.2. Affine kinematics in planar fibrous connective tissues: an experimental investigation

Participants: Jean-Sébastien Affagard, Jean-Marc Allain [correspondant].

The affine transformation hypothesis is usually adopted in order to link the tissue scale with the fibers scale in structural constitutive models of fibrous tissues. Thanks to the recent advances in imaging techniques, such as multiphoton microscopy, the microstructural behavior and kinematics of fibrous tissues can now be monitored at different stretching within the same sample. Therefore, the validity of the affine hypothesis can be investigated. In this study, the fiber reorientation predicted by the affine assumption is compared with experimental data obtained during mechanical tests on skin and liver capsule coupled with microstructural imaging using multiphoton microscopy. The values of local strains and the collagen fibers orientation measured at increasing loading levels are used to compute a theoretical estimation of the affine reorientation of collagen fibers. The experimentally measured reorientation of collagen fibers during loading could not be successfully reproduced with this simple affine model. It suggests that other phenomena occur in the stretching process of planar fibrous connective tissues, which should be included in structural constitutive modeling approaches. This work has resulted in the publication [22].

7.4.3. Improving the experimental protocol for the identification of skin mechanical behavior

Participants: Jean-Sébastien Affagard, Florent Wijanto, Jean-Marc Allain [correspondant].

Mechanical properties of the skin, the external organ of the human body, are important for many applications such as surgery or cosmetics. Due to the highly hierarchical structure of the tissue, it is interesting to develop microstructural models that have better predictability and should reduce the consequences of sample variability. However, these models generally include a quite large number of mechanical parameters. Therefore, complex assays are required to achieve a proper identification of the microstructural models. We investigated in this study the best experimental protocol to identify a nonlinear, anisotropic, model of skin behavior, namely, the Holzapfel law, using displacement field and force measurements. This was done through a sensitivity analysis of the different parameters. We determined first the optimal assay, which appears to be a biaxial test with an alternated loading: first a stretch in one direction, then in the perpendicular one, and so on. To further improve the quality of the assay, we also determined the optimal geometry. Interestingly, slightly asymmetric geometries are more adequate than symmetric ones, while being easier to realise. This work has resulted in the publication [13].

7.4.4. How aging impacts skin biomechanics: a multiscale study in mice

Participants: Jean-Sébastien Affagard, Jean-Marc Allain [correspondant].

Skin aging is a complex process that strongly affects the mechanical behavior of skin. This study aims at deciphering the relationship between age-related changes in dermis mechanical behavior and the underlying changes in dermis microstructure. To that end, we use multiphoton microscopy to monitor the reorganization of dermal collagen during mechanical traction assays in ex vivo skin from young and old mice. The simultaneous variations of a full set of mechanical and microstructural parameters are analyzed in the framework of a multiscale mechanical interpretation. They show consistent results for wild-type mice as well as for genetically-modified mice with modified collagen V synthesis. We mainly observe an increase of the tangent modulus and a lengthening of the heel region in old murine skin from all strains, which is attributed to two different origins that may act together: (i) increased cross-linking of collagen fibers and (ii) loss of water due to proteoglycans deterioration, which impedes inner sliding within these fibers. In contrast, the microstructure reorganization upon stretching shows no age-related difference, which can be attributed to opposite effects of the decrease of collagen content and of the increase of collagen cross-linking in old mice. This work has resulted in the publication [28].

7.4.5. Recent advances in studying single bacteria and biofilm mechanics

Participant: Jean-Marc Allain [correspondant].

Bacterial biofilms correspond to surface-associated bacterial communities embedded in hydrogel-like matrix, in which high cell density, reduced diffusion and physico-chemical heterogeneity play a protective role and induce novel behaviors. We made a summary of the recent advances on the understanding of how bacterial mechanical properties, from single cell to high-cell density community, determine biofilm three-dimensional growth and eventual dispersion, and we attempt to draw a parallel between these properties and the mechanical properties of other well-studied hydrogels and living systems. This work has resulted in the publication [18].

7.5. Clinical Applications

7.5.1. Assessment of atrioventricular valve regurgitation using cardiac modeling

Participants: Radomir Chabiniok [correspondant], Philippe Moireau, Dominique Chapelle.

In this work, we introduce the modeling of atrioventricular valve regurgitation in a spatially reduced-order biomechanical heart model. The model can be fast calibrated using non-invasive data of cardiac magnetic resonance imaging and provides an objective measure of contractile properties of the myocardium in the volume overloaded ventricle, for which the real systolic function may be masked by the significant level of the atrioventricular valve regurgitation. After demonstrating such diagnostic capabilities, we show the potential of modeling to address some clinical questions concerning possible therapeutic interventions for specific patients. The fast running of the model allows targeting specific questions of referring clinicians in a clinically acceptable time. The work was presented at the “Functional Imaging and Modeling of the Heart” conference (FIMH 2017, Toronto, Canada) and is included in the conference proceedings [35].

7.5.2. Model for the dobutamine response in exercise-induced failure of the Fontan circulation

Participants: Radomir Chabiniok [correspondant], Philippe Moireau, Dominique Chapelle.

Understanding physiological phenomena and mechanisms of failure in congenital heart diseases is often challenging due to the complex hemodynamics and high inter-patient variations in anatomy and function. Computational modeling techniques have the potential to greatly improve the understanding of these complex diseases and provide patient-specific clues on mechanisms of deterioration and impact of treatments. This work employs a reduced 0D biomechanical heart model coupled with venous return to capture various key pathophysiological phenomena observed in patients with completed Fontan circulation – a complex surgically established circulation used to palliate patients in whom only one of the two ventricles is functionally able to support the vascular system – with exercise-induced heart failure during dobutamine stress. The framework we propose is fast, efficient and well-suited to the type of pathology and available clinical data obtained by a

combined cardiac catheterization and magnetic resonance imaging exam. We demonstrate that the outcomes of modeling are a valuable addition to the current clinical diagnostic investigations and explain patient-specific exercise hemodynamics, identify potential mechanisms of Fontan failure, and enable evaluation of a potential new therapy – selective heart rate modulation – in the treatment of patients with Fontan circulation. The paper is currently in preparation.

7.5.3. Heart and vessels modeling with data interaction for monitoring anesthetized patients

Participants: Arthur Le Gall, Radomir Chabiniok [correspondant], Fabrice Vallée, Dominique Chapelle.

By using mathematical models of heart and vessels developed in the team, we aim at improving intra-operative cardio-vascular safety of anesthetized patients. The patient-specific models, calibrated by echocardiography images and fed by continuous monitoring of aortic arterial pressure and aortic cardiac outflow would allow us to: 1) diagnose pathophysiological modifications associated with changes in the cardio-vascular state; 2) predict the drug response of the patient before the administration of the vaso-active treatment.

7.5.4. Intra-operative monitoring of cardiac afterload

Participants: Arthur Le Gall, Fabrice Vallée [correspondant].

General anesthesia leads to alterations of the cardiovascular system. Intra-operative arterial hypotension is linked to post-operative complications, but using vasopressors to treat arterial hypotension has shown conflicting results. Vasopressors act mainly by elevating cardiac afterload, which could be deleterious in fragile patients, in case of excessive response. Moreover, differences among the most used vasopressors have been observed in vivo [34]. The choice of vasopressor could be important to improve our patients' care. Consequently, we proposed a tool (Velocity-Pressure Loops) to continuously quantify changes in cardiac afterload [33]. Although the first work involves invasive measurement of aortic blood pressure and cardiac outflow, consistent results have been observed when Velocity-Pressure Loops were obtained by a radial arterial catheter with a mathematical transform function [23]. Those findings allow the usage of the Velocity-Pressure Loop without addition of any invasive device.

7.5.5. Review on extra-corporeal circulation

Participant: Arthur Le Gall [correspondant].

This clinical review [26] aims at describing the issues of the management of extra-corporeal membrane oxygenation (ECMO) in the Intensive Care Unit (ICU). From pathophysiology to the description of the impact on mortality, this document shows a global picture of current clinical practices.

7.5.6. On the importance of consistency in cardiac timings measurements

Participants: Arthur Le Gall, Alexandre Laurin, Fabrice Vallee [correspondant].

With the contribution of Denis Chemla, professor of Cardiology at Bicêtre Hospital, we presented this work at the CinC conference in Rennes [36]. In this work, we emphasize the need for a consistent method to measure systolic period duration, which is related to cardiac afterload and could be used to quantify arterial pressure amplification phenomenon.

MAMBA Project-Team

6. New Results

6.1. Analysis and control for population dynamics

Time asymptotics for nucleation, growth and division equations

We revisited the well-known Lifshitz-Slyozov model, which takes into account only polymerisation and depolymerisation, and progressively enriched the model. Taking into account depolymerisation and fragmentation reaction term may surprisingly stabilisde the system, since a steady size-distribution of polymers may then emerge, so that "Ostwald ripening" does not happen [33].

Cell population dynamics and its control

The question of optimal control of the population dynamics, that naturally arises when dealing with anticancer drug delivery optimisation, has been specifically the object of [24], work led in common with E. Trélat (LJLL and Inria team CAGE) and published in the *J. Maths. Pures Appl*.

The asymptotic behaviour of interacting populations in a nonlocal Lotka-Volterra way is also, independently of any control, studied for two populations in this article, and for many in [49].

Mathematical models of infectious diseases

First results in this subject (which is new for the team) have been obtained for elementary models including a model of vector-borne disease [31], [29].

6.2. Reaction and motion equations for living systems

Mathematical modelling for chemotaxis

A new kinetic model of chemotaxis for angiogenesis has been developed [22].

Aggregation equation.

Based on the approach relying on weak measure-valued solutions [100], an extension to a model for two species in interaction has been proposed in [12].

Free boundary problems for tumour growth.

Motivated by numerical observations from D. Drasdo using agent-based modelling, the article [17] studies the interfaces between two cell populations described by continuous models with different motilities and recovers interface instabilities.

6.3. Model and parameter identification combining stochastic and deterministic approaches in nonlocal and multi-scale models

Data assimilation and stochastic modelling for protein aggregation

Following Carola Kruse's post-doc [57], in collaboration with Tom Banks, Aurora Armiento's Ph.D [1], co-supervised with Philippe Moireau, was devoted to the question of adapting data assimilation strategies to the specific context and difficulties of protein aggregation.

In parallel with the statistical approach to growth and division processes, the deterministic approach has been continued in collaboration with Magali Tournus [35].

Estimating cellularity and tumour heterogeneity from Diffusion-Weighted MRI based on histological data

In [25] we developed, in close collaboration with the University of Heidelberg and DKFZ, together with I. Vignon-Clementel (Inria team REO), a procedure to estimate tumour heterogeneity and cellularity from Diffusion-Weighted Imaging (DWI) with calibration using histological data. The estimate is based on the intravoxel incoherent motion (IVIM) model that relates the DWI signal to water diffusion within each image voxel, as well as on an image processing and analysis procedure we developed for automated cell counting in large histological samples after tumour removal. We recently showed that biopsies routinely taken are likely to be sufficient to construct a calibration curve to relate DWI diffusion coefficient to cell density, and thus to infer the whole tumour heterogeneity. The biopsies have to be taken in regions of largely different diffusion values.

6.4. Focus on cancer

Modelling Acute Myeloid Leukaemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations

The collaboration with the DISCO team at Inria-Saclay has been continued in conference papers [26], [27]. In one of these papers, the concept of *dormancy* in cancer as a state of coexistence between tumour and healthy stem cell populations is studied using a new model.

Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control

This topic, main subject in Camille Pouchol's ongoing PhD thesis, has already been mentioned about Axis 1. It has led to the publication [24].

The general question of drug resistance in cancer, from biological observations to mathematical modelling and optimal control, has been reviewed in [14], [15] and presented in various international conferences and workshops.

Senescence modelling by telomere shortening

This work, following Sarah Eugène's PhD thesis, has been continued in collaboration with Zhou Xu at IBPC [13].

6.5. Growth, evolution and regeneration in populations and tissues

Amyloid disease

With Wei-Feng Xue in Canterbury, we continued to investigate the intrinsic variability among identical experiments of nucleation [78], [90], with recent results in [13].

Making use of data assimilation and statistical methods [52], we proposed new models and mechanisms and most recently we predicted the existence of several coexisting species of protein fibrils [2].

Dengue fever

The release of Wolbachia-infected mosquitoes in Dengue infested zones and the study of their propagation may be represented by spatial reaction-diffusion models. When implementing such a method, an important issue concerns the spatial propagation of the mosquitoes: on releasing infected mosquitoes in a given domain (which can be part of a city), the hope is to invade the whole area. The study of this propagation phenomena falls into the study of existence of travelling waves. We proposed in [125] a mathematical model to study such phenomena and have simplified it to recover a well-known simple bistable system for which existence of traveling wave is known. The study of the probability of success of spatial invasiveness has been performed in [126], and [41] is devoted to the blocking of the propagation in heterogeneous environment presenting strong enough population gradient. In the previous works, the invasion is installed by large enough impulsive deliveries. Another approach, consisting in igniting the propagation by feedback control, has been studied in [63], [6].

Toxicity extrapolation from in vitro to in vivo

The investigation of this field has been continued by Géraldine Cellière, leading to her PhD defense in June 2017 [71].

MONC Project-Team

6. New Results

6.1. Mathematical modeling of tumor-tumor distant interactions supports a systemic control of tumor growth

Authors: *Sébastien Benzekry*, Clare Lamont, Dominique Barbolosi, Lynn Hlatky, Philip Hahnfeldt. Paper published in Cancer Research, https://hal.inria.fr/hal-01566947.

Interactions between different tumors within the same organism have major clinical implications, especially in the context of surgery and metastatic disease. Three main explanatory theories (competition, angiogenesis inhibition and proliferation inhibition) have been proposed but precise determinants of the phenomenon remain poorly understood. Here we formalized these theories into mathematical models and performed biological experiments to test them with empirical data. In syngeneic mice bearing two simultaneously implanted tumors, growth of only one of the tumors was significantly suppressed (61% size reduction at day 15, $p<0.05$). The competition model had to be rejected while the angiogenesis inhibition and proliferation inhibition models were able to describe the data. Additional models including a theory based on distant cytotoxic log-kill effects were unable to fit the data. The proliferation inhibition model was identifiable and minimal (4 parameters), and its descriptive power was validated against the data, including consistency in predictions of single tumor growth when no secondary tumor was present. This theory may also shed new light on single cancer growth insofar as it offers a biologically translatable picture of how local and global action may combine to control local tumor growth, and in particular, the role of tumor-tumor inhibition. This model offers a depiction of concomitant resistance that provides an improved theoretical basis for tumor growth control and may also find utility in therapeutic planning to avoid post-surgery metastatic acceleration. Major findings In mice bearing two tumors implanted simultaneously, tumor growth was suppressed in one of the two tumors. Three theories of this phenomenon were advanced and assessed against the data. As formalized, a model of competition for nutrients was not able to explain the growth behavior as well as indirect, angiogenesis-regulated inhibition or a third model based on direct systemic inhibition. This last model offers a depiction of concomitant resistance that provides an improved theoretical basis for tumor growth control and may also find utility in therapeutic planning to avoid post-surgery metastatic acceleration.

6.2. Precision of manual two-dimensional segmentations of lung and liver metastases and its impact on tumour response assessment using RECIST 1.1

Authors: *François Cornelis*, *Marie Martin*, *Olivier Saut*, Xavier Buy, *Michèle Kind*, Jean Palussiere, *Thierry Colin*. Paper published in European Radiology Experimental, https://hal.inria.fr/hal-01634849.

Response evaluation criteria in solid tumours (RECIST) has significant limitations in terms of variability and reproducibility, which may not be independent. The aim of the study was to evaluate the precision of manual bi-dimensional segmentation of lung, liver metastases, and to quantify the uncertainty in tumour response assessment. Methods: A total of 520 segmentations of metastases from six livers and seven lungs were independently performed by ten physicians and ten scientists on CT images, reflecting the variability encountered in clinical practice. Operators manually contoured the tumours, firstly independently according to the RECIST and secondly on a preselected slice. Diameters and areas were extracted from the segmentations. Mean standard deviations were used to build regression models and 95% confidence intervals (95% CI) were calculated for each tumor size and for limits of progressive disease (PD) and partial response (PR) derived from RECIST 1.1. Results: Thirteen aberrant segmentations (2.5%) were observed without significant differences between the physicians and scientists; only the mean area of liver tumours ($p = 0.034$) and mean diameter

of lung tumours (p = 0.021) differed significantly. No difference was observed between the methods. Inter-observer agreement was excellent (intra-class correlation >0.90) for all variables. In liver, overlaps of the 95% CI with the 95% CI of limits of PD or PR were observed for diameters above 22.7 and 37.9 mm, respectively. An overlap of 95% CIs was systematically observed for area. No overlaps were observed in lung. Conclusions: Although the experience of readers might not affect the precision of segmentation in lung and liver, the results of manual segmentation performed for tumor response assessment remain uncertain for large liver metastases.

6.3. Dose- and time-dependence of the host-mediated response to paclitaxel therapy: a mathematical modeling approach

Authors: Madeleine Benguigui, Dror Alishekevitz, Michael Timaner, Dvir Shechter, Ziv Raviv, *Sebastien Benzekry*, Yuval Shaked. Paper published in OncoTarget, https://hal.inria.fr/hal-01672568.

It has recently been suggested that pro-tumorigenic host-mediated processes induced in response to chemotherapy counteract the anti-tumor activity of therapy, and thereby decrease net therapeutic outcome. Here we use experimental data to formulate a mathematical model describing the host response to different doses of paclitaxel (PTX) chemotherapy as well as the duration of the response. Three previously described host-mediated effects are used as readouts for the host response to therapy. These include the levels of circulating endothelial progenitor cells in peripheral blood and the effect of plasma derived from PTX-treated mice on migratory and invasive properties of tumor cells in vitro. A first set of mathematical models, based on basic principles of pharmacokinetics/pharmacodynamics, did not appropriately describe the dose-dependence and duration of the host response regarding the effects on invasion. We therefore provide an alternative mathematical model with a dose-dependent threshold, instead of a concentration-dependent one, that describes better the data. This model is integrated into a global model defining all three host-mediated effects. It not only precisely describes the data, but also correctly predicts host-mediated effects at different doses as well as the duration of the host response. This mathematical model may serve as a tool to predict the host response to chemotherapy in cancer patients, and therefore may be used to design chemotherapy regimens with improved therapeutic outcome by minimizing host mediated effects.

6.4. Tumor growth model of ductal carcinoma: from in situ phase to stroma invasion

Authors: *Olivier Gallinato, Thierry Colin, Olivier Saut, Clair Poignard* Paper published in Journal of Theoretical Biology, https://hal.inria.fr/hal-01598837.

This paper aims at modeling breast cancer transition from the in situ stage –when the tumor is confined to the duct– to the invasive phase. Such a transition occurs thanks to the degradation of the duct membrane under the action of specific enzymes so-called matrix metallo-proteinases (MMPs). The model consists of advection–reaction equations that hold in the duct and in the surrounding tissue, in order to describe the proliferation and the necrosis of the cancer cells in each subdomain. The divergence of the velocity is given by the increase of the cell densities. Darcy law is imposed in order to close the system. The key-point of the modeling lies in the description of the transmission conditions across the duct. Nonlinear Kedem-Katchalsky transmission conditions across the membrane describe the discontinuity of the pressure as a linear function of the flux. These transmission conditions make it possible to describe the transition from the in situ stage to the invasive phase at the macroscopic level. More precisely, the membrane permeability increases with respect to the local concentration of MMPs. The cancer cells are no more confined to the duct and the tumor invades the surrounding tissue. The model is enriched by the description of nutrients concentration, tumor necrosis factors, and MMPs production. The mathematical model is implemented in a 3D C ++-code, which is based on well-adapted finite difference schemes on Cartesian grid. The membrane interface is described by a level-set, and the transmission conditions are precisely approached at the second order thanks to well-suited sharp stencils. Our continuous approach provides new significant insights in the macroscopic modeling of the breast cancer phase transition, due to the membrane degradation by MMP enzymes.

6.5. Superconvergent second order Cartesian method for solving free boundary problem for invadopodia formation

Authors: Olivier Gallinato, Clair Poignard. Paper published in Journal of Computational Physics, https://hal.inria.fr/hal-01483484.

In this paper, we present a superconvergent second order Cartesian method to solve a free boundary problem with two harmonic phases coupled through the moving interface. The model recently proposed by the authors and colleagues describes the formation of cell protrusions. The moving interface is described by a level set function and is advected at the velocity given by the gradient of the inner phase. The finite differences method proposed in this paper consists of a new stabilized ghost fluid method and second order discretizations for the Laplace operator with the boundary conditions (Dirichlet, Neumann or Robin conditions). Interestingly, the method to solve the harmonic subproblems is superconvergent on two levels, in the sense that the first and second order derivatives of the numerical solutions are obtained with the second order of accuracy, similarly to the solution itself. We exhibit numerical criteria on the data accuracy to get such properties and numerical simulations corroborate these criteria. In addition to these properties, we propose an appropriate extension of the velocity of the level-set to avoid any loss of consistency, and to obtain the second order of accuracy of the complete free boundary problem. Interestingly, we highlight the transmission of the superconvergent properties for the static subproblems and their preservation by the dynamical scheme. Our method is also well suited for quasistatic Hele-Shaw-like or Muskat-like problems.

6.6. A Voronoi Interface approach to cell aggregate electropermeabilization

Authors: Arthur Guittet, *Clair Poignard*, Frederic Gibou.

In this work, a Voronoi Interface approach to the study of cell electropermeabilization is presented. We consider the nonlinear electropermeabilization model of Poignard et al., which takes into account the jump in the voltage potential across cells' membrane. The jump condition is imposed in a sharp manner, using the Voronoi Interface Method of Guittet et al., while adaptive Quad-/Oc-tree grids are employed to automatically refine near the cells boundary for increased accuracy. Numerical results are provided to illustrate the accuracy of the methods. We also carry out simulations in three spatial dimensions to investigate the influence of shadowing and of the cells shape on the degree of permeabilization.

6.7. Revisiting bevacizumab + cytotoxics scheduling using mathematical modeling: proof of concept study in experimental non-small cell lung carcinoma

Authors: D.C. Imbs, R. El Cheikh, A. Boyer, J. Ciccolini, C. Mascaux, B. Lacarelle, F. Barlesi, D. Barbolosi and *S. Benzekry*. Paper published in CPT Pharmacometrics Syst Pharmacol, https://hal.inria.fr/hal-01624423.

Concomitant administration of bevacizumab and pemetrexed-cisplatin is a common treatment for advanced nonsquamous non-small cell lung cancer (NSCLC). Vascular normalization following bevacizumab administration may transiently enhance drug delivery, suggesting improved efficacy with sequential administration. To investigate optimal scheduling, we conducted a study in NSCLC-bearing mice. First, experiments demonstrated improved efficacy when using sequential vs. concomitant scheduling of bevacizumab and chemotherapy. Combining this data with a mathematical model of tumor growth under therapy accounting for the normalization effect, we predicted an optimal delay of 2.8 days between bevacizumab and chemotherapy. This prediction was confirmed experimentally, with reduced tumor growth of 38% as compared to concomitant scheduling, and prolonged survival (74 vs. 70 days). Alternate sequencing of 8 days failed in achieving a similar increase in efficacy, thus emphasizing the utility of modeling support to identify optimal scheduling. The model could also be a useful tool in the clinic to personally tailor regimen sequences.

MYCENAE Project-Team

7. New Results

7.1. Numerical and theoretical studies of slow-fast systems with complex oscillations

7.1.1. Coupled multiple timescale dynamics in populations of endocrine neurons: Pulsatile and surge patterns of GnRH secretion

Participants: Elif Köksal Ersöz, Alexandre Vidal, Frédérique Clément.

We have finalized the study of a 6D extension of our model of GnRH pulse and surge generator, which has now been published [19]. The gonadotropin releasing hormone (GnRH) is secreted by hypothalamic neurons into the pituitary portal blood in a pulsatile manner. The alternation between a frequency-modulated pulsatile regime and the ovulatory surge is the hallmark of the GnRH secretion pattern in ovarian cycles of female mammals. In this work, we aimed at modeling additional features of the GnRH secretion pattern: the possible occurrence of a two-bump surge ("camel surge") and an episode of partial desynchronization before the surge. We have proposed a six-dimensional extension of a former four-dimensional model with three timescale and introduced two mutually-coupled, slightly heterogenous GnRH subpopulations (secretors) regulated by the same slow oscillator (regulator). We have considered two types of coupling functions between the secretors, including dynamic state-dependent coupling, and we have used numerical and analytic tools to characterize the coupling parameter values leading to the generation of a two-bump surge in both coupling cases. We have revealed the impact of the slowly varying control exerted by the regulator onto the pulsatile dynamics of the secretors, which leads to dynamic bifurcations and gives rise to desynchronization. To assess the occurrence time of desynchronization during the pulsatile phase, we have introduced asymptotic tools based on quasi-static and geometric approaches, as well as analytic tools based on the H-function derived from phase equation and numerical tracking of period-doubling bifurcations. We discuss the role of coupling parameters in the two-bump surge generation and the speed of desynchronization.

7.1.2. Wild oscillations in a nonlinear neuron model with resets

Participants: Jonathan Rubin [University of Pittsburgh], Justyna Signerska-Rynkowska, Jonathan Touboul, Alexandre Vidal.

We have finalized the work undergone in a series of two studies, where we have investigated the mechanisms by which complex oscillations are generated in a class of nonlinear dynamical systems with resets modeling the voltage and adaptation of neurons. These studies have been published as a two-part article [21], [22].

The first study [21] presents a mathematical analysis showing that the system can support bursts of any period as a function of model parameters, and that are organized in a period-incrementing structure. In continuous dynamical systems with resets, such period-incrementing structures are complex to analyze. In the present context, we have used the fact that bursting patterns correspond to periodic orbits of the adaptation map that governs the sequence of values of the adaptation variable at the resets. Using a slow-fast approach, we have shown that this map converges towards a piecewise linear discontinuous map whose orbits are exactly characterized. That map shows a period-incrementing structure with instantaneous transitions. We have further shown that the period-incrementing structure persists for the full system with non-constant adaptation, yet the transitions are more complex. We have also established the presence of chaos at the transitions.

The second study [22] shows that these neuron models can generically display a form of mixed-mode oscillations (MMOs), which are trajectories featuring an alternation of small oscillations with spikes or bursts (multiple consecutive spikes). The mechanism by which these are generated relies fundamentally on the hybrid structure of the flow: invariant manifolds of the continuous dynamics govern small oscillations, while discrete resets govern the emission of spikes or bursts, contrasting with classical MMO mechanisms in ordinary differential equations involving more than three dimensions and generally relying on a timescale separation. The decomposition of mechanisms reveals the geometrical origin of MMOs, allowing a relatively simple classification of points on the reset manifold associated to specific numbers of small oscillations. We have shown that the MMO pattern can be described through the study of orbits of a discrete adaptation map, which is singular as it features discrete discontinuities with unbounded left- and right-derivatives. We have studied the orbits of the map via rotation theory for circle maps and elucidated in detail complex behaviors arising in the case where MMOs display a single small oscillation per cycle.

7.1.3. Studies of the Petrov module for a family of generalized Liénard integrable systems

Participants: Lucile Megret [UPMC], Jean-Pierre Françoise [UPMC].

In [20], we have used the Lambert function in order to study a family of integrable generalized Liénard equations X_f which display a center. We have first proven a conjugation lemma inside a continuum of nested periodic orbits. Then we have deduced an explicit operator of Gelfand-Leray associated with the Hamiltonian of equation X_f. Afterwards, we have provided a generating family for the associated Petrov module. Finally, by using the Lambert function, we have studied the monotonicity of the Abelian integral of this generating family's elements.

7.2. Non conservative transport equations for cell population dynamics

7.2.1. Dimensional reduction of a multiscale model based on long time asymptotics

Participants: Frédérique Clément, Frédéric Coquel [CMAP], Marie Postel, Kim Long Tran.

We have finalized the study on the dimensional reduction of our multiscale model of terminal follicle development, which has now been published [17]. We have considered a class of kinetic models for which a moment equation has a natural interpretation. We have shown that, depending on their velocity field, some models lead to moment equations that enable one to compute monokinetic solutions economically. We have detailed the example of a multiscale structured cell population model, consisting of a system of 2D transport equations. The reduced model, a system of 1D transport equations, is obtained from computing the moments of the 2D model with respect to one variable. The 1D solution is defined from the solution of the 2D model starting from an initial condition that is a Dirac mass in the direction removed by reduction. For arbitrary initial conditions, we have compared 1D and 2D model solutions in asymptotically large time. Finite volume numerical approximations of the 1D reduced model can be used to compute the moments of the 2D solution with proper accuracy, both in the conservative and non conservative framework. The numerical robustness is studied in the scalar case, and a full scale vector case is presented.

7.2.2. Analysis and calibration of a linear model for structured cell populations with unidirectional motion : application to the morphogenesis of ovarian follicles

Participants: Frédérique Clément, Frédérique Robin, Romain Yvinec [INRA].

We have analyzed a multi-type age dependent model for cell populations subject to unidirectional motion, in both a stochastic and deterministic framework [23]. Cells are distributed into successive layers; they may divide and move irreversibly from one layer to the next. We have adapted results on the large-time convergence of PDE systems and branching processes to our context, where the Perron-Frobenius or Krein-Rutman theorem can not be applied. We have derived explicit analytical formulas for the asymptotic cell number moments, and the stable age distribution. We have illustrated these results numerically and we have applied them to the study of the morphodynamics of ovarian follicles. We have proven the structural parameter identifiability of our model in the case of age independent division rates. Using a set of experimental biological data, we have estimated the model parameters to fit the changes in the cell numbers in each layer during the early stages of follicle development.

This work has been undergone in the framework of the PhD of Frédérique Robin. It has been the matter of a poster at ReprosSciences2017 [24] (April 10-12) and of an oral presentation (Dynamiques de populations cellulaires structurées) at the annual meeting (September 27-29) of GDR MaMovi (Mathématiques Appliquées à la MOdélisation du VIvant).

7.2.3. Mathematical modeling of progenitor cell populations in the mouse cerebral cortex

Participants: Frédérique Clément, Alice Karam [IBPS], Matthieu Perez, Marie Postel, Sylvie Schneider-Maunoury [IBPS].

We have finalized the study of our PDE-based model of structured cell populations during the development of cerebral cortex. The model accounts for three main cell types: apical progenitors (APs), intermediate progenitors (IPs), and neurons. Each cell population is structured according to the cell age distribution. Since the model describes the different phases of the cell division cycle, we could derive the numeric equivalents of many of the experimental indexes measured in experimental setups, including classical mitotic or labeling indexes targeting the cells in phase S or mitosis, and more elaborated protocols based on double labeling with fluorescent dyes. We have formulated a multi-criterion objective function which enables us to combine experimental observations of different nature and to fit the data acquired in the framework of the NeuroMathMod project (Sorbonne-Universités Émergence call with IBPS, Institut de Biologie Paris Seine). Great efforts have been put on the experimental side to provide the model with the quantitative values of cell numbers for both progenitors and neurons. With the retrieved parameters, the model can provide useful information not supplied by the data, such as the cell origin of neurons (direct neurogenesis from AP or IPgenic neurogenesis) and the proportion of IPs cells undergoing several rounds of cell cycles. In addition, we have compared the cell dynamics patterns observed in wild-type mice with respect to mutant mice used as an animal model of human ciliopathies.

In the framework of the internship of Matthieu Perez (INSA Rouen, co-supervised by Frédérique Clément and Marie Postel), we have investigated numerically the link between our deterministic, PDE-based model of progenitor and neuron cell dynamics, and possible stochastic counterparts inspired from previous work in the team [31]. The deterministic approach is averaged with respect to the deterministic one, since it does not account for the trajectories of individual cells, yet it describes in more details the progression of cells within the cell cycle since it explicitly embeds the structuring of the cell cycle into different phases. The work has consisted in comparing the main model outputs (numbers of progenitors and neurons as a function of time) obtained by numerical simulations based on characteristics, on the deterministic side, or Gillespie algorithms, on the stochastic side. A proper strategy had to be settled to deal with the main difficulties raised by this comparison, namely the time-varying rates involved in the stochastic transition rates from one cell type to another, and the matching between the average stochastic rates and the deterministic rates ruling cell kinetics, especially the cell cycle duration.

NUMED Project-Team (section vide)

REO Project-Team

7. New Results

7.1. Mathematical and numerical analysis of fluid-structure interaction problems

Participants: Matteo Aletti, Ludovic Boilevin-Kayl, Chen-Yu Chiang, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Damiano Lombardi, Marc Thiriet, Marina Vidrascu.

In [15] a reduced order modeling method is developed to simulate multi-domain multi-physics problems. In particular we considered the case in which one problem of interest, described by a generic non-linear partial differential equation is coupled to one or several problems described by a set of linear partial differential equations. In order to speed up the resolution of the coupled system, a low-rank representation of the Poincaré-Steklov operator is built by a reduced-basis approach. A database for the secondary problems is built when the interface condition is set to be equal to a subset of the Laplace-Beltrami eigenfunctions on the surface. The convergence of the method is analysed and several 3D fluid-fluid and fluid-structure couplings are presented as numerical experiments.

In [43] we study an unsteady nonlinear fluid–structure interaction problem. We consider a Newtonian incompressible two-dimensional flow described by the Navier-Stokes equations set in an unknown domain depending on the displacement of a structure, which itself satisfies a linear wave equation or a linear beam equation. We prove existence of a unique local-in-time strong solution. In the case of the wave equation or a beam equation with inertia of rotation, this is, to our knowledge the first result of existence of strong solutions for which no viscosity is added. One key point, is to use the fluid dissipation to control, in appropriate function spaces, the structure velocity.

In [26] a fluid-structure interaction solver based on 3D Eulerian monolithic formulation for an incompressible Newtonian fluid coupled with a hyperelastic incompressible solid has been implemented, verified, and validated. It is based on a Eulerian formulation of the full system. After a fully implicit discretization in time, displacement is eliminated and the variational equation is solved for the velocity and pressure. Its main application in medicine is venous flow in inferior limbs.

7.2. Numerical methods for biological flows

Participants: Chloé Audebert, Ludovic Boilevin-Kayl, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Florian Joly, Alexandre This, Marc Thiriet, Irene Vignon Clementel.

Peripheral pulmonary artery stenosis (PPS) is a congenital abnormality resulting in pulmonary blood flow disparity and right ventricular hypertension, for which optimal surgical strategies remain unclear. In [35], we conduct a pilot study to use recently refined computational simulation in the setting of multiple surgical strategies and to examine the influence of pulmonary artery reconstruction on hemodynamics in this population. Obstruction relief along with pulmonary artery vasodilation determines postoperative pulmonary flow distribution and newer methods can incorporate these physiologic changes.

Incoming velocity at open boundaries, or backflow, often yields to unphysical instabilities already for moderate Reynolds numbers. Several treatments to overcome these backflow instabilities have been proposed in the literature. In [17], we present a set of benchmark problems in order to compare different methods in different backflow regimes (with a full reversal flow and with propagating vortices after a stenosis). The examples are implemented in FreeFem++ and the source code is openly available.

The simulation of cardiac blood flow using patient-specific geometries can help for the diagnosis and treatment of cardiac diseases. Current patient-specific cardiac flow simulations requires a significant amount of human expertise and time to pre-process image data and obtain a case ready for simulations. In [38] a new procedure is proposed to alleviate this pre-processing by registering a unique generic mesh on patient-specific cardiac segmentations and transferring appropriately the spatiotemporal dynamics of the ventricle. The method is applied on real patient data acquired from 3D ultrasound imaging. Both a healthy and a pathological conditions are simulated. The resulting simulations exhibited physiological flow behavior in cardiac cavities and the experiments confirm a significant reduction in pre-processing work.

In order to reduce the complexity of heart hemodynamics simulations, one-way coupling approaches are often considered as an alternative to fluid-structure interaction (FSI) models. A possible shortcoming of these simplified approaches is the difficulty to correctly capture the pressure dynamics during the isovolumetric phases. In [39] we propose an enhanced resistive immersed surface (RIS) model of cardiac valves which overcomes this issue. The benefits of the model are investigated and tested in blood flow simulations of the left heart.

In [51], a computational model of unsteady blood flow in the cerebral venous circuit inside the skull reconstructed from medical images has been carried out. This venous network runs separately from the arterial bed perfusing the brain. The major aspects are boundary conditions and flow governing parameters.

7.3. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi, Fabien Raphel, Eliott Tixier.

In [32], we propose a model to represent the electrical potential of cardiomyocytes derived from stem cells in Multi Electrodes Arrays (MEA). This model based on the bidomain equations and a model for the MEA electrodes is used to analyze experimental signals. Our numerical algorithm is able to provide for different drugs dose-response curves which are in very good agreement with known values.

In [14], we are interested in the electrical activity of cardiomyocytes under the action of drugs in MEA devices. We present numerical simulations based on the same model as in [32] enriched with a pore block model to assay the action of drugs. The simulation results show that the model properly reflects the main effects of several drugs on the electrical potential.

In [33] the variability of phenomena in cardiac electro-physiology is investigated by using a moment matching approach. The cells activity is described by parametric systems of Ordinary Differential Equations. Given the population statistics on a system observables (which is the action potential of the cells), the probability density distribution of the parameters is sought such that the statistics of the model outputs match the observed ones. An uncertainty quantification step is solved once for all by using a non-instrusive approach, and then the inverse problem is solved by introducing an entropy regularisation. Several numerical experiments are considered to validate the approach on realistic datasets.

In [34] a realistic application on the classification of the drugs effect on cardiac cells is investigated. In particular, the electrical activity of the cells is recorder by Micro Electrode Arrays in normal conditions and under drugs, at different concentrations. In order to perform a classification of a drug in terms of promoting or inhibit the activity of certain ion channels a machine learning approach is used (support vector machine). Since the data amount is not big and the variability and alea sources have a large impact on the signals recorded, the data set is augmented by in silico experiments. Several tests on realistic data are performed.

7.4. Lung and respiration modeling

Participants: Céline Grandmont, Dena Kazerani, Nicolas Pozin, Marina Vidrascu, Marc Thiriet, Irene Vignon Clementel.

In [30] we use the coupled model tree-parenchyma model introduced in [31] to study the impact of asthma on effort and ventilation distribution along with the effect of Heliox compared to air. Indeed, in spite of numerous clinical studies, there is no consensus on the benefit Heliox mixtures can bring to asthmatic patients in terms of work of breathing and ventilation distribution. For this study, lung surface displacement fields extracted from computed tomography medical images are used to prescribe realistic boundary conditions to the system. Asthma is simulated by imposing bronchoconstrictions to some airways of the tracheo-bronchial tree based on statistical laws deduced from the literature. This study illuminates potential mechanisms for patient responsiveness to Heliox when affected by obstructive pulmonary diseases. Responsiveness appears to be function of the pathology severity, as well as its distal position in the tracheo-bronchial tree and geometrical position within the lung. Moreover, as already stated, in asthma and COPD, some airways of the tracheo-bronchial tree can be constricted, from moderate narrowing up to closure. These pathological patterns affect the lung ventilation distribution. While some imaging techniques enable visualization and quantification of constrictions in proximal generations, no non-invasive technique provides precise insights on what happens in more distal areas. In [44] we propose a process that exploits dynamical lung ventilation measurements to access positions of airways closures in the tree. This identification approach combines our lung ventilation model along with a machine learning approach. Based on synthetic data generated with typical temporal and spatial resolutions as well as reconstruction errors, we obtain encouraging results with a detection rate higher than 90%.

The human tracheobronchial tree surface is covered with mucus that ensures clearance of foreign material. An alteration of mucus or its environment such as in cystic fibrosis dramatically impacts the mucociliary clearance. In [48] the numerical method is able to manage variations of more than 5 orders of magnitude in the shear rate and viscosity. It leads to a cartography that enables to discuss major issues on defective mucociliary clearance in cystic fibrosis. In addition, cystic fibrosis is associated with a shear-thinning mucus that tends to aggregate in regions of lower clearance. However, a rarefaction of periciliary fluid has a greater impact than the mucus shear-thinning.

7.5. Miscellaneous

Participants: Damiano Lombardi, Irene Vignon Clementel.

In [27] an adaptive tensor method is developed to build a parsimonious discretization for the kinetic equations, starting from separated, arbitrary and a priori chosen discretizations for the space and the velocity variables. The method automatically adapts the rank of the decomposition in order to ensure that a criterion on the residual of the equations is satisfied, and the proof of the convergence is provided. The method is tested on the Vlasov-Poisson equation but can be extended to other kinetic equations and to systems in which the domain is the cartesian product of separated domains.

In [42] an a posteriori error estimator for hermitian positive eigenvalue problem is proposed. This estimator, which is based on a residual formulation, is constructed by shifting the operators in such a way that the error between the exact eigenvalues and the approximated ones can be estimated efficiently. It is conditionally certified and sharp.

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumor treatment assessment; yet its relation to the underlying tissue structure is not clear. In [36], in order to link low-resolution but non-invasive DWI data with high resolution (invasive) histological information, we developed an image processing and analysis chain, which was used to study the correlation between the DWI diffusion coefficient and tumor cellularity from serial histological slides of a resected non-small cell lung cancer tumor.

SISTM Project-Team

7. New Results

7.1. Statistical and mechanistic modeling

- Prague M, Commenges D, Gran JM, Ledergerber B, Young J, Furrer H, Thiébaut R. Dynamic models for estimating the effect of HAART on CD4 in observational studies: Application to the Aquitaine Cohort and the Swiss HIV Cohort Study. Biometrics. 2017;73:294-304. [26]
  Comparison of descriptive models (Marginal structural models) and mechanistic models (Ordinary differential equations with mixed effect models on parameters) performances for estimating treatment effect from observational studies.
- Turner EL, Li F, Gallis JA, Prague M, Murray DM. Review of recent methodological developments in group-randomized trials: part 1 design. American Journal of Public Health 107(6), 907-915. [27]
- Mélanie Prague, Elisabeth Turner, Gallis John, Li Fan, Murray David. Review of recent Methodological Developments in group-randomized trials: Part 2 - Analysis. American Journal of Public Health, American Public Health Association, 2017. [28]
  Review of the literature on how to analyse data from a cluster randomised trial

  Review of literature regarding Design of cluster randomised trials.
- Jarne Ana, Daniel Commenges, Mélanie Prague, Yves Levy, Rodolphe Thiébaut and inspire 2/3 study group. Modelling CD4+ T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous interleukin 7. Annals of applied statistics 11(3) 1593-1616. [21]
  Modelling the disability of CD4 restauration by repeated cycles of Intereukine-7 injections using mechanistic models.
- Prague M., Wang R., Stephens A., Tchetgen Tchetgen E, DeGruttola V. Accounting for interference variables using semi-parametric augmentation for improving efficiency in clustered randomized trials with missing at random outcomes. Biometrics 72(4) 1066-1077. [29]
  Doubly robust approach to estimate the treatment effect in Cluster randomised trials.

7.2. Statistical learning methods for high-dimensional data

- Genuer R, Poggi J-M, Tuleau-Malot C, Villa-Vialaneix N. Random Forests for Big Data. Big Data Research, 9 (2017). [18]
  Addresses the analysis of Big Data with Random Forests, review of existing algorithms, simulation study and recommandations.
- Agniel D and Hejblum BP, Variance component score test for time-course gene set analysis of longitudinal RNA-seq data, Biostatistics, 18(4):589–604, 2017.[16]
  We propose tcgsaseq, a principled, model-free, and efficient method for detecting longitudinal changes in RNA-seq gene sets defined a priori. Applied to both simulated data and two real datasets, tcgsaseq is shown to exhibit very good statistical properties, with an increase in stability and power when compared to state-of-the-art methods
- Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaut R. Sequential Dirichlet Process Mixtures of Multivariate Skew t-distributions for Model-based Clustering of Flow Cytometry Data, preprint on ArXiv. [39]

  We propose to use a Bayesian nonparametric approach with Dirichlet process mixture of multivariate skew t-distributions to perform model based clustering of flow-cytometry data, robustly estimating the number of cell populations from the data.

7.3. Software tools

- Mougin F, Auber D, Bourqui R, Diallo G, Dutour I, Jouhet V, Thiessard F, Thiébaut R, Thébault P. Visualizing omics and clinical data: Which challenges for dealing with their variety? Methods. 2017. [23]
  This is a review on the methods to visualize the big data in the context of clinical research.
- Prague M., Wang R. and de Grutolla V. CRTgeeDR: An R package for generalized estimating equations with missing data in cluster randomized trials. R journal (in press). [29]
  Diffusion of a package to estimate estimate the intervention effect of a prevention strategy agains epidemics in cluster randomised trials. Estimation is based on GEE

7.4. Analysis of results from Clinical trials and cohorts in HIV

- Bouteloup V, Sabin C, Mocroft A, Gras L, Pantazis N, Le Moing V, d'Arminio Monforte A, Mary-Krause M, Roca B, Miro JM, Battegay M, Brockmeyer N, Berenguer J, Morlat P, Obel N, De Wit S, Fätkenheuer G, Zangerle R, Ghosn J, Pérez-Hoyos S, Campbell M, Prins M, Chêne G, Meyer L, Dorrucci M, Torti C, Thiébaut R; Standard Reference Distribution of CD4 Response to HAART Project Team for the Collaboration of Observational HIV Epidemiological Research Europe (COHERE) in EuroCoord. Reference curves for CD4 T-cell count response to combination antiretroviral therapy in HIV-1-infected treatment-naïve patients. HIV Medicine. 2017;18:33-44.
  This is a tool that should help clinicians to evaluate the immunological response to antiretroviral therapy in HIV infected patients. Thanks to the analyse of one of the largest observational database in the world, we provide with an online tool references on the CD4 count during the first year of antiretroviral therapy.
- Picat MQ, Pellegrin I, Bitard J, Wittkop L, Proust-Lima C, Liquet B, Moreau JF, Bonnet F, Blanco P, Thiébaut R; ANRS CO3 Aquitaine Cohort. Integrative Analysis of Immunological Data to Explore Chronic Immune T-Cell Activation in Successfully Treated HIV Patients. PLoS One. 2017;12:e0169164. [25]
  This is an analysis of data on gene expression and factors associated to the immune activation in HIV-infected patients. Using structural models, we disentangle the effect of factors such as CMV and the mediation through type 1 interferon pathway.
- Thiébaut R, Hue S, Le Marec F, Lelièvre JD, Dupon M, Foucat E, Lazaro E, Dabis F, Duffau P, Wittkop L, Surenaud M, Pellegrin I, Lacabaratz C, Bonnet F, Lévy Y; ANRS CO3 Aquitaine Cohort. Serum ST2 level is an independent predictor of all-cause mortality in HIV-infected patients. Aquitaine Cohort, France. AIDS. 2017 [32]
  In this work, we have demonstrated the independent effect of the biomarker ST2 on the overall mortality in a large cohort of HIV infected patients.
- Vladimir Novitsky, Mélanie Prague, Sikhulile Moyo, Tendani Gaolathe, Mompati Mmalane, et al.. High HIV-1 RNA among Newly Diagnosed People in Botswana. AIDS Research and Human Retroviruses, Mary Ann Liebert, To appear.

7.5. Analysis of results from Clinical trials and cohorts in Ebola

- Rechtien A, Richert L, Lorenzo H, Martrus G, Hejblum B, Dahlke C, Kasonta R, Zinser M, Stubbe H, Matschl U, Lohse A, Krähling V, Eickmann M, Becker S; VEBCON Consortium, Thiébaut R, Altfeld M, Addo MM. Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV. Cell Report. 2017;20:2251-2261. [30]
  In this work, we have analyzed high-dimensional gene expression and cell characterization data. We showed the predictive capacity of the innate immune response to the Ebola vaccine to define the antibody response established beyond one month. This is a successful application of integrative analyses tools on high dimensional immunogenicity data from an Ebola vaccine trial with identification of early correlates of later antibody responses.

7.6. Analysis of results from clinical trials and cohorts in other fields (Epidemiology, Medical Sciences, Neuroimaging, Sport Sciences)

- Zago L, Hervé PY, Genuer R, Laurent A, Mazoyer B, Tzourio-Mazoyer N, Joliot M. Predicting hemispheric dominance for language production in healthy individuals using support vector machine. Hum Brain Mapp. 2017 Dec;38(12):5871-5889. doi: 10.1002/hbm.23770. [34]
  Joint work with the GIN-IMN team, application of a variable selection procedure based on SVM method to analyze functional MRI data.
- Tabue-Teguo M, Grasset L, Avila-Funes JA, Genuer R, Proust-Lima C, Péres K, Féart C, Amieva H, Harmand MG, Helmer C, Salles N, Rainfray M, Dartigues JF. Prevalence and Co-Occurrence of Geriatric Syndromes in People Aged 75 Years and Older in France: Results From the Bordeaux Three-city Study. J Gerontol A Biol Sci Med Sci. (2017) [31]
  Application of Multiple Correspondence Analysis which enlights frailty and dependent profile of people from the Three-city study.
- Née M, Avalos M, Luxcey A, Contrand B, Salmi LR, Fourrier-Réglat A, Gadegbeku B, Lagarde E, Orriols L. Prescription medicine use by pedestrians and the risk of injurious road traffic crashes: A case-crossover study. PLoS Medicine. Jul 18;14(7):e1002347 (2017) [24]
  Exploration of the association between the use of medicinal drugs and the risk of being involved in a road traffic crash as a pedestrian. We applied the Lasso methodology that we previously developed for the case-crossover design in a high-dimensional setting. This design controls for time-invariant factors by using each case as its own control. This study highlights the necessity of improving awareness of the effect of medicines on pedestrians.
- Hellard P, Scordia C, Avalos M, Mujika I, Pyne DB. Modelling of optimal training load patterns during the 11 weeks preceding major competition in elite swimmers. Applied Physiology, Nutrition, and Metabolism. Jun 26 (2017) [20]
  Quantification of the relationships between the effects of periodization variables and competitive performance in elite swimmers using semiparametric mixed effects models. In the framework of the 2014-2016 R&D project "Quels schémas de périodisation pour la préparation des Jeux Olympiques à Rio ?" with the French Swimming Federation.
- Hejblum BP, Cui J, Lahey LJ, Cagan A, Sparks JA, Sokolove J, Cai T, Liao KP, Association between anti-citrullinated fibrinogen antibodies and coronary artery disease in rheumatoid arthritis, Arthritis Care & Research, in press, 2017. [19].
  We show that anti-cit-fibrinogen antibodies as a group were associated with CAD outcomes in our RA cohort, with the strongest signal for association arising from a subset of the autoantibodies.
- Liao KP, Sparks JA, Hejblum BP, Kuo IH, Cui J, Lahey LJ, Cagan A, Gainer VS, Liu W, Cai TT, Sokolove J, Cai T, Phenome-wide association study of autoantibodies to citrullinated and non-citrullinated epitopes in rheumatoid arthritis, Arthritis & Rheumatology, 69: 742–749, 2017. [22]
  We demonstrated application of a bioinformatics method, the PheWAS, to screen for the clinical significance of RA-related autoantibodies. Using the PheWAS approach, we identified potentially significant links between variations in the levels of autoantibodies and comorbidities of interest in RA.

7.7. Conferences

Members of the team were involved in 12 talks during conferences and colloquium.

Mélanie Prague has her work presented in 2017 in 2 peer-reviewed international conferences (Society of clinical trials Liverpool UK and Keystone symposium of mathematical modeling of virus infection, Este Park, May 2017).
Robin Genuer presented his work in the peer-reviewed International Conference of the European Research Consortium for Informatics and Mathematics Working Group (ERCIM WG) on Computational and Methodological Statistics, University of London, UK.

Boris Hejblum presented his work in the peer-reviewed 38th Annual Conference of the International Society for Clinical Biostatistics.
Chloé Pasin presented her work in the peer-reviewed Systems Immunology & Vaccine Design symposium, Heidelberg, Germany and the French Applied and Industrial Mathematics Society (SMAI) conference (Ronce-les-Bains).

Members of the team participated in French conferences: GDR Stat santé Bordeaux, GDR mathematical modelling of life Lyon and Journées de la statistique Francaise, Avignon (Perrine Soret, Mélanie Prague, Boris Hejblum). Mélanie Prague and Boris Hejblum also presented 4 posters in workshops.

XPOP Project-Team

7. New Results

7.1. Sampling from a log-concave distribution with compact support with proximal Langevin Monte Carlo

A detailed theoretical analysis of the Langevin Monte Carlo sampling algorithm was conducted when applied to log-concave probability distributions that are restricted to a convex body K. This method relies on a regularisation procedure involving the Moreau-Yosida envelope of the indicator function associated with K. Explicit convergence bounds in total variation norm and in Wasserstein distance of order 1 are established. In particular, we show that the complexity of this algorithm given a first order oracle is polynomial in the dimension of the state space.

7.2. Clustering and Model Selection via Penalized Likelihood for Different-sized Categorical Data Vectors

In this study, we consider unsupervised clustering of categorical vectors that can be of different size using mixture. We use likelihood maximization to estimate the parameters of the underlying mixture model and a penalization technique to select the number of mixture components. Regardless of the true distribution that generated the data, we show that an explicit penalty, known up to a multiplicative constant, leads to a non-asymptotic oracle inequality with the Kullback-Leibler divergence on the two sides of the inequality. This theoretical result is illustrated by a document clustering application. To this aim a novel robust expectation-maximization algorithm is proposed to estimate the mixture parameters that best represent the different topics. Slope heuristics are used to calibrate the penalty and to select a number of clusters.

7.3. Low-rank Interaction Contingency Tables

ontingency tables are collected in many scientific and engineering tasks including image processing, single-cell RNA sequencing and ecological studies. Low-rank methods have proved useful to analyze them, by facilitating visualization and interpretation. However, common methods do not take advantage of extra information which is often available, such as row and column covariates. We propose a method to denoise and visualize high-dimensional count data which directly incorporates the covariates at hand. Estimation is done by minimizing a Poisson log-likelihood and enforcing a low-rank structure on the interaction matrix with a nuclear norm penalty. We also derive theoretical upper and lower bounds on the Frobenius estimation risk. A complete methodology is proposed, including an algorithm based on the alternating direction method of multipliers, and automatic selection of the regularization parameter. The simulation study reveals that our estimator compares favorably to competitors. Then, analyzing environmental science data, we show the interpretability of the model using a biplot visualization. The method is available as an R package.

7.4. Online EM for functional data

A novel approach to perform unsupervised sequential learning for functional data is proposed. The goal is to extract reference shapes (referred to as templates) from noisy, deformed and censored realizations of curves and images. The proposed model generalizes the Bayesian dense deformable template model, a hierarchical model in which the template is the function to be estimated and the deformation is a nuisance, assumed to be random with a known prior distribution. The templates are estimated using a Monte Carlo version of the online Expectation–Maximization (EM) algorithm. The designed sequential inference framework is significantly more computationally efficient than equivalent batch learning algorithms, especially when the missing data is high-dimensional. Some numerical illustrations on curve registration problem and templates extraction from images are provided to support the methodology