RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2016

Section New Results

Edition: 2017-08-25

ABS Project-Team

6. New Results

6.1. Modeling interfaces and contacts

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

6.1.1. Predicting binding poses and affinities for protein - ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation

Participants: Frédéric Cazals, Simon Marillet.

In collaboration with Sergei Grudinin, Maria Kadukova and Andreas Eisenbarth (Univ. Grenoble Alpes / CNRS / Inria, France).

The 2015 D3R Grand Challenge provided an opportunity to test our new model for the binding free energy of small molecules [17], as well as to assess our protocol to predict binding poses for protein-ligand complexes. Our pose predictions were ranked 3-9 for the HSP90 dataset, depending on the assessment metric. For the MAP4K dataset the ranks are very dispersed and equal to 2-35, depending on the assessment metric, which does not provide any insight into the accuracy of the method. The main success of our pose prediction protocol was the re-scoring stage using the recently developed Convex-PL potential. We make a thorough analysis of our docking predictions and discuss the effect of the choice of rigid receptor templates, the number of flexible residues in the binding pocket, the binding pocket size, and the subsequent re-scoring.

However, the main challenge was to predict experimentally determined binding affinities for two blind test sets. Our affinity prediction model consisted of two terms, a pairwise-additive enthalpy, and a non pairwise-additive entropy. We trained the free parameters of the model with a regularized regression using affinity and structural data from the PDBBind database. Our model performed very well on the training set, however, failed on the two test sets. We explain the drawback and pitfalls of our model, in particular in terms of relative coverage of the test set by the training set and missed dynamical properties from crystal structures, and discuss different routes to improve it.

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

Participants: Frédéric Cazals, Simon Marillet.

In collaboration with Pierre 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 [23], 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.

6.2. Modeling macro-molecular assemblies

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

No new result on this topic in 2016.

6.3. Modeling the flexibility of macro-molecules

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

6.3.1. Energy landscapes and persistent minima

Participants: Frédéric Cazals, Dorian Mazauric.

In collaboration with David Wales and Joanne Carr, from Cambridge University (UK).

In this work [15], we consider a coarse-graining of high-dimensional potential energy landscapes based upon persistences—which correspond to lowest barrier heights to lower-energy minima. Persistences can be calculated efficiently for local minima in kinetic transition networks that are based on stationary points of the prevailing energy landscape. The networks studied here represent peptides, proteins, nucleic acids, an atomic cluster, and a glassy system. Minima with high persistence values are likely to represent some form of alternative structural morphology, which, if appreciably populated at the prevailing temperature, could compete with the global minimum (defined as in- finitely persistent). Threshold values on persistences (and in some cases equilibrium occupation probabilities) have therefore been used in this work to select subsets of minima, which were then analysed to see how well they can represent features of the full network. Simplified disconnectivity graphs showing only the selected minima can convey the funnelling (including any multiple-funnel) characteristics of the corresponding full graphs. The effect of the choice of persistence threshold on the reduced disconnectivity graphs was considered for a system with a hierarchical, glassy landscape. Sets of persistent minima were also found to be useful in comparing networks for the same system sampled under different conditions, using minimum oriented spanning forests.

6.3.2. Hybridizing rapidly growing random trees and basin hopping yields an improved exploration of energy landscapes

Participants: Frédéric Cazals, Tom Dreyfus, Andrea Roth.

In collaboration with Charles Robert (IBPC / CNRS, Paris, France).

The number of local minima of the potential energy landscape (PEL) of molecular systems generally grows exponentially with the number of degrees of freedom, so that a crucial property of PEL exploration algorithms is their ability to identify local minima which are low lying and diverse. In this work [18], we present a new exploration algorithm, retaining the ability of basin hopping (BH) to identify local minima, and that of transition based rapidly exploring random trees (T-RRT) to foster the exploration of yet unexplored regions. This ability is obtained by interleaving calls to the extension procedures of BH and T-RRT, and we show tuning the balance between these two types of calls allows the algorithm to focus on low lying regions. Computational efficiency is obtained using state-of-the art data structures, in particular for searching approximate nearest neighbors in metric spaces. We present results for the BLN69, a protein model whose conformational space has dimension 207 and whose PEL has been studied exhaustively. On this system, we show that the propensity of our algorithm to explore low lying regions of the landscape significantly outperforms those of BH and T-RRT.

6.4. Algorithmic foundations

Keywords: computational geometry, Computational topology, Voronoi diagrams, α-shapes, Morse theory, graph algorithm, combinatorial optimization, statistical learning.

6.4.1. The Structural Bioinformatics Library: modeling in biomolecular science and beyond

Participants: Frédéric Cazals, Tom 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, http://sbl. inria.fr) [20], 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). 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 tools can also be combined to tackle integrated analysis problems. For developers, the SBL provides a broad C++ toolbox with modular design, involving (2) core 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.

6.4.2. Optimal transportation problems with connectivity constraints

Participants: Frédéric Cazals, Dorian Mazauric.

The earth mover distance (EMD) or the Mallows distance are example optimal transportation (OT) problems reducing to linear programs. In this work [21], we study a generalization of these problems when the supply and demand nodes are the vertices of two graphs called the supply and the demand graphs. The novel problems embed connectivity constraints in the transport plans computed, using a Lipschitz-like condition involving distances between certain subgraphs of the supply graph and certain subgraphs of the demand graph. More precisely, we make three contributions.

First, we formally introduce two optimal transportation problems generalizing EMD, namely Minimum-cost under flow, transport size, and connectivity constraints problem (problem EMD-FCC) and Maximum-flow under cost, transport size, and connectivity constraints problem (problem EMD-CCC). We prove that problems EMD-CCC and EMD-FCC are NP-complete, and that EMD-FCC is hard to approximate within any given constant. Second, we develop a greedy heuristic algorithm returning admissible solutions, of time complexity $O(n^3m^2)$ with n and m the numbers of vertices of the supply and demand graphs, respectively. Third, on the experimental side, we apply our novel OT algorithms for two applications, namely the comparison of clusterings, and the analysis of so-called potential energy landscapes in molecular science. These experiments show that optimizing the transport plan and respecting connectivity constraint can be competing objectives. Implementations of our algorithms are available in the Structural Bioinformatics Library at http://sbl.inria.fr.

6.4.3. Clustering stability revealed by matchings between clusters of clusters

Participants: Frédéric Cazals, Dorian Mazauric, Romain Tetley.

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 [22], 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. Altogether, these pieces of information help assessing the coherence between two clusterings.

More specifically, 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 (k,D) and D-family-matching problems on intersection graphs, with k the number of meta-clusters and D the upper-bound on the diameter of the graph induced by the clusters of any meta-cluster. First we prove hardness and inapproximability results. Second, we design exact polynomial time dynamic programming algorithms for some classes of graphs (in particular trees). Then, we prove efficient (exact, approximation, and heuristic) algorithms, based on spanning

trees, for general graphs. Practically, we present extensive experiments in two directions. First, we illustrate the ability of our algorithms to identify relevant meta-clusters between a given clustering and an edited version of it. Second, we show how our methods can be used to identify notorious instabilities of the k-means algorithm.

6.4.4. Experimental evaluation of a branch and bound algorithm for computing pathwidth

Participant: Dorian Mazauric.

In collaboration with David Coudert and Nicolas Nisse (COATI project-team, Université Côte D'Azur, Inria, I3S / CNRS).

Path-decompositions of graphs are an important ingredient of dynamic programming algorithms for solving efficiently many NP-hard problems. Therefore, computing the pathwidth and associated path-decomposition of graphs has both a theoretical and practical interest. In [16], we design a Branch and Bound algorithm that computes the exact pathwidth of graphs and a corresponding path-decomposition. Our main contribution consists of several non-trivial techniques to reduce the size of the input graph (pre-processing) and to cut the exploration space during the search phase of the algorithm. We evaluate experimentally our algorithm by comparing it to existing algorithms of the literature. It appears from the simulations that our algorithm offers a significant gain with respect to previous work. In particular, it is able to compute the exact pathwidth of any graph with less than 60 nodes in a reasonable running-time (≤ 10 minutes on a standard laptop). Moreover, our algorithm achieves good performance when used as a heuristic (i.e., when returning best result found within bounded time-limit). Our algorithm is not restricted to undirected graphs since it actually computes the directed pathwidth which generalizes the notion of pathwidth to digraphs.

6.4.5. Extracting the core structural connectivity network: guaranteeing network connectedness through a graph-theoretical approach

Participant: Dorian Mazauric.

In collaboration with Demian Wassermann, Guillermo Gallardo-Diez and Rachid Deriche (ATHENA project-team, Université Côte d'Azur, Inria).

In this work [19], we present a graph-theoretical algorithm to extract the connected core structural connectivity network of a subject population. Extracting this core common network across subjects is a main problem in current neuroscience. Such network facilitates cognitive and clinical analyses by reducing the number of connections that need to be explored. Furthermore, insights into the human brain structure can be gained by comparing core networks of different populations. We show that our novel algorithm has theoretical and practical advantages. First, contrary to the current approach our algorithm guarantees that the extracted core subnetwork is connected agreeing with current evidence that the core structural network is tightly connected. Second, our algorithm shows enhanced performance when used as feature selection approach for connectivity analysis on populations.

6.4.6. On the complexity of the representation of simplicial complexes by trees

Participant: Dorian Mazauric.

In collaboration with Jean-Daniel Boissonnat (DataShape team, Université Côte d'Azur, Inria).

In this paper [14], we investigate the problem of the representation of simplicial complexes by trees. We introduce and analyze local and global tree representations. We prove that the global tree representation is more efficient in terms of time complexity for searching a given simplex and we show that the local tree representation is more efficient in terms of size of the structure. The simplicial complexes are modeled by hypergraphs. We then prove that the associated combinatorial optimization problems are very difficult to solve and to approximate even if the set of maximal simplices induces a planar graph of maximum degree at most three or a bounded degree hypergraph. However, we prove polynomial time algorithms that compute constant factor approximations and optimal solutions for some classes of instances.

6.4.7. Well balanced designs for data placement

Participant: Dorian Mazauric.

In collaboration with Jean-Claude Bermond (COATI project-team, Université Côte D'Azur, Inria, I3S / CNRS), Alain Jean-Marie (MAESTRO project-team, Université Côte D'Azur, Inria) and Joseph Yu (Department of Mathematics, UFV, Abbotsford, BC, Canada).

The problem we consider in [13] is motivated by data placement, in particular data replication in distributed storage and retrieval systems. We are given a set V of v servers along with b files (data, documents). Each file is replicated on exactly k servers. A placement consists in finding a family of b subsets of V (representing the files) called blocks, each of size k. Each server has some probability to fail and we want to find a placement that minimizes the variance of the number of available files. It was conjectured that there always exists an optimal placement (with variance better than any other placement for any value of the probability of failure). We show that the conjecture is true, if there exists a well-balanced design, that is a family of blocks, each of size k, such that each j-element subset of V, $1 \leq j \leq k$, belongs to the same or almost the same number of blocks (difference at most one). The existence of well-balanced designs is a difficult problem as it contains, as a subproblem, the existence of Steiner systems. We completely solve the case math formula and give bounds and constructions for math formula and some values of v and b.

AMIB Project-Team

5. New Results

5.1. RNA design

We extended our previous results on RNA design [29], obtained in collaboration with J. Hales, J. Manuch and L. Stacho (Simon Fraser University/Univ. British Columbia, Canada).

Our results provided complete characterizations for the structures that can be designed using restricted alphabets. We provided a complete characterization of designable structures without unpaired bases. When unpaired bases are allowed, we provided partial characterizations for classes of designable/undesignable structures, and showed that the class of designable structures is closed under the stutter operation. Membership of a given structure to any of the classes can be tested in linear time and, for positive instances, a solution could be found in linear time. Finally, we considered a structure-approximating version of the problem that allows to extend helices and, assuming that the input structure avoids two motifs, we provided a linear-time algorithm that produces a designable structure with at most twice more base pairs than the input structure, as illustrated by Fig. 3 .

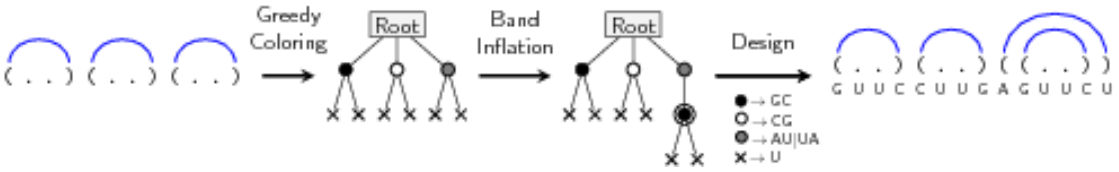


Figure 3. Principle of our structure-approximating version of RNA design: Starting from a potentially undesignable structure, a greedy coloring can be performed and corrected such that the final structure is provably designable in linear time.

In a manuscript accepted for publication in Algorithmica [4], we have shown that our previous results [29] 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.

We also initiated a collaboration with Danny Barash's group at Ben-Gurion university (Israel). We contributed a review of existing tools and techniques for RNA design, to appear as an article within the *Briefings in Bioinformatics* series [2]. We also combined previously contributed methods for design into a new method and web-server for the design of RNAs [3]. This collaboration stemed from the observation that `IncaRNAtion` [36], a random generation algorithm for RNA design recently developed in collaboration with Jérôme Waldispühl's group at McGill University (Montreal, Canada), produced excellent starting points (seed) for classic algorithms based on local-search. In particular, the combination of `IncaRNAtion` and `RNAfbInv` [45] was found to yield particularly promising candidates for design.

5.2. Algorithmics and combinatorics of motifs occurrences

We have developed a new algorithm to compute minimal absent words in external memory. Minimal absent words are used in sequence comparison [23] or to detect biologically significant events. For instance, it was

shown that there exist three minimal words in Ebola virus genomes which are absent from human genome [42]. The identification of such specific-species sequences may prove to be useful for the development of diagnosis and therapeutics. We have already provide an $O(n)$-time and $O(n)$-space algorithm to compute minimal absent words, with an implementation that can be executed in parallel. However these implementations require a large amount of RAM, thus they cannot be used for the human genome on a desktop computer. In our new contribution we developed an external memory implementation, it can compute minimal absent words of length at most 11 for the human genome using only 1GB of RAM in less than 4 hours (manuscript submitted [16]).

Repetitive patterns in genomic sequences have a great biological significance. This is a key issue in several genomic problems as many repetitive structures can be found in genomes. One may cite microsatellites, retrotransposons, DNA transposons, long terminal repeats (LTR), long interspersed nuclear elements (LINE), ribosomal DNA, short interspersed nuclear elements (SINE). Knowledge about the length of a maximal repeat also has algorithmic implications, most notably the design of assembly algorithms that rely upon de Bruijn graphs.

Analytic combinatorics allowed us to derive formula for the expected length of repetitions in a random sequence [9]. The originality of the approach is the demonstration of a Large Deviation principle and the use of Lagrange multipliers. This allowed for a generalization of previous works on a binary alphabet. Simulations on random sequences confirmed the accuracy of our results. As an application, the sample case of Archaea genomes illustrated how biological sequences may differ from random sequences, and in turns provides tools to extract repetitive sequences.

5.3. Integrative RNA structural modeling

To circumvent expensive, low-throughput, 3D experimental techniques such as X-ray crystallography, a low resolution/high throughput technology called SHAPE is increasingly favored for structural modeling by structural biologists.

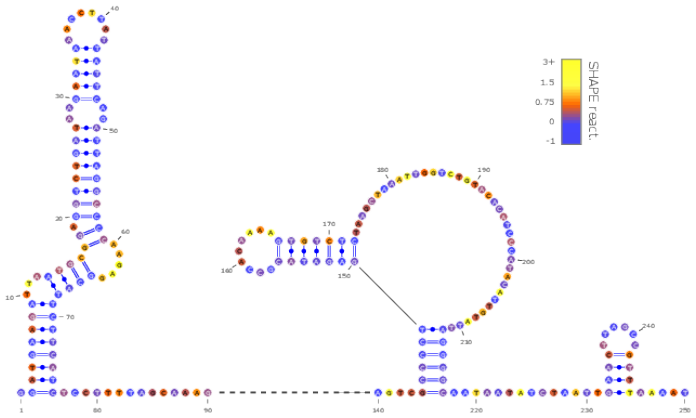
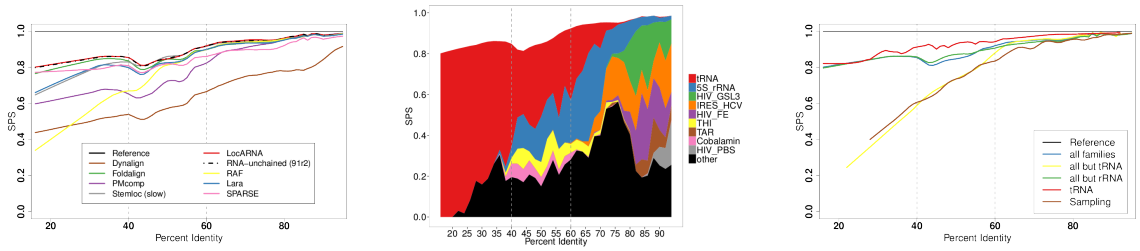


Figure 4. Conserved and thermodynamically-stable structure elements revealed by our analysis of an Ebola UTR region.

Within Afaf Saaidi's thesis, funded by the *Fondation pour la Recherche Médicale* and co-supervised by Bruno Sargueil at Faculté de Pharmacie of Université Paris V, we have developed integrative modeling strategies based on Boltzmann sampling. Preliminary results, obtained by applying these methods to model the structures of 3'UTR regions in Ebola, were presented at JOBIM 2016 [14].

Moreover, in collaboration with McGill University (Canada), we cross-examined mutate-and-map data (MaM [30]) in the light of evolutionary data. MaM data consist in the sequential SHAPE probing of a set of mutant RNAs, obtained by systematic point-wise mutations, to highlight structurally-dependent nucleotides, later to use dependent pairs as constraints in (an automated) structural modeling. We chose to adopt an alternative perspective on MaM data, and used the perturbation of the SHAPE profiles as a proxy for the structural disruption induced by a mutation. Disruptive mutations are rescued within homologs, *i.e.* compensated to re-establish the structure. However, our analysis also revealed the existence of non-structurally local (neither on the 2D or 3D levels) nucleotides which have significant mutual-information with highly disruptive positions, despite not being involved in any obvious compensatory relationship.

We hypothesized that such mutations are revealing of interactions involving RNA. In a manuscript published in *Nucleic Acids* journal, we tested and validated this hypothesis by showing its capacity to discriminate discriminative positions that are known to be in contact with specific ligands (proteins, DNA, small molecules...) [10].



a. *b.* *c.*

Figure 5. Typical software for comparative RNA structure prediction exhibit a dent in performance within the 40%-60% sequence identity range when benchmarked using the popular Bralibase data set (a.). However, this is due to the overrepresentation of a well-predicted type of RNA (tRNAs, red area) for low-identity ranges (b.). A re-evaluation of state-of-the-art software on an unbiased (c., brown line) reveals much more modest predictive capacities than initially believed in the community.

A fruitful line of research for RNA structure prediction is based on a comparative approach. Whenever homologous RNAs are identified, a classic strategy is to perform a simultaneous alignment and folding of several RNAs. Many software (30+) have been contributed over the past decades for this problem, leading to the introduction of benchmarks, one of the most prominent being the `Bralibase`, to position new developments and identify axes of progression. One such desired improvement, as illustrated in Figure 5 , was the difficulties experienced by most software around the 40-60% sequence identity range, which was believed to arise from deep algorithmic reasons. In collaboration with Cedric Chauve (Simon Fraser University, Canada) Benedikt Löwes and Robert Giegerich (Bielefeld University, Germany), we showed that this perceived difficulty was simply the consequence of a strong bias towards tRNAs in the 40-60% sequence identity region. Moreover, we argued that the overall performance of existing tools for low sequence identities were largely overestimated [8].

Finally, we presented at JOBIM 2016 an efficient implementation, called `LicoRNA`, of our parameterized complexity algorithm based on tree-decomposition for the sequence/structure alignment of RNA [15]. Specifically, we showed that our `LicoRNA`, by including an expressive scoring scheme and capturing pseudoknots of arbitrary complexity, generally outperforms previously contributions for the problem.

5.4. Combinatorial foundations

Pairwise ordered tree alignment are combinatorial objects that appear in RNA secondary structure comparison. However, the usual representation of tree alignments as supertrees is ambiguous, i.e. two distinct supertrees

may induce identical sets of matches between identical pairs of trees. This ambiguity is uninformative, and detrimental to any probabilistic analysis. In a recent collaboration with Cédric Chauve (SFU Vancouver, Canada) and Julien Courtiel (LIPN, Paris XII) presented at the ALCOB'16 conference, we considered tree alignments up to equivalence [11]. Our first result was a precise asymptotic enumeration of tree alignments, obtained from a context-free grammar by means of basic analytic combinatorics. Our second result focused on alignments between two given ordered trees. By refining our grammar to align specific trees, we obtained a decomposition scheme for the space of alignments, and used it to design an efficient dynamic programming algorithm for sampling alignments under the Gibbs-Boltzmann probability distribution. This generalizes existing tree alignment algorithms, and opens the door for a probabilistic analysis of the space of suboptimal RNA secondary structures alignments.

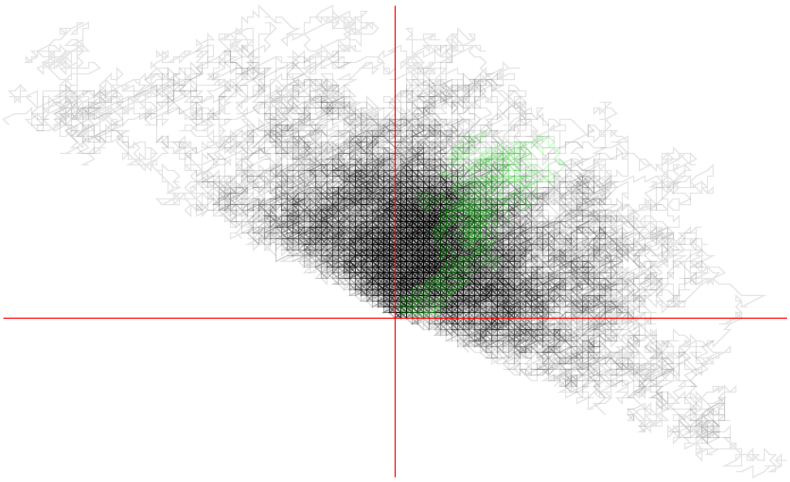


Figure 6. Random 2D walks (green walks) confined in the positive can be generated efficiently by performing rejection from a well-chosen 1D model (black walks) [12].

Finally, in collaboration with Marni Mishna (Simon Fraser University, Canada) and Jérémie Lumbroso (Princeton University, USA), we considered the uniform random generation of *difficult*, or *reluctant*, 2D discrete walks that remain confined in the positive quarter plane. We proposed a naive dynamic programming algorithm having complexity $O(n^4)$ for any step set. We also exploited the remark that any quarterplane walks can be transformed into a well-chosen 1D model having the same exponential growth factor. However, such a 1D model takes irrational steps, leading us to explore new avenues for the random generation. This work was presented at the GASCOM'16 conference [12].

5.5. Comparative genomics

D. Iakovishina defended in 2015 a PhD thesis co-advised by M. Régnier and V. Boeva (Curie Institute). She proposed a new computational method to detect structural variants using whole genome sequencing data. It combines two techniques that are based either on the detection of paired-end mapping abnormalities or on the detection of the depth of coverage. SV-BAY relies on a probabilistic Bayesian approach and includes a modelization of possible sequencing errors, read mappability profile along the genome and changes in the GC-content. Keeping only somatic SVs is an additional option when matched normal control data are provided. SV-BAY compares favorably with existing tools on simulated and experimental data sets [6] Software SV-BAY is freely available https://github.com/InstitutCurie/SV-Bay.

As a side product, a novel exhaustive catalogue of SV types -to date the most comprehensive SV classification- was built. On the grounds of previous publications and experimental data, seven new SV types, ignored by the existing SV calling algorithms, were exhibited.

We also contributed, in collaboration with Céline Scornavacca's group (ISEM, Montpellier) to the algorithmic foundations of the `EcceTERA` software [7] for the reconciliation of gene and species phylogenetic trees. This software adopts a maximum parsimony approach to predict in an evolutionary model that includes duplications, losses and transfers of genes.

BEAGLE Project-Team

6. New Results

6.1. Open-Ended Novelty: Requirements, Guidelines, and Challenges

Participants: G. Beslon

We started in 2014 a collective reflexion on the concept of "Open-Endedness". This reflexion led to a collective paper published this year in "Theory in Biosciences" [12]. The open-endedness of a system is often defined as a continual production of novelty. In this paper we pin down this concept more fully by defining several types of novelty that a system may exhibit, classified as variation, innovation, and emergence. We then provide a meta-model for including levels of structure in a system's model. From there, we define an architecture suitable for building simulations of open-ended novelty-generating systems and discuss how previously proposed systems fit into this framework. We discuss the design principles applicable to those systems and close with some challenges for the community.

6.2. Endocannabinoid dynamics gate spike timing dependent depression and potentiation

Participants: I. Prokin and H. Berry, in collaboration with L. Venance lab, CIRB, Collège de France, Paris.

Learning and memory depend on processes that alter the connections – or synapses – between neurons in the brain. For example, molecules called endocannabinoids can alter synapses to decrease the influence that one neuron has on another neuron's activity. This "synaptic depression" is an important mechanism through which the brain can adapt to an experience. However, recent research also suggests that endocannabinoids might also increase the influence one neuron has on another neuron's activity by strengthening the synaptic connection between neurons. This opposite process is known as synaptic potentiation, and is also important for learning from experience. But how do endocannabinoids manage to produce opposing effects? Using a combination of electrophysiological recording experiments from our experimental collaborator lab and mathematical modeling, we have deciphered the molecular mechanisms that govern the action of endocannabinoids at key synapses in rat brain slices. This revealed that both the levels and timing of endocannabinoid release control changes in the strength of the synaptic connections. Electrical stimulations that produced moderate amounts of endocannabinoids over a prolonged period led to synaptic depression. However, stimulation that produced short but large endocannabinoid peaks caused synaptic potentiation. The enzymes that control endocannabinoid levels thus play a crucial role in determining whether a given stimulation leads to the strengthening or weakening of a synaptic connection. In the type of synapses studied, changes to synaptic strength also depend on another chemical called dopamine. Abnormal dopamine production is implicated in a number of disorders, including Parkinson's disease and addiction. These results have been published in eLife [16].

6.3. Quantitative convergence towards a self similar profile in an age-structured renewal equation for subdiffusion

Participants: A. Mateos Gonzalez and H. Berry, in collaboration with T. Lepoutre, EPI Dracula, Inria.

Continuous-time random walks are generalisations of random walks frequently used to account for the consistent observations that many molecules in living cells undergo anomalous diffusion, i.e. subdiffusion. We described the subdiffusive continuous-time random walk using age-structured partial differential equations with age renewal upon each walker jump, where the age of a walker is the time elapsed since its last jump. In the spatially-homogeneous (zero-dimensional) case, we followed the evolution in time of the age distribution. An approach inspired by relative entropy techniques allows us to obtain quantitative explicit rates for the convergence of the age distribution to a self-similar profile, which corresponds to convergence to a stationnary profile for the rescaled variables. An important difficulty arises from the fact that the equation in self-similar variables is not autonomous and we do not have a specific analyitcal solution. Therefore, in order to quantify the latter convergence, we estimate attraction to a time-dependent "pseudo-equilibrium", which in turn converges to the stationnary profile. These results have been published in Acta Applicandae Mathematicae [13].

6.4. Modulation of Synaptic Plasticity by Glutamatergic Gliotransmission

Participants: M. De Pittà in collaboration with N. Brunel, Dept of Neuroscience and Statistics, University of Chicago, USA.

Glutamatergic gliotransmission, that is the release of glutamate from perisynaptic astrocyte processes in an activity-dependent manner, has emerged as a potentially crucial signal-ing pathway for regulation of synaptic plasticity, yet its modes of expression and function in vivo remain unclear. We focused on two experimentally well-identified gliotransmitter patwhays: (i) modulations of synaptic release and (ii) postynaptic slow inward currents mediated by glutamate released from astrocytes, and investigate their possible functional relevance on synaptic plasticity in a biophysical model of an astrocyte-regulated synapse. Our model predicts that both pathways could profoundly affect both short-and long-term plasticity. In particular, activity-dependent glutamate release from astrocytes, could dramatically change spike-timing–dependent plasticity, turning potentiation into depression (and vice versa) for the same protocol. These results have been published in Neural plasticity [17] and in a review targetting a biologist audience in the journal Neuroscience [18].

6.5. Comparative Genomics and artificial life

Participants: P Biller, C Knibbe, G Beslon, E Tannier

Molecular evolutionary methods and tools are difficult to validate as we have almost no direct access to ancient molecules. Inference methods may be tested with simulated data, producing full scenarios they can be compared with. But often simulations design is concomitant with the design of a particular method, developed by a same team, based on the same assumptions, when both should be blind to each other. In silico experimental evolution consists in evolving digital organisms with the aim of testing or discovering complex evolutionary processes. Models were not designed with a particular inference method in mind, only with basic biological principles. As such they provide a unique opportunity to blind test the behavior of inference methods. We give a proof of this concept on a comparative genomics problem: inferring the number of inversions separating two genomes. We use Aevol, an in silico experimental evolution platform, to produce benchmarks, and show that most combinatorial or statistical estimators of the number of inversions fail on this dataset while they were behaving perfectly on ad-hoc simulations. We argue that biological data is probably closer to the difficult situation.

This work has been published in the article [23] and presented at the Jobim conference [25] and provided the inspiration for a new estimator of the evolutionary distance between two genomes (see below).

6.6. Breaking good

Participants: P Biller, C Knibbe, E Tannier, in collaboration with L Guéguen, University of Lyon 1.

Models of evolution by genome rearrangements are prone to two types of flaws: one is to ignore the diversity of susceptibility to breakage across genomic regions, the other is to suppose that susceptibility values are given. Without necessarily supposing their precise localization, we call "solid" the regions that are improbably broken by rearrangements and "fragile" the regions outside solid ones. We propose a model of evolution by inversions where breakage probabilities vary across fragile regions and over time. It contains as a particular case the uniform breakage model on the nucleotidic sequence, where breakage probabilities are proportional to fragile region lengths. This is very different from the frequently used pseudo-uniform model where all fragile regions have the same probability to break. Estimations of rearrangement distances based on the pseudo-uniform model completely fail on simulations with the truly uniform model. On pairs of amniote genomes, we show that identifying coding genes with solid regions yields incoherent distance estimations, especially with the pseudo-uniform model, and to a lesser extent with the truly uniform model. This incoherence is solved when we co-estimate the number of fragile regions with the rearrangement distance. The estimated number of fragile regions is surprisingly small, suggesting that a minority of regions are recurrently used by rearrangements. Estimations for several pairs of genomes at different divergence times are in agreement with a slowly evolvable co-localization of active genomic regions in the cell.

This work has been published in an article for a reference biology journal [14].

6.7. Subspace clustering

Participants: S Peignier, C Rigotti

We developed an algorithm to tackle the subspace clustering problem over a data stream containing clusters than change over time. Very few subspace clustering algorithms can handle such streams. Our starting point was the work made in the team on evolution of evolution mechanisms and on a preliminary bio-inspired algorithm that we have proposed last year. This previous algorithm included many bio-like features like variable genome length and organization, functional and non-functional elements, and variation operators including chromosomal rearrangements. It achieved satisfying results on standard benchmark data sets but was not designed to process dynamic streams. The new algorithm finds and adapts changing clusters over such streams, while preserving high cluster quality. It has been successfully used to build the evolving music generation system EvoMove.

BIGS Project-Team

7. New Results

7.1. Stochastic modeling

7.1.1. Spatial and spatio-temporal modeling

Participants: A. Gégout-Petit
External collaborators: Y. Cao, S. Li, L. Guerin-Dubrana (Inra Bordeaux)

In the framework of a collaboration with INRA Bordeaux about the esca-illness of vines, Anne Gégout-Petit with Shuxian Li 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 test based on the join count statistics, a paper is accepted for publication [5]. We also developped a two-step centered autologistic model for the study of the dynamic of propagation. This work has been presented as invited paper in [20] and is in preparation for publication.

7.1.2. Modelisation of response to chemotherapy in gliomas

Participant: S. Wantz-Mézières
External collaborator: J.-M. Moureaux, Y. Gaudeau, M. Ben Abdallah, M. Ouqamra (CRAN, Université de Lorraine), L. Taillandier, M. Blonski (CHU Nancy)
The collaboration with neurologists (CHU Nancy) and automaticians (CRAN) has carried on this year and led to the PhD presentation of M. Ben Abdallah, on December 12, 2016 [17], [16]. We completed the modeling approach by a data analysis one. In the framework of a master 2 project, supervised and non supervised methods have confirmed our results on our local data base. This encourages us to continue our work in extending the data base via a collaboration with Montpellier CHU. Our perspectives are to validate multi-factor models, including biological and anatomopathological factors, and to design a decision-aid tool for praticians.

7.1.3. Time-changed extremal process as a random sup measure

Participant: Céline Lacaux
External collaborator: Gennady Samorodnistky

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

7.1.4. Fast and Exact synthesis of some operator scaling Gaussian random fields

Participant: Céline Lacaux
External collaborator: Hermine Biermé

Operator scaling Gaussian random fields, as anisotropic generalizations of self-similar fields, know an increasing interest in the literature. Up to now, such models were only defined through stochastic integrals, without knowing explicitly their covariance functions. In link with this misunderstanding, one of the drawbacks is that no exact method of simulation has been proposed. In view to fill this lack, Hermine Biermé and Céline Lacaux have recently exhibit explicit covariance functions, as anisotropic generalizations of fractional Brownian fields ones. This allows them to propose a fast and exact method to synthetise an operator scaling Gaussian random fields with such a covariance function. Their algorithm is based on the famous circulant embedding matrix method. This is a first piece of work to popularized operator scaling Gaussian random field in anisotropic spatial data modeling.

7.1.5. DNA sequences analysis

Participants: P. Vallois
External collaborators: A. Lagnoux and S. Mercier (Toulouse)
In an article accepted at Bioinformatics, the goal is to illustrate different results on the local score distribution assuming an i.i.d. model, especially the one based on the pair (local score,length) and the one on the local score position. We measure with statistical tests how different approximations of the local score distribution fit to simulated sequences. In particular, our simulations show that the popular Karlin & Altschul approximation is not accurate in a wide range of situations. We add to the local score the length of the segment that realises it and we study the induced changes with numerical simulations. We also study specificity and sensitivity for the different methods. We introduce a new one dimensional statistic which is a function of H_n^* and L_n^* and we test its distribution. Finally, we estimate the probability that $H_n^* = H_n$ in different settings.

7.2. Estimation and control for stochastic processes

7.2.1. Piecewise-deterministic Markov processes

Participants: Romain Azaïs, Florian Bouguet, Anne Gégout-Petit, Florine Greciet, Aurélie Muller-Gueudin
External participants: Michel Benaïm (Université de Neuchâtel), Bertrand Cloez (Inra-SupAgro MISTEA), Alexandre Genadot (Inria CQFD, Université de Bordeaux)
A piecewise-deterministic Markov process is a stochastic process whose behavior is governed by an ordinary differential equation punctuated by random jumps occurring at random times. This class of stochastic processes offers a wide range of applications, especially in biology (kinetic diatery exposure model and growth of bacteria for example). BIGS' members mainly work on statistical inference techniques for these stochastic processes [2], [29], which is an essential step to build relevant application models. We also investigate the probabilistic properties of these processes [32], [31] as well as the application in reliability to crack growth in some alloy in the industrial context of the PhD thesis of Florine Greciet with SAFRAN Aircraft Engines [33].
In a preprint recently accepted for publication in Electronic Journal of Statistics [2], we focus on the nonparametric estimation problem of the jump rate for piecewise-deterministic Markov processes observed within a long time interval under an ergodicity condition. More precisely, we introduce an uncountable class (indexed by the deterministic flow) of recursive kernel estimates of the jump rate and we establish their strong pointwise consistency as well as their asymptotic normality. In addition, we propose to choose among this class the estimator with the minimal variance, which is unfortunately unknown and thus remains to be estimated. We also discuss the choice of the bandwidth parameters by cross-validation methods. In [29], we state a new characterization of the jump rate when the transition kernel only charges a discrete subset of the state space. We deduce from this result a competitive nonparametric technique for estimating this feature of interest. We state the uniform convergence in probability of the estimator. Both the methodologies have been illustrated on numerical examples and real data.
The article [32] deals with a class of conservative growth-fragmentation equations with a deterministic viewpoint. With the help of Foster-Lyapunov criteria, we study the long-time behavior of some associated piecewise-deterministic Markov process, which represents a typical individual following the dynamics of the equation. If the growth and the fragmentation are balanced, it is possible to provide existence and unicity for the stationary distribution on the process, as well as precise bounds for its tails of distributions in the neighborhoods of both 0 and $+\infty$. Our probabilistic results are systematically compared to estimates already obtained with deterministic methods.

In [31], we are interested by the long-time behavior of inhomogeneous-time Markov chains. We put forward an original and unified approach to relate some of their asymptotic properties (stationary distribution, speed of convergence, ...) to the ones of an auxiliary homogeneous-time Markov process. Such results are close to traditional functional limit theorems, but our method differs from the standard "Tightness/Identification" argument; it is based on the notion of asymptotic pseudotrajectories on the space of probability measures. We recover classical results, such as normalized bandit algorithms converging to a piecewise-deterministic Markov process, or weighted random walks or decreasing step Euler schemes approximated with solutions of stochastic differential equations.

7.2.2. Statistics of Markov chains

Participant: Romain Azaïs

External participants: Bernard Delyon (Université Rennes 1), François Portier (Télécom ParisTech)

Suppose that a mobile sensor describes a Markovian trajectory in the ambient space. At each time the sensor measures an attribute of interest, e.g., the temperature. Using only the location history of the sensor and the associated measurements, the aim of the paper [27] is to estimate the average value of the attribute over the space. In contrast to classical probabilistic integration methods, e.g., Monte Carlo, the proposed approach does not require any knowledge on the distribution of the sensor trajectory. Probabilistic bounds on the convergence rates of the estimator are established. These rates are better than the traditional "root n"-rate, where n is the sample size, attached to other probabilistic integration methods. For finite sample sizes, the good behaviour of the procedure is demonstrated through simulations and an application to the evaluation of the average temperature of oceans is considered.

7.2.3. Realtime Tracking of the Photobleaching Trajectory during Photodynamic Therapy

Participant: T. Bastogne

Photodynamic therapy (PDT) is an alternative treatment for cancer that involves the administration of a photosensitizing agent, which is activated by light at a specific wavelength. This illumination causes after a sequence of photoreactions, the production of reactive oxygen species responsible for the death of the tumor cells but also the degradation of the photosensitizing agent, which then loose the fluorescence properties. The phenomenon is commonly known as photobleaching process and can be considered as a therapy efficiency indicator. In [8], we present the design and validation of a real time controller able to track a preset photobleaching trajectory by modulating the light impulses width during the treatment sessions. This innovative solution was validated by in vivo experiments that have shown a significantly improvement of reproducibility of the inter-individual photobleaching kinetic. We believe that this approach could lead to personalized photodynamic therapy modalities in the near future.

7.2.4. Stochastic simulation and design of numerical experiments for the prediction of nanoparticles/X-ray interactions in radiotherapy.

Participant: T. Bastogne

The increase of computational environments dedicated to the simulation of nanoparticles (NP)-X-Rays interactions has opened new perspectives in computer-aided-design of nanostructured materials for biomedical applications. Several published studies have shown a crucial need of standardization of these numerical simulations [92]. That is why, we proposed to perform a robustness multivariate analysis in [8]. A gold nanoparticle (GNP) of 100 nm diameter was selected as a standard nano-system activated by a X-ray source placed just below the NP. Two response variables were examined: the dose enhancement in seven different spatial regions of interest around the NP and the duration of the experiments. 9 factors were pre-identified as potentially critical. A Plackett-Burman design of numerical experiments was applied to estimate and test the effects of each simulation factors on the examined responses. Four factors: the working volume, the spatial resolution, the spatial cutoff and the computational mode (parallelization) do not significantly affect the dose deposition results and none except the last one may reduce the computational duration. The energy cutoff may cause significant variations of the dose enhancement in some specific regions of interest: the higher the cutoff, the closer the secondary particles will stop from the GNP. By contrast, the Auger effect as well as the choice of the physical medium and the fluence level clearly appear as critical simulation parameters. Consequently, these

four factors may be compulsory examined before comparing and interpreting any simulation results coming from different simulation sessions.

In [9], we address the prediction issue of organometallic nanoparticles (NPs)-based radiosensitization enhancement. The goal was to carry out computational experiments to quickly identify efficient nanostructures and then to preferentially select the most promising ones for the subsequent in vivo studies. To this aim, this interdisciplinary article introduces a new theoretical Monte Carlo computational ranking method and tests it using 3 different organometallic NPs in terms of size and composition. While the ranking predicted in a classical theoretical scenario did not fit the reference results at all, in contrast, we showed for the first time how our accelerated in silico virtual screening method, based on basic in vitro experimental data (which takes into account the NPs cell biodistribution), was able to predict a relevant ranking in accordance with in vitro clonogenic efficiency. This corroborates the pertinence of such a prior ranking method that could speed up the preclinical development of NPs in radiation therapy.

This in-silico approach was tested in [25] to screen radiosensitizing nanoparticles and the results have been validated by in vitro assays.

7.2.5. Complexity analysis of Policy Iteration

Participant: Bruno Scherrer

Given a Markov Decision Process (MDP) with n states and a total number m of actions, we study in [10] the number of iterations needed by Policy Iteration (PI) algorithms to converge to the optimal γ-discounted policy. We consider two variations of PI: Howard's PI that changes the actions in all states with a positive advantage, and Simplex-PI that only changes the action in the state with maximal advantage. We show that Howard's PI terminates after at most $O\left(\frac{m}{1-\gamma}\log\left(\frac{1}{1-\gamma}\right)\right)$ iterations, improving by a factor $O(\log n)$ a result by Hansen et al., while Simplex-PI terminates after at most $O\left(\frac{nm}{1-\gamma}\log\left(\frac{1}{1-\gamma}\right)\right)$ iterations, improving by a factor $O(\log n)$ a result by Ye. Under some structural properties of the MDP, we then consider bounds that are independent of the discount factor γ: quantities of interest are bounds τ_t and τ_r—uniform on all states and policies—respectively on the *expected time spent in transient states* and *the inverse of the frequency of visits in recurrent states* given that the process starts from the uniform distribution. Indeed, we show that Simplex-PI terminates after at most $\widetilde{O}\left(n^3m^2\tau_t\tau_r\right)$ iterations. This extends a recent result for deterministic MDPs by Post & Ye, in which $\tau_t \leq 1$ and $\tau_r \leq n$; in particular it shows that Simplex-PI is strongly polynomial for a much larger class of MDPs. We explain why similar results seem hard to derive for Howard's PI. Finally, under the additional (restrictive) assumption that the state space is partitioned in two sets, respectively states that are transient and recurrent for all policies, we show that both Howard's PI and Simplex-PI terminate after at most $\widetilde{O}(m(n^2\tau_t + n\tau_r))$ iterations.

7.2.6. Approximate Dynamic Programming for Markov Games

Participant: Bruno Scherrer

We have made two contributions to the analysis of Approximate Dynamic Programming algorithms for Markov Games.

First, we extend in [21] several non-stationary Reinforcement Learning (RL) algorithms and their theoretical guarantees to the case of discounted zero-sum Markov Games (MGs). As in the case of Markov Decision Processes (MDPs), non-stationary algorithms are shown to exhibit better performance bounds compared to their stationary counterparts. The obtained bounds are generically composed of three terms: 1) a dependency over gamma (discount factor), 2) a concentrability coefficient and 3) a propagation error term. This error, depending on the algorithm, can be caused by a regression step, a policy evaluation step or a best-response evaluation step. As a second contribution, we empirically demonstrate, on generic MGs (called Garnets), that non-stationary algorithms outperform their stationary counterparts. In addition, it is shown that their performance mostly depends on the nature of the propagation error. Indeed, algorithms where the error is due to the evaluation of a best-response are penalized (even if they exhibit better concentrability coefficients and dependencies on gamma) compared to those suffering from a regression error.

Furthermore, we report in [22] theoretical and empirical investigations on the use of quasi-Newton methods to minimize the Optimal Bellman Residual (OBR) of zero-sum two-player Markov Games. First, it reveals that state-of-the-art algorithms can be derived by the direct application of Newton's method to different norms of the OBR. More precisely, when applied to the norm of the OBR, Newton's method results in the Bellman Residual Minimization Policy Iteration (BRMPI) and, when applied to the norm of the Projected OBR (POBR), it results into the standard Least Squares Policy Iteration (LSPI) algorithm. Consequently, new algorithms are proposed, making use of quasi-Newton methods to minimize the OBR and the POBR so as to take benefit of enhanced empirical performances at low cost. Indeed, using a quasi-Newton method approach introduces slight modifications in term of coding of LSPI and BRMPI but improves significantly both the stability and the performance of those algorithms. These phenomena are illustrated on an experiment conducted on artificially constructed games called Garnets.

7.3. Algorithms and estimation for graph data

7.3.1. Modelisation of networks of multiagent systems

Participants: Aurélie Muller-Gueudin
We relate here a collaboration with researchers in Automatic in Nancy (CRAN).

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

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

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

We have an accepted paper in the journal IEEE Transactions on Automatic Control [6].

7.3.2. Compression and analysis of trees

Participant: Romain Azaïs
External participants: Jean-Baptiste Durand (ENSIMAG, Inria MISTIS), Christophe Godin (Inria Virtual Plants), Benoît Henry (Inria TOSCA puis Madynes), Alexandre Genadot (Université de Bordeaux, Inria CQFD)
Tree-structured data naturally appear in various fields, particularly in biology where plants and blood vessels may be described by trees, but also in computer science because XML documents form a tree structure. Among trees, the class of self-nested trees presents remarkable compression properties because of the systematic repetition of subtrees in their structure. In a recent preprint [28], we provide a better combinatorial characterization of this specific family of trees. We show that self-nested trees may be considered as a good approximation class of unordered trees. In addition, we compare our approximation algorithms with a competitive approach of the literature on a simulated dataset. On the other hand, the paper [30] 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. An application to the analysis of revisions of Wikipedia articles is also considered through real data.

7.4. Regression and machine learning

7.4.1. Aggregated methods for covariates selection in high-dimensional data under dependence

Participants: A. Gégout-Petit, A. Muller-Gueudin, Y. Shi
External collaborators: B. Bastien (Transgene, Strasbourg)

In the purpose to select 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 the lung cancer. A paper on the subject is in preparation.

7.4.2. Clustering of the values of a response variable and simultaneous covariate selection using a stepwise algorithm

Participant: J.-M. Monnez
External collaborator: O. Collignon (LIH Luxembourg)
In supervised learning the number of values of a response variable to predict can be very high. Grouping these values in a few clusters can be useful to perform accurate supervised classification analyses. On the other hand selecting relevant covariates is a crucial step to build robust and efficient prediction models. We propose in this paper an algorithm that simultaneously groups the values of a response variable into a limited number of clusters and selects stepwise the best covariates that discriminate this clustering. These objectives are achieved by alternate optimization of a user-defined selection criterion. This process extends a former version of the algorithm to a more general framework. Moreover possible further developments are discussed in detail [3].

7.4.3. Death or hospitalization scoring for heart failure patients

Participant: J.-M. Monnez, K. Duarte
External collaborator: E. Albuisson (CHU, Nancy)
The purpose of this study was to define a short term event (death or hospitalization) score for heart failure patients based on the observation of biological, clinical and medical historical variables. Some of them were transformed or winsorized. Two methods of statistical learning were performed, logistic regression and linear discriminant analysis, different variable selection methods were used, on bootstrap samples. Aggregation of classifiers and out-of-bag validation were used. Finally a score taking values between 0 and 100 was established and an odds-ratio was defined in order to support medical decision (writing in progress).

7.4.4. Sequential linear regression with online standardized data

Participant: J.-M. Monnez, K. Duarte
External collaborator: E. Albuisson
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 (paper to be submitted).

7.4.5. Mixed-effects ARX Model Identification of Dynamical Biological Systems

Participants: T. Bastogne, L. Batista

System identification is a data-driven modeling approach more and more used in biology and biomedicine [26]. In this application context, each assay is always repeated to estimate the response variability. The inference of the modeling conclusions to the whole population requires to account for the inter-individual variability within the modeling procedure. One solution consists in using mixed effects models but up to now no similar approach exists in the field of dynamical system identification. In [23], we propose a new solution based on an ARX (Auto Regressive model with eXternal inputs) structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameters. Simulations show the relevance of this solution compared with a classical procedure of system identification repeated for each subject.

In [24], we propose a solution to firstly estimate the Fisher information matrix using the Louis' method and secondly to determine the parameters confidence intervals of an ARX model structure. We show relevance of the proposed solution in simulation and using real in-vitro data coming from realtime cell impedance measurements.

In parallel, we applied the mixed-effect modeling approach to the analysis in vivo responses in order to identify pronostic biomarkers of tumor regrowth after photodynamic therapy [11]. This application corroborated the practical relevance of our model-based approach.

7.4.6. Uniform asymptotic certainty bands for the conditional cumulative distribution function

Participants: S.Ferrigno, A. Muller-Gueudin, M. Maumy-Bertrand (IRMA, Strasbourg)

In this work, we study the conditional cumulative distribution function and a nonparametric estimator associated to this function. The conditional cumulative distribution function has the advantages of completely characterizing the law of the random considered variable, allowing to obtain the regression function, the density function, the moments and the conditional quantile function. As a nonparametric estimator of this function, we focus on local polynomial techniques described in Fan and Gijbels [ref]. In particular, we use the local linear estimation of the conditional cumulative distribution function.

The objective of this work is to establish uniform asymptotic certainty bands for the conditional cumulative distribution function. To this aim, we give exact rate of strong uniform consistency for the local linear estimator of this function (writing in progress).

7.4.7. Omnibus tests for regression models

Participants: R.Azaïs, S.Ferrigno, M-J Martinez Marcoux (LJK, Grenoble)

The aim of this collaboration begins is to compare, through simulations, several methods to test the validity of a regression model. These tests can be "directional" in that they are designed to detect departures from mainly one given assumption of the model (for example the regression function, the variance or the error) or global (for example the conditional distribution function). The establishment of such statistical tests require the use of nonparametric estimators various functions (regression, variance, cumulative distribution function). The idea would then be able to build a tool (package R) that allows a user to test the validity of the model it uses through different methods and varying parameters associated with modeling. This work is currently in progress.

BONSAI Project-Team

7. New Results

7.1. Approximate pattern matching

The problem of measuring the similarity between two strings arises in many areas of sequence analysis. A common metric for it is the *Levenshtein distance*. This distance is defined as the smallest number of substitutions, insertions, and deletions of symbols required to transform one of the words into the other. We have investigated the basic problem of the size of the neighborhood of a given pattern P: count how many strings are within a bounded distance of a fixed reference string. There has been no efficient algorithm for calculating it so far. We have proposed a dynamic programming algorithm that scales linearly with the size of the pattern P. For that, we have introduced a new variant of the universal Levenshtein automaton, that is interesting by itself and that can have many other applications in text algorithms [31].

We have also addressed the related problem of approximate pattern matching: Given a text T and a pattern P, find all locations in T that differ by at most k errors (in the sense of the Levenshtein distance) from P. We have proposed a new kind of seeds (the 01^*0 seeds) that combines exact parts and parts with a fixed number of errors, and that are specifically well-suited for short DNA motifs with high error-rate. We have demonstrated the applicability of those seeds on two main case studies : pattern matching on a genomic scale with a Burrows-Wheeler transform, and multi-pattern matching with indexation of the set of patterns [30].

7.2. Parallel algorithm for de Bruijn graph compaction

Constructing a *de Bruijn graph* is an important step in the analysis of NGS data. This data structure is used in several applications, such as *de novo* assembly, variant detection, and transcriptome quantification. However, the representation of this graph often consumes prohibitive amounts of memory for large datasets. An operation, called compaction, enables to represent the graph more efficiently. However, so far, there was no algorithm for compacting the graph quickly and in low memory.

Along with colleagues at Inria Rennes and at Penn State University, we introduced a parallel algorithm and an implementation, BCALM 2, for constructing directly a compacted de Bruijn graph given a set of reads. Our results show that this algorithm enables to construct the graph for very large datasets, such as the spruce and pine genomes, in reasonable time and memory on a single machine. This represents a performance improvement of two orders of magnitude compared to previously available methods. BCALM 2 is open-source and was published at ISMB 2016 [20].

7.3. Range minimum query

The *range minimum query* problem consists in finding the minimum value inside any queried range of a preprocessed integer sequence. Several methods exist to compute the minimum in constant time, using almost the theoretical minimal amount of space. Those methods consist in splitting the problem in several subproblems and precomputing the solutions for them.

With Alice Héliou (AMIB Inria team, Saclay), Martine Léonard and Laurent Mouchard (LITIS, Rouen), we designed a new method, which is worse in terms of time complexity [24]. Our solution relies on a totally different concept as previous ones: We only store the values that are local minima. This approach is therefore simple and can, on specific inputs, require much less memory than the general theoretical minimal bound. Moreover the simplicity of the method can be easily adapted to allow updates in the original integer sequence.

7.4. Coding isoform structures

Our researches on gene isoform structures started in 2014 with the CG-Alcode Associated Team and in collaboration with Anne Bergeron from the LACIM (Montréal, Canada). We aimed at defining better definitions of isoform orthology at the coding level, which are based on the preservation of all the exon junctions in two orthologous isoforms. This estimation is achieved at the gene level, where sequence homology is detected for both exons and their flanking intronic splice sites [19]. The approach largely outperforms competing programs in terms of precision and recall. Using the successive releases of the ten years old CCDS database, we show that the discovery rate of orthologous isoforms between human and mouse is growing continuously and that it displays no sign of completion.

7.5. Nonribosomal peptides

We were invited to contribute in a volume of "Methods in Molecular Biology" by authoring a chapter focusing on NRPS biosynthesis. This chapter [32] was about the use of the Norine platform (developed by the team) and other bioinformatics tools for the analysis of nonribosomal peptide synthetases and their products. We invited our collaborator from Denmark, Tilmann Weber, to complete this chapter with the introduction of his tool, antiSMASH.

We annotated 48 genomes of *Burkholderia* species using our annotation protocol, that starts from a genome sequence and goes to the predicted nonribosomal peptides. We have predicted 161 gene clusters producing nonribosomal peptides, leading to the synthesis of not only already known peptides, but also new ones [22] with potential applications in biocontrol.

A new version of the Norine interface is now available. The form to query the annotations is now flexible and dynamic. The user can build his own query to search for annotations in several fields combined by boolean operators. Moreover, the database structure has been modified to allow, among others, a hierarchical representation of the NRPS taxonomy. Finally, the MyNorine tool has been enhanced and updated to take into account these changes and the description page of the peptides has been reorganized.

7.6. 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 [21].

In 2016, with our colleagues at Lille hospital, we published two articles in haematological journals to detail our method for the diagnosis [23] and for the follow-up [28] of the acute lymphoblastic leukemia using high-throughput sequencing. Our results also show what those new techniques, together with bioinformatics software, bring in a routine practice. Being a full platform with metadata storage, Vidjil is used on a regular basis by about 20 laboratories around the world. In France, the majority of diagnosis samples from acute lymphoblastic leukemia patients are now analyzed using Vidjil.

7.7. Assembly of the giraffe genome and the gorilla Y-chromosome

We collaborated with two labs from the Pennsylvania State Institute (Cavener Lab and Makova Lab) for practical analysis of DNA sequencing data. The first collaboration led to the publication of the giraffe genome in Nature Communication [18]. In this article our contribution was to provide the first draft-quality whole-genome sequences of the giraffe and the okapi. The second collaboration was about assembling the Y-chromosome of the gorilla using a novel sequencing strategy as well as novel computational tools. This work was published in Genome Research [29].

CAPSID Project-Team

7. New Results

7.1. Correlating Adverse Drug Side Effects

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 Prof. Michel Dumontier of the 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 [26], [27]. A paper describing this work has been submitted to the Journal of Biomedical Semantics.

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 [19]. An article describing the results obtained when using Sam to dock several symmetrical protein complexes from the "CAPRI" docking experiment has also been published [13].

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. We have recently extended our "Kpax" protein structure alignment algorithm to flexibly align pairs of structures that cannot be completely superposed by a single rigid-body transformation, and to calculate multiple alignments of several similar structures flexibly. A journal article describing the approach has been published [20].

7.4. Annotating 3D Protein Domains

Many protein chains in the Protein Data Bank (PDB) are cross-referenced with EC numbers and Pfam domains. However, these annotations do not explicitly indicate any relation between EC numbers and Pfam domains. In order to address this limitation, we developed EC-DomainMiner, a recommender-based approach for associating EC (Enzyme Commission) numbers with Pfam domains [29]. EC-DomainMiner is able to infer automatically 20,179 associations between EC numbers and Pfam domains from existing EC-chain/Pfam-chain associations from the SIFTS database as well as EC-sequence/Pfam-sequence associations from UniProt databases. A manuscript describing this work has been provisionally accepted by the journal *BMC-Bioinformatics.*

7.5. Identifying New Anti-Fungal Agents

In this collaboration with several Brasilian laboratories (at University of Mato Grosso State, University of Maringá, Embrapa, and University of Brasilia), we identified several novel small-molecule drug leads against *Trypanosoma cruzi*, a parasite responsible for Chagas disease [21]. We also proposed several small-molecule inhibitors against *Fusarium graminearum*, a fungal threat to global wheat production [15], [12].

DYLISS Project-Team

7. New Results

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

Participants: Meziane Aite, Marie Chevallier, Olivier Dameron, Aurélie Evrard, Clémence Frioux, Xavier Garnier, Jeanne Got, François Moreews, Yann Rivault, Anne Siegel, Pierre Vignet, Denis Tagu, Camille Trottier.

Integration and query of biological datasets with Semantic Web technologies. The purpose of this work is to obtain quick answers to biological questions demanding currently hours of manual search in several spreadsheet results files. We introduce an integration and interogation framework using an RDF model and the SPARQL query language. It allows biologists to transparently integrate and query their data without any a priori proficiency about RDF and SPARQL. [*O. Dameron, A. Evrard, X. Garnier*] [37], [45]

Handling the heterogeneity of genomic and metabolic networks data within flexible workflows with the PADMet toolbox A main challenge of the era of fast and massive genome sequencing is to transform sequences into biological knowledge. The high diversity of input files and tools required to run any metabolic networks reconstruction protocol represents an important drawback: it appears very difficult to ensure that input files agree among them. Such a heterogeneity produces loss of information during the use of the protocols and generates uncertainty in the final metabolic model. Here we introduce the PADMet-toolbox which allows conciliating genomic and metabolic network information. The toolbox centralizes all this information in a new graph-based format: PADMet (PortAble Database for Metabolism) and provides methods to import, update and export information. For the sake of illustration, the toolbox was used to create a workflow, named AuReMe, aiming to produce high-quality genome-scale metabolic networks and eventually input files to feed most platforms involved in metabolic network analyses. We applied this approach to two exotic organisms and our results evidenced the need of combining approaches and reconciling information to obtain a functional metabolic network to produce biomass. [*M. Chevallier, M. Aite, C. Frioux, J. Got, A. Siegel, C. Trottier, P. Vignet*] [34]

PEPS: a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases We showed that Semantic Web technologies are technically adapted for representing patients' data from medico-administrative databases as RDF and querying them using SPARQL. We also demonstrated that this approach is relevant as it supports the combination of patients' data with hierarchical knowledge in order to address the problem of reconciling precise patients data with more general query criteria. [*O. Dameron, Y. Rivault*] [33], [31], [30]

Telemedicine : ontology-based reasoning and data integration 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. The methods may be extended to other types of connected devices. We also worked on a telemedicine application for monitoring patients with chronic diseases. We proposed an architecture supporting data exchange in the context of multiple chronic diseases [*O. Dameron*] [26], [25], [18]

7.2. Data and knowledge integration based on combinatorial optimization

Participants: Marie Chevallier, Damien Eveillard, Jeanne Got, Julie Laniau, François Moreews, Jacques Nicolas, Anne Siegel.

Packing graphs with ASP for landscape simulation This study is part of a more general research track on graph compression, a fundamental issue for the analysis of biological networks that we address with Answer Set Programming (ASP)modelling. The general issue is to cover a given graph by a set of subgraphs. The IJCAI paper describes an application to crop allocation for generating realistic landscapes. The aim is to cover optimally a bare landscape, represented by its plot graph, with spatial patterns describing local arrangements of crops. This problem belongs to the hard class of graph packing problems. The approach provides a compact solution to the basic problem and at the same time allows extensions such as a flexible integration of expert knowledge. Particular attention is paid to the treatment of symmetries, especially due to sub-graph isomorphism issues. Experiments were conducted on a database of simulated and real landscapes. Currently, our program can process graphs of medium size, a size that enables studies on real agricultural practices. [*J. Nicolas*] [29]

Deciphering transcriptional regulations coordinating the response to environmental changes We introduce a method that extracts from a transcriptional regulatory network determined from a set of predicted transcription factors (TF) and binding site (BS) a subnetwork explaining a given set of observed co-expressions, highlighting those TFs and BSs most likely involved in the co-regulation. The method solves an optimization problem on a graph to select confident paths within the given transcriptional regulatory network joining a putative common regulator with two co-expressed genes via regulatory cascades. It provides a useful modeling scheme for deciphering the regulatory mechanisms that underly the phenotypical response of an organism to environmental challenges and can be used as a reliable tool for further research on genome scale transcriptional regulation studies. [*M. Chevallier, D. Eveillard, A. Siegel*] [13]

Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach. Our software tool *shogen* [62] was used to decipher functional roles within a consortium of five mining bacteria through the integration of genomic and metabolic knowledge at genome scale. We first reconstructed a global metabolic network. Next, using a parsimony assumption, we deciphered sets of genes, called Sets from Genome Segments (SGS), that (i) are close on their respective genomes, (ii) take an active part in metabolic pathways and (iii) whose associated metabolic reactions are also closely connected within metabolic networks. The use of SGS (shogen) pinpoints a functional compartmentalization among the investigated species and exhibits putative bacterial interactions necessary for promoting these pathways. [*M. Chevallier, D. Eveillard, A. Siegel*] [15]

Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location Our methods based on the integration of metabolic and regulatory regulations [61] were combined to statistical approaches and applied to the understanding of fatty acid metabolism in porcs and chicken. The analyses evidenced that a high-fat high-fiber diet depressed glucose and lipid anabolic molecular pathways, thus counteracting adipose tissue expansion. Interaction effects between dietary intake of fiber and lipids on gene expression may modulate innate immunity and inflammation, a response which is of interest with regard to chronic inflammation and its adverse effects on health and performance. [*F. Moreews, A. Siegel*] [20]

7.3. Systems biology

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

Representation of symbolic dynamical systems generated by a substitution. Iterated morphisms are combinatorial processes which are related to several classes of dynamical systems appearing in several fields of computer sciences and mathematics: numeration, ergodic theory, discrete geometry. They may be associated to fractal sets called "Rauzy fractals" whose topological properties are linked to the properties of the underlying dynamical system. We have introduced a generic algorithm framework to check such topological properties within a complete family of iterated morphism. This makes efficient the verification of conjectures on several families of substitutions related to multi-dimensional continued fraction algorithms. [*A. Siegel*] [14]

Identification of logical models for signaling pathways. Logical models of signaling pathways are a promising way of building effective in silico functional models of a cell. The automated learning of Boolean logic models describing signaling pathways can be achieved by training to phosphoproteomics data. In our work, combinatorial optimization methods based on recent logic programming paradigm allow to enumerate, and discriminate the family of logical models explaining data. Together, these approaches enable a robust understanding of the system response. The results are implemented in the *caspo* software. The main weakness of ASP-based learning algorithm is that they focus on the comparison of two time-points and assumes that the system has reached an early steady state. We have generalized such a learning procedure in order to discriminate Boolean networks according to their transient dynamics. To that goal, we exhibit a necessary condition that must be satisfied by a Boolean network dynamics to be consistent with a discretized time series trace. [*A. Siegel*] [23], [28]

Model of the Delayed Translation of Cyclin B Maternal mRNA After Sea Urchin Fertilization. An extended model of the numerical model introduced in [74] was developed to have a better understanding of the role of cyclin B in protein synthesis within minutes after fertilization of sea urchin eggs. The model confirms that regulation of cyclin B biosynthesis is an example of a select protein whose translation is controlled by pathways that are distinct from housekeeping proteins, even though both involve the same cap-dependent initia- tion pathway. Therefore, this model should help provide insight to the signaling utilized for the biosynthesis of cyclin B and other select proteins. [*J. Bourdon, A. Siegel*] [24]

Deciphering pathways involved in TGF-β signalling network. TGF-β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis. As a growth inhibitor of epithelial, endothelial, and hematopoietic cells, TGF-β is a potent anticancer agent in normal tissue. At the opposite TGF-β acts as a promoter of tumor by inducing the hallmarks of the cancer. Consequently targeting the deleterious effects of TGF-β without affecting its physiological role is the common goal of therapeutic strategies. While several strategies based on blocking TGF-β antibodies or small inhibitors of TGF-β receptors have been investigated, they did not take into account the impact of the (extracellular matrix) ECM remodeling that regulates TGF-β bioavailability and the complexity of TGF-β-dependent signaling pathways which regulate both physiological and pathological processes depending on context. In accordance with this, we recently demonstrated the benefical anti-tumor effect of the interplay between TGF-β signaling and the CD103 integrin pathway. At the opposite we have previously demonstrated that the disintegrin ADAMTS1 promotes TGF-β activation in chronic liver disease and we recently characterized interaction with inhibitor peptide to block such effects, using in silico approach. Importantly, we need to take into account a system-wide view and develop predictive models for therapeutic benefit. In that context we demonstrated that the ratio of TGFBR2 to TGFBR1 receptors concentrations can be used to discriminate between metastable regimes of TGF-β signaling model and predic the tumor cell aggressiveness [*N. Théret*][27], [16], [21].

7.4. Sequence and structure annotation

Participants: Aymeric Antoine-Lorquin, Catherine Belleannée, François Coste, Jacques Nicolas.

Detection of mutated primers on metagenomics sequences to detect more species. 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 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) [*A. Antoine-Lorquin, C. Belleannée*] [32], [11].

VIRALpro: a tool to identify viral capsid and tail sequences. Not only sequence data continues to outpace annotation information, but the problem is further exacerbated when organisms are underrepresented in the annotation databases. This is the case with non human-pathogenic viruses which occur frequently in metagenomic projects. Thus there is a need for tools capable of detecting and classifying viral sequences.

Based on machine learning techniques, we have proposed a new effective tool for identifying capsid and tail protein sequences, which are the cornerstones toward viral sequence annotation and viral genome classification. The software and corresponding web server are publicly available as part of the SCRATCH suite. [*F. Coste, C. Galiez*] [19]

Learning substitutable context-free grammars to model protein families. In the first experiments on learning substitutable context-free grammars to model protein families, an identified bottleneck for larger scale experimentation was parsing time. We have implemented a new parsing strategy enabling to handle efficiently the ambiguity of 'gap loops', enabling a factor 20 speedup in practice. We have also begun to investigate the inference of more expressive classes, said contextually substitutable, and have proposed a refined graph approach to learn smaller contextually substitutable grammars from smaller training samples in the framework that we have initiated with ReGLiS. [*F. Coste*] [43], [35]

How to measure the topological quality of protein grammars? To assess the quality of grammars modelling protein families, one is interested in their performances to predict new members of the families, classically measured on the basis of recall and precision in the machine learning framework, but also by their modelling power, which is more difficult to evaluate. We propose here to address this later point by measuring the consistency of grammar's parse trees with 3D structures of proteins, when they are available, by the introduction of a set of measures based on respective internal distances. [*F. Coste*] [36]

Tutorial chapter: Learning the Language of Biological Sequences. Learning the language of biological sequences is an appealing challenge for the grammatical inference research field. While some first successes have already been recorded, such as the inference of profile hidden Markov models or stochastic context-free grammars which are now part of the classical bioinformatics toolbox, it is still a source of open and nice inspirational problems for grammatical inference, enabling us to confront our ideas to real fundamental applications. As an introduction to this field, we survey here the main ideas and concepts behind the approaches developed in pattern/motif discovery and grammatical inference to characterize successfully the biological sequences with their specificities. [*F. Coste*] [40]

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2016. Some were already in preparation or submitted at the end of 2015. This will be indicated whenever it is the case.

We tried to organise the results following the five main axes of research of the team. Clearly, in some cases, a result obtained overlaps more than one axis. 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 rumour spreading (by Pierluigi Crescenzi) [9] or on general network analysis (by Pierluigi Crescenzi or Roberto Grossi) [31], [36], [40], [39], [37], [38], [10], [42] could in the future become relevant for life sciences (biology or ecology). In the other direction, algorithmic ideas that were developed in the context of a problem in life sciences could prove useful for solving more general problems (possibly with other applications). This was the case of some of the ideas explored in previous years to deal with de Bruijn graphs in the context of NGS analysis that led to the team fruitfully collaborating with a group of researchers at the ETH in Switzerland on a problem related to transport systems [34].

Below however, we preferred to only indicate the theoretical results related to problems closely resembling questions that have already been addressed by us in computational biology. Notice that such CS results 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 are not mentioned either 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

RNA-seq NGS algorithms and data analysis
SNPs (Single Nucleotide Polymorphisms) are genetic markers whose precise identification is a prerequisite for association studies. Methods to identify them are currently well developed for model species, but rely on the availability of a (good) reference genome, and therefore cannot be applied to non-model species. They are also mostly tailored for whole genome (re-)sequencing experiments, whereas in many cases, transcriptome sequencing can be used as a cheaper alternative which already enables to identify SNPs located in transcribed regions. In a paper accepted this year [18], we proposed the use of a previously developed method, KISSPLICE, that identifies, quantifies and annotates SNPs without any reference genome, using RNA-seq data only. 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 recall of our method and discussed them with respect to other methods 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. KISSPLICE can be used for any species to annotate SNPs and predict their impact on the protein sequence. We further enable to test for the association of the identified SNPs with a phenotype of interest.

We participated also in two other works, one computational and the other biological, on alternative splicing in Human.

The first is associated to the ANR Colib'read project in which we were one of the partners. A Colib'read Galaxy tools suite was developed that should enable a broad range of life science researchers to analyse raw NGS data, allows the maximum biological information to be retained in the data, and uses a very low memory footprint [17]. The algorithms implemented in the tools are based on the use of a de Bruijn graph and of a bloom filter. The analyses can be performed in a few hours, using small amounts of memory. Applications using real data further demonstrate the good accuracy of these tools compared to classical approaches.

KISSPLICE was also used in the context of myotonic dystrophy (DM), which is caused by the expression of mutant RNAs containing expanded CUG repeats that sequester muscleblind-like (MBNL) proteins, leading to alternative splicing changes. Cardiac alterations, characterised by conduction delays and arrhythmia, are the second most common cause of death in DM. Using RNA sequencing, the authors of [14] identified novel splicing alterations in DM heart samples, including a switch from adult exon 6B towards fetal exon 6A in the cardiac sodium channel, SCN5A. They found that MBNL1 regulates alternative splicing of SCN5A mRNA and that the splicing variant of SCN5A produced in DM presents a reduced excitability compared to the control adult isoform. Importantly, reproducing splicing alteration of Scn5a in mice is sufficient to promote heart arrhythmia and cardiac-conduction delay, two predominant features of myotonic dystrophy. Misregulation of the alternative splicing of SCN5A may therefore contribute to a subset of the cardiac dysfunctions observed in myotonic dystrophy.

We introduced CIDANE, a novel framework for genome-based transcript reconstruction and quantification from RNA-seq reads [8]. CIDANE assembles transcripts efficiently with significantly higher sensitivity and precision than existing tools. Its algorithmic core not only reconstructs transcripts *ab initio*, but also allows the use of the growing annotation of known splice sites, transcription start and end sites, or full-length transcripts, which are available for most model organisms. CIDANE supports the integrated analysis of RNA-seq and additional gene-boundary data and recovers splice junctions that are invisible to other methods.

Landscape of somatic mutations in breast cancer whole-genome sequences
In the context of the International Cancer Genome Consortium (ICGC), we conducted a whole-genome, exome, RNASeq and methylome characterisation of 560 breast cancers. The results were published this year in three main papers.

The first one describes the general landscape of somatic mutations and rearrangements in all subtypes of breast cancers [21]. This allowed to extend our current repertoire of probable breast cancer drivers to 93 genes. The mutational signature analysis was extended to genome rearrangements as well and revealed six typical rearrangement signatures. Three of them, characterised by tandem duplications or deletions, appear associated with defective homologous- recombination-based DNA repair (BRCA1/2). This analysis highlighted the repertoire of cancer genes and mutational processes operating in human, and represented a progress towards obtaining a comprehensive account of the somatic genetic basis of breast cancer.

This first analysis was then used to link known and novel drivers and mutational signatures to gene expression (transcriptome) of 266 cases [28]. One important and still debated question is to know to what extend somatic aberrations could trigger an immune-response. Our data suggested that substitutions of a particular type could be more effective in doing so than others.

Finally, in the context of ICGC, France was in charge of the analysis of a clinically specific subgroup of breast cancers, called HER2-positive, characterised by the HER2/ERBB2 amplification and over-expression. This is a subgroup for which several efficient targeted therapies (trastuzumab) are now available. However, resistance to treatment has been observed, revealing the underlying diversity of these cancers. An in-depth genomic and transcriptomic characterisation of 64 HER2-positive breast tumour was carried out. We delineated four subgroups, based on the expression data, each of them with distinctive genomic features in terms of somatic mutations, copy-number changes or structural variations [12]. The results suggested that, despite being clinically delineated by a specific gene amplification, HER2-positive tumours actually melt into the luminal-basal breast cancer spectrum rather, probably following their "cell-of-origin" fate and suggesting that the ERBB2 amplification is an embedded event in the natural history of these tumours. Finally, WGS data allowed us to gain more information about the amplification process itself and brought some indications about

how (and maybe when) it arose. Whole genome paired-end sequencing provides two important experimental clues to this purpose: a) high dynamics and resolution analysis of copy numbers, and b) ability to pinpoint large scale structural rearrangements by using clipping and abnormal mapping of read pairs. We could show that, in several cases, the observed sequence of copy numbers as well as the orientation of clipped reads was consistent with a breakage-fusion-bridge folding mechanism (BFB). However, the observation of long distance and inter-chromosomal rearrangements further showed that the amplification is a complex event (or sequence of events), likely involving several amplicons on the same or different chromosomes and several intertwined mechanisms. Indeed one of the features of HER2+ tumours is the ubiquitous presence of firestorms, corresponding to multiple closely spaced amplicons on highly rearranged chromosomal arms. It is therefore tempting to combine two mechanisms to explain the complex amplification patterns observed: chromothripsis, which will generate a mosaic of fragments (but no amplification per se), followed by a BFB amplification of chromosomal arm(s). This work was done at the "Plateforme Bioinformatique Gilles Thomas" located at Centre Léon Bérard (Lyon).

Sequence comparison
Sequence comparison is a fundamental step in many important computational biology tasks, in particular the reconstruction of genomes, a first key step before being able to identify the molecular elements present in them.

Traditional algorithms for measuring approximation in sequence comparison are based on the notions of distance or similarity, and are generally computed through sequence alignment techniques. As circular molecular structures are a common phenomenon in nature, a caveat of the adaptation of alignment techniques for circular sequence comparison is that they are computationally expensive, requiring from super-quadratic to cubic time in the length of the sequences. We introduced a new distance measure based on q-grams, and showed how it can be applied effectively and computed efficiently for circular sequence comparison [15]. Experimental results, using real DNA, RNA, and protein sequences as well as synthetic data, demonstrated orders-of-magnitude superiority of our approach in terms of efficiency, while maintaining an accuracy very competitive in relation to the state of the art.

Data structures for text indexing and string (sequence) comparison
Suffix trees are important data structures for text indexing and string algorithms. For any given string w of length $n = |w|$, a suffix tree for w takes $O(n)$ vertices and links. It is often presented as a compacted version of a suffix trie for w, where the latter is the trie (or digital search tree) built on the suffixes of w. The compaction process replaces each maximal chain of unary vertices with a single arc. For this, the suffix tree requires that the labels of its arcs are substrings encoded as pointers to w (or equivalent information). On the contrary, the arcs of the suffix trie are labeled by single symbols but there can be $\Theta(n^2)$ vertices and links for suffix tries in the worst case because of their unary vertices. It was an interesting question if the suffix trie can be stored using $O(n)$ vertices. We addressed it and thus presented the linear-size suffix trie, which guarantees $O(n)$ vertices [11]. We used a new technique for reducing the number of unary vertices to $O(n)$, that stems from some results on anti-dictionaries. For instance, by using the linear-size suffix trie, we are able to check whether a pattern p of length $m = |p|$ occurs in w in $O(m \log |\Sigma|)$ time and we can find the longest common substring of two strings w_1 and w_2 in $O((|w_1| + |w_2|) \log |\Sigma||)$ time for an alphabet $\Sigma|$.

Haplotype assembly
Haplotype assembly is the computational problem of reconstructing haplotypes in diploid organisms and is of fundamental importance for characterising the effects of single-nucleotide polymorphisms on the expression of phenotypic traits. Haplotype assembly highly benefits from the advent of "future-generation" sequencing technologies and their capability to produce long reads at increasing coverage. Existing methods are not able to deal with such data in a fully satisfactory way, either because accuracy or performances degrade as read length and sequencing coverage increase or because they are based on restrictive assumptions.

By exploiting a feature of future-generation technologies – the uniform distribution of sequencing errors – we designed an exact algorithm, called HAPCOL, that is exponential in the maximum number of corrections for each single-nucleotide polymorphism position and that minimises the overall error-correction score [22]. We performed an experimental analysis, comparing HAPCOL to the current state-of-the-art combinatorial methods both on real and simulated data. On a standard benchmark of real data, we showed that HAPCOL is competitive with state-of-the-art methods, improving the accuracy and the number of phased positions. Furthermore, experiments on realistically simulated datasets revealed that HAPCOL requires significantly less computing resources, especially memory. Thanks to its computational efficiency, HAPCOL can overcome the limits of previous approaches, allowing to phase datasets with higher coverage and without the traditional all-heterozygous assumption.

HAPCOL is based on MEC (Minimum error correction) which is computationally hard to solve. However, some approximation-based or fixed-parameter approaches have been proved capable of obtaining accurate results on real data. In another work [5], we then attempted to expand the current characterisation of the computational complexity of MEC from such approximation and fixed-parameter tractability points of view. We showed that MEC is not approximable within a constant factor, whereas it is approximable within a logarithmic factor in the size of the input. Furthermore, we answered open questions on the fixed-parameter tractability for parameters of classical or practical interest: the total number of corrections and the fragment length. In addition, we presented a direct 2-approximation algorithm for a variant of the problem that has also been applied in the framework of clustering data. Finally, since polyploid genomes, such as those of plants and fishes, are composed of more than two copies of the chromosomes, we introduced a novel formulation of MEC, namely the k-ploid MEC problem, that extends the traditional problem to deal with polyploid genomes. We showed that the novel formulation remains both computationally hard and hard to approximate. Nonetheless, from the parameterised point of view, we proved that the problem is tractable for parameters of practical interest such as the number of haplotypes and the coverage, or the number of haplotypes and the fragment length.

6.3. Inferring and analysing the networks of molecular elements

Metamodules in transcriptomic analysis

The human microbiome plays a key role in health and disease. Thanks to comparative metatranscriptomics, the cellular functions that are deregulated by the microbiome in disease can now be computationally explored. Unlike gene-centric approaches, pathway-based methods provide a systemic view of such functions; however, they typically consider each pathway in isolation and in its entirety. They can therefore overlook the key differences that (i) span multiple pathways, (ii) contain bidirectionally deregulated components, (iii) are confined to a pathway region. To capture these properties, computational methods that reach beyond the scope of predefined pathways are needed.

By integrating an existing module discovery algorithm into comparative metatranscriptomic analysis, we developed METAMODULES, a novel computational framework for automated identification of the key functional differences between health- and disease-associated communities [20]. Using this framework, we recovered significantly deregulated subnetworks that were indeed recognised to be involved in two well-studied, microbiome-mediated oral diseases, such as butanoate production in periodontal disease and metabolism of sugar alcohols in dental caries. More importantly, our results indicated that our method can be used for hypothesis generation based on automated discovery of novel, disease-related functional subnetworks, which would otherwise require extensive and laborious manual assessment.

Metabolic environmental dialog

What an organism needs at least from its environment to produce a set of metabolites, *e.g.* target(s) of interest and/or biomass, has been called a minimal precursor set. Early approaches to enumerate all minimal precursor sets took into account only the topology of the metabolic network (topological precursor sets). Due to cycles and the stoichiometric values of the reactions, it is often not possible to produce the target(s) from a topological precursor set in the sense that there is no feasible flux. Although considering the stoichiometry makes the problem harder, it enables to obtain biologically reasonable precursor sets that we call stoichiometric. Recently a method to enumerate all minimal stoichiometric precursor sets was proposed in the literature. The relationship between topological and stoichiometric precursor sets had however not yet been studied.

Such relationship was explored in a recently accepted paper [3]. In there, we also presented two algorithms that enumerate all minimal stoichiometric precursor sets. The first one is of theoretical interest only and is based on the above mentioned relationship. The second approach solves a series of mixed integer linear programming (MILP) problems. We compared the computed minimal precursor sets to experimentally obtained growth media of several *Escherichia coli* strains using genome-scale metabolic networks.

The results showed that the second approach, called SASITA, efficiently enumerates minimal precursor sets taking stoichiometry into account, and allows for broad *in silico* studies of strains or species interactions that may help to understand *e.g.* pathotype and niche-specific metabolic capabilities.

This work was also part of the PhD of Martin Wannagat, defended in June 2016 [2].

Metabolic hyperstories
In the context of a PhD in the team (whose defence took place in Dec 8, 2016) [1] and using metabolomics data, we focused on inferring the metabolic behaviour of an organism when it is subjected to a change in conditions. In this case, one can infer the reactions impacted when the changes are controlled and known (*e.g.* exposition to toxic compounds, changes in culture conditions). However, understanding how the metabolism of an organism changes of equilibrium is also of interest to infer the processes related for example to a transition between a commensal or beneficial bacterium to a pathogenic one. This question led to two different methods. The first, that we called TOTORO (for TOpological analysis of Transient metabOlic RespOnse), is based on the topology of metabolic networks to infer the reactions involved in a transient state, when an organism goes from one state of growth to another. We proposed a novel definition using the directed hypergraph representation and discussed its application on a dataset of Yeast exposed to cadmium. We showed that this method suggests more complete solutions of the reactions impacted during the metabolic shift. The second method, called KOTOURA (for Kantitative analysis Of Transient metabOlic and regUlatory Response And control), offers a constraint-based perspective in a more quantitative approach. We applied it to a simulated dataset and we are currently trying to infer the possible quantitative responses to mutations with a more complete kinetic model. An image previously used is that condition-specific models provide a snapshot of the metabolism of an organism, whether it is at the evolutionary-time scale or at the scale of a specific environmental condition describing a physiological process. Our idea here is thus to infer the transitions between those snapshots.

Besides the PhD manuscript, two papers are in preparation to present this work. They should be submitted in early 2017. A prototype for the two methods is available at: http://hyperstories.gforge.inria.fr/.

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

Robustness of the parsimonious reconciliation method in cophylogeny
The currently most used method in cophylogenetic studies is the so-called *phylogenetic tree reconciliation*. In this model, we are given the phylogenetic tree of the hosts H, the one of the symbionts S, and a mapping ϕ from the leaves of S to the leaves of H indicating the known symbiotic relationships among present-day organisms. The common evolutionary history of the hosts and of their symbionts is then explained through a number of macroevolutionary events (four in general). A reconciliation is then a function λ which is an extension of the mapping ϕ between leaves to a mapping that includes all internal nodes and that can be constructed using the different types of events considered. An optimal reconciliation is usually defined in a parsimonious way: a cost is associated to each event and a solution of minimum total cost is searched for.

An important issue in this model is that it makes strong assumptions on the input data which may not be verified in practice. We examine two cases where this situation happens. The first is related to a limitation in the currently available methods for tree reconciliation where the association ϕ of the leaves is for now required to be a function. This is not realistic as a single symbiont species can infect more than one host. For each present-day symbiont involved in a multiple association, one is currently forced to choose a single one. The second case addresses a different type of problem related to the phylogenetic trees of hosts and symbionts. These indeed are assumed to be correct, which may not be the case. In this work, we addressed the problem of correctly rooting a phylogenetic tree.

We thus explored the robustness of the parsimonious tree reconciliation method under "editing" (multiple associations) or "small perturbations" of the input (rooting problem) [29].

An extended version of this paper has been submitted to *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Insights on the virulence of swine respiratory tract mycoplasmas through genome-scale metabolic modelling
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* was shown to be virtually asymptomatic, *M. hyopneumoniae* is known to be the causative agent of enzootic pneumonia and *M. hyorhinis* to be present in cases of pneumonia, polyserositis and arthritis. Nonetheless, the elevated genomic resemblance among these three mycoplasmas combined with their different levels of pathogenicity is an indication that they have unknown mechanisms of virulence and differential expression. In 2015, we performed whole-genome metabolic network reconstructions for these three mycoplasmas. The results obtained were then submitted for publication to *BMC Genomics*. The paper has since been published [13].

Maximal chain subgraphs and covers of bipartite graphs motivated by analysis of cytoplasmic incompatibility
In a previous work of the team (Nor *et al.American Naturalist*, 182(1):15-24, 2013; Nor*et al.Information and Computation*, 213:23-32, 2012), we showed that a minimum chain subgraph cover of a given bipartite graph provides a good model for identifying the minimum genetic architecture enabling to explain one type of manipulation, called *cytoplasmic incompatibility*, by some parasite bacteria on their hosts. This phenomenon results in the death of embryos produced in crosses between males carrying the infection and uninfected females. The observed cytoplasmic compatibility relationships, can then be represented by a bipartite graph with males and females in different classes. Moreover, as different minimum (resp. minimal) covers may correspond to solutions that differ in terms of their biological interpretation, the capacity to enumerate all such minimal chain covers becomes crucial.

We recently addressed three related problems that bear some interest for the above problem besides raising interesting theoretical questions [35]. One is the enumeration of all the maximal *edge induced* chain subgraphs of a bipartite graph, for which we provided a polynomial delay algorithm. We gave bounds on the number of maximal chain subgraphs for a bipartite graph and used them to establish the input-sensitive complexity of the enumeration problem. The second problem we treated was the one of finding the minimum number of chain subgraphs needed to cover all the edges a bipartite graph. For this, we provided an exact exponential algorithm with a non trivial complexity. Finally, we approached the problem of enumerating all minimal chain subgraph covers of a bipartite graph and showed that it can be solved in quasi-polynomial time.

An extended version of the conference paper has been submitted to a journal in December 2016.

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

On the Complexity of Quadratic-Time Solvable Problems

Quadratic-time solvable problems may be classified into two classes: problems that are solvable in *truly subquadratic* time (that is, in time $(n^{2-\epsilon})$ for some $\epsilon > 0$) and problems that are not, unless the well known Strong Exponential Time Hypothesis (in short, SETH) is false. We proved that some quadratic-time solvable problems are indeed easier than expected [6]. We provided an algorithm that computes the transitive closure of a directed graph in time $(mn^{\frac{\omega+1}{4}})$, where m denotes the number of edges in the transitive closure and ω is the exponent for matrix multiplication. As a side effect of our analysis, we were able to prove that our algorithm runs in time $(n^{\frac{5}{3}})$ if the transitive closure of the graph is sparse. The same time bounds hold if we want to check whether a graph is transitive, by replacing m with the number of edges in the graph itself. As far as we know, this gives us the fastest algorithm for checking whether a sparse graph is transitive. Finally, we applied our algorithm to the comparability graph recognition problem (which dates back to 1941): also in this case, we obtained the first truly subquadratic algorithm. We then dealt with some hardness results. In particular, we started from an artificial quadratic-time solvable variation of the k-SAT problem and constructed a graph of Karp reductions, proving that a truly subquadratic-time algorithm for any of the problems in the graph falsifies SETH. More specifically, the analysed problems were the following: computing the subset graph, finding dominating sets, computing the betweenness centrality of a vertex, computing the minimum closeness centrality, and computing the hyperbolicity of a pair of vertices. We were also able to include in our framework three proofs that had already appeared in the literature, concerning the problems of distinguishing between split graphs of diameter 2 and diameter 3, of solving the local alignment of strings, and of finding two orthogonal binary vectors inside a collection.

Enumeration of solutions produced by closure operations
In enumeration problems, we are interested in listing a set of elements, which can be of exponential cardinality in the size of the input. The complexity of such problems is thus measured in terms of their input and output sizes. An enumeration algorithm with a complexity polynomial in both sizes is called output polynomial or total polynomial time. Another more precise notion of complexity is related to the *delay*, that is to the time between the production of two consecutive solutions. We are especially interested in problems solvable with a delay polynomial in the input size. These are considered as the tractable problems in enumeration complexity.

We addressed the problem of generating all elements obtained by the saturation of an initial set by some operations [41]. More precisely, we proved that we can generate the closure by polymorphisms of a boolean relation with a polynomial delay. This implies for instance that we can compute with polynomial delay the closure of a family of sets by any set of "set operations" (*e.g.* union, intersection, difference, symmetric difference, etc.). To do so, we proved that for any set of operations $\mathcal{F}$, one can decide in polynomial time whether an element belongs to the closure by $\mathcal{F}$ of a family of sets. When the relation is over a domain larger than two elements, our generic enumeration method fails for some cases since the associated decision problem is NP-hard, and we then provide an alternative algorithm.

6.6. Going towards control

Combinatorial approach for microbial consortia synthetic design
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. This was motivated by the fact that such consortia may perform more complicated functions than could single populations and be more robust to environmental fluctuations. Success is however not always guaranteed. In particular, establishing which consortium is best for the production of a given compound or set thereof remains a great challenge. This is the problem we addressed in a paper accepted this year [16].

We thus 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 the interaction among the species in the consortium could improve the production line. In mathematical terms, given a weighted directed hypergraph $\mathcal{H}$, the problem is to enumerate all directed sub-hypergraphs whose sets of vertices and of hyperarcs are included in those of $\mathcal{H}$, enable to produce the set of targets of interest from a subset of the sources of $\mathcal{H}$, and are of minimum weight. We called this the Directed Steiner Hypertree (DSH) problem.

We showed that the main issue in terms of the complexity of the problem comes from the hyperarcs with multiple source vertices (we called those the *tentacular hyperarcs*), not from the possible multiplicity of the target vertices. This is not the only issue though, and we thus further demonstrated that even when there is only one target that needs to be reached, the problem remains NP-hard. When both parameters, number of tentacular hyperarcs and of targets, are fixed, the problem becomes tractable. We then explored two methods for addressing it. One is a dynamic programming approach, and the other logic programming using ASP (Answer Set Programming). The second was more efficient for now, and the software MULTIPUS is thus based on it.

As initial validations of the model and of the method, we applied it to two case-studies taken from the literature.

This work was also part of the PhD of Alice Julien-Laferrière defended in December 2016 [1].

GENSCALE Project-Team

7. New Results

7.1. HTS data processing

7.1.1. Providing end-user solutions, example from the Colib'read on galaxy project

Participants: Claire Lemaitre, Camille Marchet, Pierre Peterlongo.

Colib'read tools suite uses optimized reference-free algorithms for various analyses of NGS datasets, such as variant calling or read set comparisons. To facilitate data analysis and tools dissemination, we developed Galaxy tools and tool shed repositories. The galaxy package, facilitates the analyzis of raw NGS data for a broad range of life scientists [16].

7.1.2. Assembly of Streptococcus Bacteria

Participant: Dominique Lavenier.

With the microbiological and bacteriological group of the Rennes hospital, we design a new strategy to assemble the genomes of 40 *Streptococcus* bacteria. Each strain has been sequenced and independently assembled using different assembly tools. For a specific strain, a merge of the contigs is done using the MIX software. This step allows the number of contigs to be significantly reduced, resulting in a better final assembly compared to each individual assembly. The comparison with other known *Streptococcus* genomes indicates where phages are located in the genome [20].

7.1.3. Data-mining applied to GWAS

Participants: Pham Hoang Sun, Dominique Lavenier.

Identifying variant combination association with disease is a bioinformatics challenge. This problem can be solved by discriminative pattern mining that uses statistical functions to evaluate the significance of individual biological patterns. There is a wide range of such measures. However, selecting an appropriate measure as well as a suitable threshold in some specific practical situations is a difficult task. We propose to use the skypattern technique which allows combinations of measures to be used to evaluate the importance of variant combinations without having to select a given measure and a fixed threshold. Experiments on several real variant datasets demonstrate that the skypattern method effectively identifies the risk variant combinations related to diseases [28].

7.1.4. Variant detection in transcriptomic data

Participant: Camille Marchet.

We defined a method to identify, quantify and annotate SNPs (Single Nucleotide Polymorphisms) using RNA-seq reads only. Organisms with a poor quality or no reference genome can take benefit of this approach, as well as studies where not enough material is available for sequencing from one individual, where samples can be pooled. The method relies on motifs discovery and post-treatment in de Bruijn graphs built from the reads. It can be used for any species to annotate SNPs and predict their impact on proteins as well as test their association to a phenotype of interest. The approach has been validated using well known human RNA-seq data. Results have been compared with state of the art approaches for variant calling. We showed that the methods perform similarly in terms of precision and recall. Then we focused on the main target of the study, namely the non-model species. We finally validated experimentally the predictions of our method [18].

7.1.5. Faster de Bruijn graph compaction

Participant: Antoine Limasset.

We developed a new algorithm, called BCALM2, for the compaction of de Bruijn graphs. BCALM2 is a parallel algorithm based on minimizer repartition of sequences. This repartition allows the compaction of extremely large graphs with moderate memory usage and time. The compaction of a human sequencing graph can be done in 1 hour with only 3GB of memory and huge genomes, such as the pine and white spruce ones (more than 20Gbp each), can be handled using our approach on a regular server (2 days and 40GB of memory). Those results argue that BCALM2 is one order of magnitude more efficient than available approaches and can tackle the assembly bottleneck of constructing a compacted de Bruijn graph [14].

7.1.6. Scaffolding

Participants: Rumen Andonov, Sebastien François, Dominique Lavenier.

We developed a method for solving genome scaffolding as a problem of finding a long simple path in a graph defined by the contigs that satisfies additional constraints encoding the insert-size information. Then we solved the resulting mixed integer linear program to optimality using the Gurobi solver. We tested our algorithm on several chloroplast genomes and showed that it outperforms other widely-used assembly solvers by the accuracy of the results [25].

7.2. Sequence comparison

7.2.1. Metagenomics datasets comparison

Participants: Gaetan Benoit, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo.

We developed a new method, called Simka, to compare simultaneously numerous large metagenomics datasets. The method computes pairwise distances based on the amount of shared k-mers between datasets. The method scales to a large number of datasets thanks to an efficient kmer-counting step that processes all datasets simultaneoulsy. Additionnally, several distance definitions were implemented and compared, including some originating from the ecological domain. The method is currently applied to the TARA oceans project (more than 2000 datasets) which aims at comparing worldwide sea water samples (ANR HydrGen project) [12].

7.2.2. Read similarity detection

Participants: Camille Marchet, Antoine Limasset, Pierre Peterlongo.

Retrieving similar reads inside or between read sets is a fundamental task either for algorithmic reasons or for analyses of biological data. This task is easy in small datasets, but becomes particularly hard when applied to millions or billions of reads. In [24] we used a straightforward indexing structure that scales to billions of elements. We proposed two direct applications in genomics and metagenomics. These applications consist in either approximating the number of similar reads between dataset(s) or to simply retrieve these similar reads. They can be applied on distinct read sets or on a read set against itself.

7.3. Parallelism

7.3.1. Processing-in-Memory

Participants: Charles Deltel, Dominique Lavenier.

The concept of PIM (Processor 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 two bioinformatics algorithms for this new type of memory: sequence alignment and mapping [34] [33]. The first results show that blast-like algorithms or mapping algorithms can highly benefit from such memory and speed-up of more than 25 can be achieved [26].

7.3.2. GPU for graph algorithms

Participants: Rumen Andonov, Dominique Lavenier.

We describe three algorithms and their associated GPU implementations for two types of shortest path problems. These implementations target computations on graphs with up to millions of vertices and executions on GPU clusters. The first two algorithms solve the All-Pairs Shortest Path (APSP) problem. The first of these two algorithms allows computations on graphs with negative edges while the second trades this ability for better parallel scaling properties and improved memory access. The third algorithm solves the Single-Pair Shortest Path (SPSP) query problem. Our implementations efficiently exploit the computational power of 256 GPUs simultaneously. All shortest paths of a million vertex graph can be computed in 6 minutes and shortest path queries on the same graph are answered in a quarter of a millisecond. These implementations proved to be orders of magnitude faster than existing parallel approaches[30].

7.4. Data representation

7.4.1. Computational pan-genomics: status, promises and challenges

Participant: Pierre Peterlongo.

We took part to the Computational Pan-Genomics Consortium producing a “white paper” dedicated to computational pan-genomic. A pan-genome is a representation of the union of the genomes of closely related individuals (eg from a same species). Computational pan-genomics is a new sub-area of research in computational biology. In [19], we generalized existing definitions and we examined already available approaches to construct and use pan-genomes, discussed the potential benefits of future technologies and methodologies and reviewed open challenges from the vantage point of the above-mentioned biological disciplines.

7.4.2. Mapping reads on graphs

Participants: Pierre Peterlongo, Antoine Limasset.

Many published genome sequences remain in the state of a large set of contigs. Each contig describes the sequence found along some path of the assembly graph, however, the set of contigs does not record all the sequence information contained in that graph. Although many subsequent analyses can be performed with the set of contigs, one may ask whether mapping reads on the contigs is as informative as mapping them on the paths of the assembly graph.

In [17], we proposed a formal definition of mapping a sequence on a de Bruijn graph, we analysed the problem complexity, and we provided a practical solution. The proposed tool can map millions of reads per CPU hour on a de Bruijn graph built from a large set of human genomic reads. Results show that up to 22 % more reads can be mapped on the graph but not on the contig set.

7.5. Applications

7.5.1. Study of the rapeseed genome structure

Participants: Sebastien Letort, Pierre Peterlongo, Dominique Lavenier, Claire Lemaitre, Fabrice Legeai.

In collaboration with IGEPP (Institut de Génétique, Environnement et Protection des Plantes), INRA, and through two national projects, PIA Rapsodyn and France-Génomique Polysuccess, we are involved in the genome analysis of several rapeseed varieties. The Rapsodyn project has the ambition to insure long-term competitiveness of the rapeseed production through improvement of the oil yield and reduction of nitrogen inputs during the crop cycle. Rapeseed varieties must thus be selected from genotypes that favor low nitrogen input. DiscoSnp++ is here used to locate new variants among the large panel of rapeseed varieties which have been sequenced during the project.

The PolySuccess project aims to answer the following question: how a polyploid, such as the oilseed rape plant, becomes a new species? Oilseed rape (Brassica napus) being a natural hybrid between B.rapa and B.oleracea, different genomes of these three species have been sequenced to study their structures. The Minia assembly pipeline provides a fast way to generate contigs that are used for studying gene specificities.

7.5.2. GATB Production Pipeline

Participants: Patrick Durand, Charles Deltel.

The entire set of libraries and tools related to the GATB Software have been introduced within a professional environment to support high-quality C++ developments. It relies on the use of technology platforms available at Inria: OpenStack and Jenkins. Considering the latter, we have setup more than 50 Jenkins tasks to automate the entire software development based on GATB: C++ code compiling and testing, documentation creation, packaging and preparation of official releases, mirroring on public Github repositories. Code compilation and tests are done on Linux and MacOSX VMs. https://ci.inria.fr/gatb-core/

7.5.3. Variant predictions in the pea genome

Participant: Pierre Peterlongo.

Progress in genetics and breeding in pea suffered from the limited availability of molecular resources. SNP markers that can be identified through affordable sequencing processes without the need for prior genome reduction or a reference genome allow the discovery of thousands of molecular markers.

We have been involved with IGEPP (Institut de Génétique, Environnement et Protection des Plantes, INRA) in the application of the discoSnp++ tool, discovering SNPs on HiSeq whole genome sequencing of four pea lines. Validation of a subset of predicted SNPs showed that almost all generated SNPs are highly designable and that most (95 %) deliver highly qualitative genotyping result [13].

7.5.4. Analysis of insect pest genomes

Participant: Fabrice Legeai.

Within a large international network of biologists, GenScale has contributed to various projects for identifying important components involved in the adaptation of major agricultural pests to their environment. We provided the assemblies, the annotations and the comparisons of various insects genomes [29]. Following specific agreement or policy, these results are available for browsing and consulting to a restricted consortium or a large community through the BioInformatics platform for Agro-ecosystems Arthropods (http://bipaa.genouest.org/is). In particular, this year our work helped to identify aphid genes involved in the adaptation to their favorite plant [15], or genes that are differentially expressed between leaf- and root-feeding phylloxera [21]. Furthermore, in order to help scientists to consult and cross genomics and postgenomics data, we are developping AskOmics, an integration and interrogation software for (linked) biological data, within a strong partnership, with Dyliss and GenOuest [36], [27].

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 this year, have developed a qualitative model of the osmotic stress response in the model bacterium *Escherichia coli* for which more information is available in the literature. 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* [15] and in *Data in Brief* [16]. 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.

The tool used for the qualitative modeling and simulation of the regulatory mechanism underlying osmotic stress is GENETIC NETWORK ANALYZER (GNA). This tool describes the dynamics of gene regulatory networks by means of PLDE models, as described in Section 5.1 . GNA has been integrated with the other bioinformatics tools distributed by Genostar (http://www.genostar.com/). Version 8.7.2 of GNA was released by IBIS and Genostar this year and has been deposited at the Agence pour la Protection des Programmes (APP). Some bugs have been corrected in the new version and the program has been adapted to the latest versions of Java and the software platform of Genostar. Version 8.7.2 supports the SBML standard and is also capable of exporting its models to the newly-developed standard for qualitative models, SBML Qual. This standard has been elaborated by the community of developers of logical and related modeling tools (CoLoMoTo), in which the GNA developers participate.

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. This work was presented at the major bioinformatics conference ISMB/ECCB and published in the special issue of *Bioinformatics* associated with the conference last year. The linear inversion methods have been implemented in the Python package WELLFARE and integrated in the web application WELLINVERTER (Section 5.3). Funded by the Institut Français de Bioinformatique (IFB), Yannick Martin is currently extending WellInverter into a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. This web service will be deployed on the IFB platform and accompanied by extensive user documentation, online help, and a tutorial.

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 software for image analysis (segmentation, tracking, lineage reconstruction, ...) adapted to the requirements of mother machine applications are still missing. IBIS therefore collaborates 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 is 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. 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, much less is known about adaptation during growth transitions. In particular, only very few data are available on changes in the concentration and activity of the transcription and translation machineries and almost no data exost for the dynamic response of the degradation machinery.

In the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique INRA-Inria (2012-2015), the collaboration of Delphine Ropers with Muriel Cocaign-Bousquet and Brice Enjalbert at INRA/INSA de Toulouse has allowed to disentangle the role of post-transcriptional regulation from other regulatory interactions in the dynamic adaptation of central carbon metabolism in *E. coli*. In a multi-scale analysis of a wild-type strain and its isogenic mutant attenuated for the protein CsrA, a variety of experimental data have been acquired in relevant conditions, including growth parameters, gene expression levels, metabolite pools, enzyme activities and metabolic fluxes. Data integration, metabolic flux analysis and regulation analysis revealed the pivotal role of post-transcriptional regulation for shaping carbon metabolism. In particular, the work has shed light on *csrA* essentiality and has provided an explanation for the glucose-phosphate stress observed in the mutant strain. A paper summarizing the work has been published in *Molecular Microbiology* this year [14]. A follow-up study conducted with various mutant strains of the carbon storage regulator system has 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. A paper summarizing the work is being prepared for publication.

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.

6.4. 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.

In the context of the response of yeast cells to osmotic shocks, in collaboration with the LIFEWARE project team and colleagues from Université Paris Descartes and University of Pavia (Italy), Eugenio Cinquemani has investigated the use of mixed effects-modelling and identification techniques to characterize individual cell dynamics in isogenic cell populations. Mixed-effects models are hierarchical models where parametric response profiles of individuals is subject to inter-individual parameter variability following a common population distribution. Starting from identification approaches in pharmacokinetics, we have developed and applied inference methods to microfluidics data, with a focus on the response of budding yeast to osmotic shocks. Results were described in a publication in *PLoS Computational Biology* [13]. A study of statistical validation methods for mixed-effects and alternative stochastic modelling paradigms has been presented at the *IFAC Conference on Foundations of Systems Biology in Engineering (FOSBE)* in Magdeburg [19]. In collaboration with the project-team BIOCORE at Inria Sophia-Antipolis - Méditerranée, the approach is now being investigated for the joint modelling of growth and gene expression in *E. coli*, based on single-cell microfluidics data from growth arrest-and-restart experiments. Further challenges stemming from this activity toward modelling and identification of extrinsic noise in individual cells are part of the recently started ANR project MEMIP (Section 8.2).

Work on identification and state estimation for single-cell gene network dynamics has been focused on the reconstruction of promoter activity profiles from fluorescent reporter data. In a stochastic, intrinsic noise modelling context, Eugenio Cinquemani addressed the problem of inferring promoter activity statistics over a cell population, such as mean and variance, from analogous statistics of the reporter output, as obtained from so-called population snapshot data. This nontrivial extension of the deterministic promoter activity deconvolution problem from population-average data is the first, crucial step toward reconstruction of promoter activity regulation and inference of stochastic network models. Earlier results, concerning parameter identifiability of stochastic promoter activity models and reconstruction of promoter activity distributions in the special case of single-switch systems, were further developed in a contribution to the HSB conference this year [18]. The relationship between the spectrum of the promoter process (cross-correlation function) and the mean-variance profiles of fluorescent reporter readouts was derived and demonstrated on examples, laying down the bases for a full-blown observability analysis and the development of spectrum estimation methods.

The collaboration of Eugenio Cinquemani with Marianna Rapsomaniki (IBM Zurich Research Lab, Switzerland), Zoi Lygerou (University of Patras, Greece) and John Lygeros (ETH Zurich, Switzerland) is moving on to applications of joint work published in *Bioinformatics* last year. Deployment of the methods developed into an efficient cluster-based software for the inference of protein kinetics in single cells from Fluorescence Recovery After Photobleaching (FRAP) experiments is under study. Exploitation of the same methods for the simulation and analysis of more general biochemical processes in single cells is part of the ongoing research efforts.

6.5. 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* last year. 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 is being continued in several directions in the context of the RESET project by Célia Boyat. In order to further explore the possibility of transferring this technology to biotechnology companies, we participated in the Challenge Out of Labs (http://www.linksium.fr/lancez-vous/resultat-challenge-out-of-labs/) organized by Linksium, the local incubator for technology transfer and start-up building. The presentation by Hans Geiselmann was selected for further development by Linksium.

In a review recently accepted for publication in *Trends in Microbiology* [11], 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.

Whereas the synthetic growth switch has been designed for biotechnological purposes, the question can be asked how resource allocation is organized in wild-type strains that have naturally evolved. Recent work has shown that coarse-grained models of resource allocation can account for a number of empirical regularities relating the the macromolecular composition of the cell to the growth rate. Some of these models hypothesize control strategies enabling microorganisms to optimize growth. While these studies focus on steady-state growth, such conditions are rarely found in natural habitats, where microorganisms are continually challenged by environmental fluctuations. The aim of the PhD thesis of Nils Giordano is to extend the study of microbial growth strategies to dynamical environments, using a self-replicator model. In collaboration with the BIOCORE project-team, we formulate dynamical growth maximization as an optimal control problem that can be solved using Pontryagin's Maximum Principle. We compare this theoretical gold standard with different possible implementations of growth control in bacterial cells. We find that simple control strategies enabling growth-rate maximization at steady state are suboptimal for transitions from one growth regime to another, for example when shifting bacterial cells to a medium supporting a higher growth rate. A near-optimal control strategy in dynamical conditions is shown to require information on several, rather than a single physiological variable. Interestingly, this strategy has structural analogies with the regulation of ribosomal protein synthesis by ppGpp in *E. coli*. It involves sensing a mismatch between precursor and ribosome concentrations, as well as the adjustment of ribosome synthesis in a switch-like manner. Our results show how the capability of regulatory systems to integrate information about several physiological variables is critical for optimizing growth in a changing environment. A paper describing the above results was published in *PLoS Computational Biology* this year [12].

LIFEWARE Project-Team

7. New Results

7.1. Analog computation in the cell: specifications, compilation into biochemical reactions and computational complexity

Participants: François Fages, Guillaume Le Guludec.

The continuous nature of many protein interactions leads us to consider mixed analog-digital computation methods, for which the recent results in the theory of analog computability and complexity obtained by Amaury Pouly [0] and Olivier Bournez, establish fundamental links with digital computation. In [18], we derive from these results a Turing completeness result for elementary reaction systems (without polymerization) under the differential semantics, and present a compiler of behavioural specifications into biochemical reactions which can be compared to natural circuits acquired through evolution. We illustrate this approach through the example of the MAPK signaling module which has a function of analog-digital converter in the cell, and through the cell cycle control.

The biochemical compiler is implemented in BIOCHAM v4.0 which will be soon released. We plan to use it in the ANR-MOST project BIOPSY on "Biochemical Programming" for the design of artificial biosensors and the programming of (non-living) protocells in collaboration with Franck Molina, CNRS Sys2diag lab, Montpellier.

7.2. Influence systems vs reaction systems

Participants: François Fages, Thierry Martinez, David Rosenblueth, Sylvain Soliman, Denis Thieffry.

In Systems Biology, modelers develop more and more reaction-based models to describe the mechanistic biochemical reactions underlying cell processes. They may also work, however, with a simpler formalism of influence graphs, to merely describe the positive and negative influences between molecular species. The first approach is promoted by reaction model exchange formats such as SBML, and tools like CellDesigner, while the second is supported by other tools that have been historically developed to reason about boolean gene regulatory networks. In practice, modelers often reason with both kinds of formalisms, and may find an influence model useful in the process of building a reaction model. In [11], we introduce a formalism of influence systems with forces, and put it in parallel with reaction systems with kinetics, in order to develop a similar hierarchy of boolean, discrete, stochastic and differential semantics. We show that the expressive power of influence systems is the same as that of reaction systems under the differential semantics, but weaker under the other interpretations, in the sense that some discrete behaviours of reaction systems cannot be expressed by influence systems. This approach leads us to consider a positive boolean semantics which we compare to the asynchronous semantics of gene regulatory networks à la Thomas. We study the monotonicity properties of the positive boolean semantics and derive from them an efficient algorithm to compute attractors.

7.3. What population reveals about individual cell identity: Single-cell parameter estimation of models of gene expression in yeast

Participants: Grégory Batt, Pascal Hersen, Artémis Llamosi.

[0] Amaury Pouly, "Continuous models of computation: from computability to complexity", PhD Thesis, Ecole Polytechnique, Nov. 2015.

Significant cell-to-cell heterogeneity is ubiquitously observed in isogenic cell populations. Consequently, parameters of models of intracellular processes, usually fitted to population-averaged data, should rather be fitted to individual cells to obtain a population of models of similar but non-identical individuals. In [6], we propose a quantitative modeling framework that attributes specific parameter values to single cells for a standard model of gene expression. We combine high quality single-cell measurements of the response of yeast cells to repeated hyperosmotic shocks and state-of-the-art statistical inference approaches for mixed-effects models to infer multidimensional parameter distributions describing the population, and then derive specific parameters for individual cells. The analysis of single-cell parameters shows that single-cell identity (e.g. gene expression dynamics, cell size, growth rate, mother-daughter relationships) is, at least partially, captured by the parameter values of gene expression models (e.g. rates of transcription, translation and degradation). Our approach shows how to use the rich information contained into longitudinal single-cell data to infer parameters that can faithfully represent single-cell identity. This is the first demonstration that biologically-meaningful values for parameter of intracellular processes can be attributed to individual cells.

7.4. The cost of cellular adaptation to stress: tradeoff between short-term and long-term adaptation to osmotic stress in yeast

Participants: Grégory Batt, Ewen Corre, Pascal Hersen, Artémis Llamosi.

Upon stress, cells have evolved complex adaptation strategies to environmental variations which include sensing, information processing and modification of metabolic and transcriptional activity. The reaction of yeast cells to hyperosmotic stress spans several timescales and includes massive gene-expression changes, bio-compatible osmolyte production, and direct action on the cell cycle. Despite a detailed knowledge of molecular events, the impact of stress-response on cellular resources is poorly known. In particular, strong and fast adaptation which alter proliferation in the short term while conferring advantage on the long term are important drivers of stress-response evolution.

In this study, we use microfluidics to vary dynamically both the source of cost (osmotic stress) and the available metabolic resources (glucose) while monitoring cellular proliferation. We show that, under hyper-osmotic stress, metabolic resources are rerouted towards the production of glycerol through activation of an essential enzyme in glycerol production. This reveals the nature of the burden imposed by osmotic stress and, more generally, allows us to better understand the evolutionary tradeoffs between stress response and proliferation.

7.5. Balancing a genetic inverted pendulum

Participants: Grégory Batt, Pascal Hersen, Jean-Baptiste Lugagne, Jean-Baptiste Caron.

The ability to routinely control complex genetic circuits in vivo and in real-time promises quantitative understanding of cellular processes of unprecedented precision and quality. With combined efforts in microfluidic design, microscope automation, image analysis, modeling and control theory, we propose a platform for real-time, single-cell, *in silico* control of genetic networks in *E. coli*. The circuit we are trying to control is a genetic toggle switch, a foundational circuit in synthetic biology, which consists of two genes that repress each other. This genetic system features two stable equilibrium points where one of the genes has taken over. Our objective is to dynamically balance the circuit in single cells around a third, unstable equilibrium point at which no gene dominates and their mutual repression strengths are balanced. This is similar to the landmark problem in control theory of stabilizing an inverted pendulum in its upright position. Although our work indicates that this real-time control approach can drive convoluted genetic networks towards states that are inaccessible to traditional genetic perturbations such as knock-outs and promoter induction, the a priori quantitative knowledge of the system required for achieving this control is minimal. We show that even a simple Proportional-Integral controller can maintain in a balanced state the toggle switch in single cells. Finally, we demonstrate that similar results can be obtained by applying periodic inductions, identical to all cells in the population. Given the fact that all cells behave differently, this result was highly unexpected. It can however be understood as an example of dynamic stabilization, analogous to the solution proposed by Kapitza for the inverted pendulum.

These results are presented in the PhD thesis of Jean-Baptiste Lugagne [2].

7.6. A look-ahead simulation algorithm for DBN models of biochemical pathways

Participants: Grégory Batt, Sucheendra Palaniappan.

Dynamic Bayesian Networks (DBNs) have been proposed as an efficient abstraction formalism of biochemical models. They have been shown to approximate well the dynamics of biochemical models, while offering improved efficiency for their analysis. In [14], we compare different representations and simulation schemes on these DBNs, testing their efficiency and accuracy as abstractions of biological pathways. When generating these DBNs, many configurations are never explored by the underlying dynamics of the biological systems. This can be used to obtain sparse representations to store and analyze DBNs in a compact way. On the other hand, when simulating these DBNs, singular configurations may be encountered, that is configurations from where no transition probability is defined. This makes simulation more complex. We initially evaluate two simple strategies for dealing with singularities: first, re-sampling simulations visiting singular configurations; second filling up uniformly these singular transition probabilities. We show that both these approaches are error prone. Next, we propose a new algorithm which samples only those configurations that avoid singularities by using a look-ahead strategy. Experiments show that this approach is the most accurate while having a reasonable run time.

7.7. Logical model specification aided by model-checking techniques: application to the mammalian cell cycle regulation

Participants: François Fages, Sylvain Soliman, Denis Thieffry, Pauline Traynard.

Understanding the temporal behaviour of biological regulatory networks requires the integration of molecular information into a dynamical model. However, the analysis of model dynamics faces a combinatorial explosion as the number of regulatory components and interactions increases. In [8], we use model-checking techniques to verify sophisticated dynamical properties resulting from the model influence structure in the absence of kinetic assumption. We demonstrate the power of this approach by analysing a Boolean influence model of the molecular network controlling mammalian cell cycle. This approach enables a systematic analysis of model properties, the delineation of model limitations, and the assessment of various refinements and extensions based on recent experimental observations. The resulting logical model accounts for the main irreversible transitions between cell cycle phases, the sequential activation of cyclins, and the inhibitory role of Skp2, and further emphasizes the multifunctional role for the cell cycle inhibitor Rb.

7.8. Model-based investigation of the circadian clock and cell cycle coupling in mouse embryonic fibroblasts: prediction of RevErb–α up-regulation during mitosis

Participants: François Fages, Sylvain Soliman, Pauline Traynard.

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. 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 [9], in order to explain these observations, we study a possible entrainment of the circadian clock by the cell cycle through a regulation of clock genes around the mitosis phase. We develop a computational model and a formal specification of the observed behavior to investigate the conditions of entrainment in period and phase. We show that either the selective activation of RevErb-α or the selective inhibition of Bmal1 transcription during the mitosis phase, allow 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. We conclude on the arguments favouring the RevErb-α up-regulation hypothesis and on some further predictions of the model.

7.9. Stochastic continuous optimization backend for the constraint modelling language MiniZinc with applications to geometrical placement problems

Participants: François Fages, Thierry Martinez, Sylvain Soliman.

MiniZinc is a solver-independent constraint modeling language which is increasingly used in the constraint programming community. It can be used to compare different solvers which are currently based on either Constraint Programming, Boolean satisfiability, Mixed Integer Linear Programming, and recently Local Search. In [12], [13] we present a stochastic continuous optimization backend for MiniZinc models over real numbers. More specifically, we describe the translation of FlatZinc models into objective functions over the reals, and their use as fitness functions for the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) solver. We illustrate this approach with the declarative modeling and solving of hard geometrical placement problems [16], motivated by packing applications in logistics [10] involving mixed square-curved shapes and complex shapes defined by Bézier curves.

Beyond these applications to packing problem, our real motivation for these developments is the solving of parameter search problems in computational systems biology and its implementation in BIOCHAM.

7.10. Mixture model CMA-ES

Participants: François Fages, Nicolas Vasselin.

In [19], we report on our attempt to improve the CMA-ES global optimization algorithm based on two ideas: first the use of Sobol's quasi-random low discrepancy numbers instead of pseudo-random numbers, second the design of a mixture model extension of CMA-ES (MM-CMA-ES) which, instead of doing restarts with an important loss of information at each restart, evolves a dynamic set of multivariate normal distributions in parallel, using an EM clustering algorithm at each step to decide of population splittings and mergings. On the standard Coco benchmark for evaluating global stochastic optimization methods, the use of Sobol numbers shows a quite uniform improvement by 30% as was already shown by Teytaud last year [0]. On the other hand, MM-CMA-ES does not show speed-up w.r.t. CMA-ES with IBOP restart strategy, even on objective functions with many local minima such as the Rastragin function. The reason is the overhead in the number of evaluation of the objective functions, introduced by the MM strategy, and the very subtle effect of the adaptive step size strategy of CMA-ES to escape from the covering of several local minima by one (large) normal distribution.

7.11. Metro energy optimization through rescheduling

Participants: François Fages, Thierry Martinez.

The use of regenerative braking is a key factor to reduce the energy consumption of a metro line. In the case where no device can store the energy produced during braking, only the metros that are accelerating at the same time can benefit from it. Maximizing the power transfers between accelerating and braking metros thus provides a simple strategy to benefit from regenerative energy without any other hardware device. In [15], we use a mathematical timetable model to classify various metro energy optimization problems studied in the literature and prove their NP-hardness by polynomial reductions of SAT. We then focus on the problem of minimizing the global energy consumption of a metro timetable by modifying the dwell times in stations. We present a greedy heuristic algorithm which aims at locally synchronizing braking trains along the timetable with accelerating trains in their time neighbourhood, using a non-linear approximation of energy transfers. On a benchmark of the litterature composed of six small size timetables, we show that our greedy heuristics performs better than CPLEX using a MILP formulation of the problem with a linear approximation of the objective function. We also show that it runs ten times faster than a state-of-the-art evolutionary algorithm, called the covariance matrix adaptation evolution strategy (CMA-ES), using the same non-linear objective function on these small size instances. On real data leading to 10000 decision variables on which both MILP and CMA-ES do not provide solutions, our dedicated algorithm computes solutions with a reduction of energy consumption ranging from 5% to 9%.

[0] O. Teytaud. Quasi-random numbers improve the CMA-ES on the BBOB testbed. Artificial Evolution (EA2015), 2015, Lyon, France. Springer Verlag, pp.13

This work done in 2014 in the Cifre PhD Thesis of David Fournier with General Electric Transportation has received this year the Gold Medal of the *Annual Alstom Contest* “I Nove You” in the “Green Innovation” category.

MORPHEME Project-Team

5. New Results

5.1. Multi-Angle TIRF reconstruction for studying the cell adhesion phenomenon

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

This work is made in collaboration with Agata Radwanska and Ellen Van Obberghen-Schilling from Institut de Biologie Valrose (iBV) at Nice.

Understanding cell adhesion mechanism is of a major importance in biology for example in the context of tumoral angiogenesis [0]. However, this process occurs at the vicinity of the cell membrane within a layer of a few hundred nanometer making classical microscopy devices unable to image such biological structures due to their lack of resolution in the axial direction. An interesting alternative would be to use a multi-angle total internal reflection illumination together with numerical reconstruction algorithms in order to reach a nanoscale precision in the axial direction.

Following this idea, we made use of our previous work on MA-TIRF reconstruction to produce color-coded maps (see the example on Figure 1), with an axial resolution of 20 nm, of biological samples provided by Agata Radwanska and Ellen Van Obberghen-Schilling from the Institut de Biologie Valrose. The information obtained from the study of the reconstructed images have confirmed known behaviors of some proteins involved in the cell adhesion process allowing us, by this way, to complete the validation of our reconstruction method. Moreover, the 3D reconstructions have provided new information concerning the axial position of the observed biological proteins, information which was unavailable for previous studies conducted with other microscopy systems.

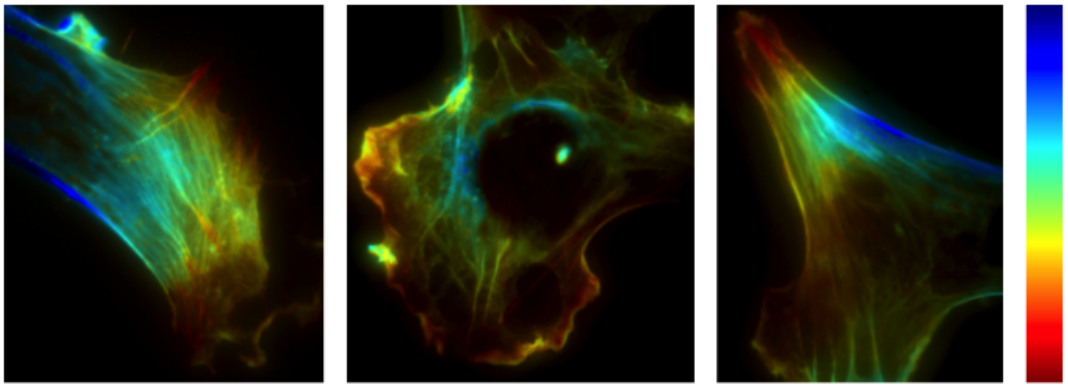


Figure 1. Example of color-coded representations for the reconstructed samples. The colors correspond to the depth (in the axial direction) of the biological structures (in the colorbar on the right: red = 0 nm and dark blue = 400 nm).

[0] Process of blood vessels creation from existing ones.

5.2. Exact continuous penalties for ℓ_2-ℓ_0 minimization: Application to Photo Activated Localization Microscopy (PALM)

Participants: Simon Gazagnes, Emmanuel Soubies, Laure Blanc-Féraud.

In conventional microscopy techniques, the spatial resolution of an image is limited by the diffraction phenomena. Recent methods like photo-activated localization microscopy (PALM) allow high-precision molecule localization by sequentially activating and imaging a small random set of fluorescent molecules in the sample. However, the quality of this super-resolved image is related to the density of emitters activated at each acquisition and the numerical method used to locate molecules.

Applications for these microscopy techniques are then mainly restricted by the number of acquisitions required to obtain the superresolved image. One way to overcome this limitation is to increase the density of emitters activated at each acquisition. Nevertheless, it will cause overlapping for a certain number of spots on the acquired image which makes the localization of the underlying molecules a harder task. Considering such a high density setting, we have proposed to perform the molecules localization by solving a ℓ_0-penalized least squares criteria through the minimization of the Continuous Exact ℓ_0 (CEL0) relaxation that we have previously proposed. The method has provided interesting results, competing with state of the art methods, as shown on Figure 2 . This work has been submitted for the conference ISBI 2017.

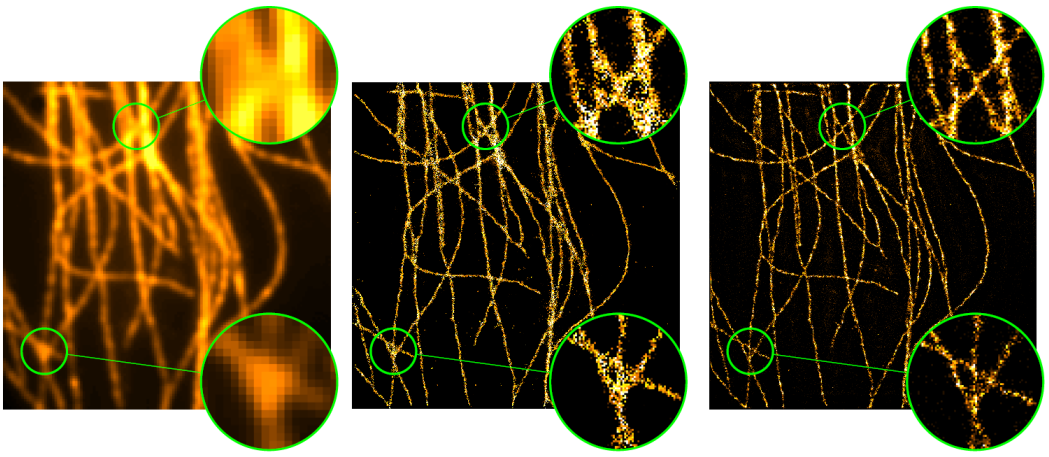


Figure 2. From left to right: Conventional Wide Field image, PALM with DAOSTORM (state of the art algorithm) reconstruction, PALM with the proposed reconstruction.

5.3. Exact continuous penalties for ℓ_2-ℓ_0 minimization: Application to Channel and Direction Of Arrival (DOA) estimation problems

Participants: Emmanuel Soubies, Laure Blanc-Féraud.

This work is made in collaboration with Adilson Chinatto, Cynthia Junqueira, João M. T. Romano (University of Campinas, Brazil) and Pascal Larzabal, Jean-Pierre Barbot (ENS Cachan, SATIE Lab).

In this work, we have proposed to extend the Continuous Exact ℓ_0 (CEL0) penalty, which we initially introduced for the real single measurement vector (SMV) case, to complex SMV and complex multiple measurement vector (MMV) situations involving structured sparsity. Such an extension is necessary to address sparse signal processing estimation problems like augmented resolution channel estimation and direction of

arrival (DOA) estimation for which the mixture matrix do not verify restrict isometry property (RIP) and incoherence conditions. We thereby have derived a row-structured version of the CEL0 penalty and showed that the relations between minimizers of the resulting relaxation and those of the initial ℓ_0-penalised least squares criteria, that we previously showed in the real SMV case, are still valid for complex SMV and MMV situations using the proposed row-structured CEL0 penalty. Finally, we have employed state of the art nonsmooth nonconvex algorithms to minimize the proposed relaxation and we have compared the results obtained by our method with those provided by the well known iterative hard thresholding (IHT) algorithm as well as some classical algorithms for the studied problems. We have shown that minimizing the row-structured CEL0 relaxation provides better estimation results than IHT, which minimizes directly with the initial ℓ_0-penalized least-squares criteria, and than classical algorithms used for such problems where the mixture matrix is highly correlated. Moreover, the proposed method is able to reach the oracle RMSE in some cases. This work has been submitted to the IEEE Transaction on Signal Processing journal.

5.4. Phase estimation in Differential Interference Contrast (DIC) microscopy

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

We address the problem of estimating the phase from color images acquired with differential-interference-contrast microscopy. In particular, we consider the nonlinear and nonconvex optimization problem obtained by regularizing a least-squares-like discrepancy term with a total variation functional, possibly smoothed with the introduction of a positive constant. We deeply investigate the analytic properties of the resulting objective function, proving the existence of minimum points, and several optimization methods able to address the minimization problem. Besides revisiting the conjugate gradient method proposed in the literature for this problem and comparing it with standard conjugate gradient approaches, we introduce more recent effective optimization tools able to obtain both in the smooth and in the non smooth case accurate reconstructions with a reduced computational demand.

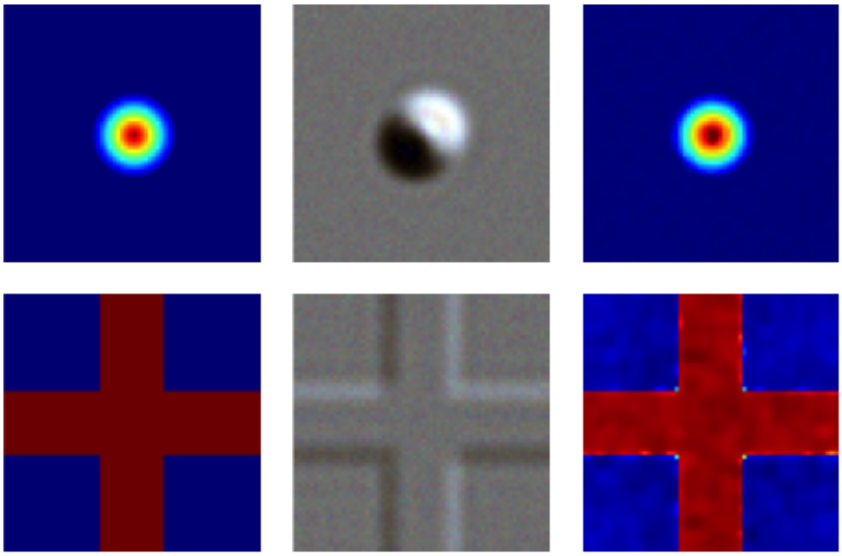


Figure 3. Data and results for the cone (top row) and cross (bottom row) objects. From left to right: true object, noisy DIC color image taken at shear angle $\frac{\pi}{4}$ rad *and corrupted with white Gaussian noise at SNR = 4.5 dB, and reconstructed phase with the LMSD method from observations at shear angles equal to* $-\pi/4$ rad *and* $\pi/4$ rad.

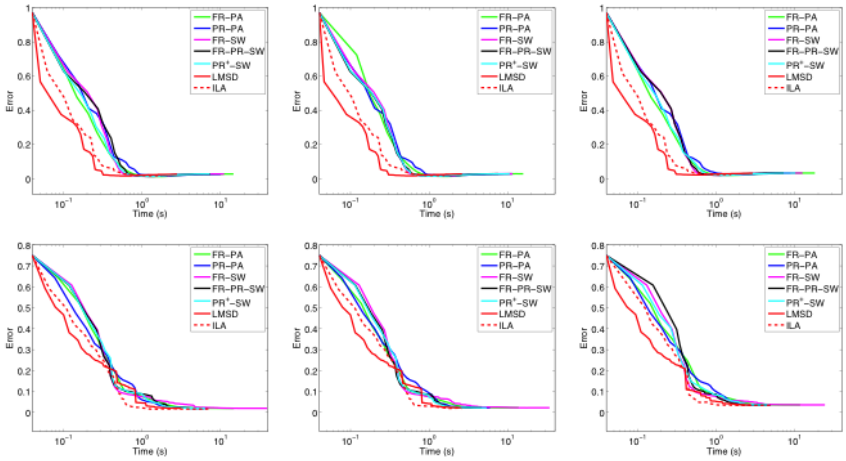


Figure 4. Error versus computational time plots for the cone (top row) and cross (bottom row) objects. From left to right: noise-free data, SNR = 9 dB and SNR = 4.5 dB.

5.5. White Blood Cells Segmentation and Classification in Bone Marrow Images

Participants: Mohammed Lamine Benomar, Xavier Descombes.

This work is made in collaboration with Chikh Amine and Mourtada Benazzouz from GBM Lab. (Tlemcen University). Our experiments were performed on an image database acquired in the Hemobiology service of the Tlemcen Hospital (Algeria).

The differential count of white blood cells (WBC) for medical diagnosis requires a careful observation in peripheral blood and bone marrow microscopic images in order to detect abnormal or suspicious cells. However, this process (screening) is time consuming, requiring concentration, experience and competence of the expert. The diagnosis depends on the correct recognition of cells. For that, computer analysis image system is required to automate the process in order to help experts, reduce the time and increase the accuracy. The main important steps in such systems are segmentation and classification of white blood cells.

The proposed approach to locate WBC in bone marrow microscopic smear could be divided into three main steps: pre-processing, segmentation and classification. The main concept of the segmentation and classification algorithm employed uses WBCs color, texture and morphological properties.

The first step is to reveal chromatic characteristics of the WBC by applying decorrelation stretch to multi-channel RGB image, simple color transformation and Otsu thresholding to suppress background and most of the red blood cells. In the segmentation step, two techniques have been used which are Marker Controlled Watershed followed by MLE (Maximum Likelihood Estimator) to differentiate between WBC, the grouped red blood cells and artifacts using shape, color and texture features. Then Otsu thresholding based on HSL color space to separate WBC nucleus and cytoplasm (see Figure 5). Finally, white blood cells were classified into two categories related to the type of Myeloma, this step is based on features extraction and then applying a classifier.

5.6. Classification of the extracellular matrix

Participants: Raphael Meunier, Anca-Ioana Grapa, Laure Blanc-Féraud, Xavier Descombes, Sébastien Schaub.

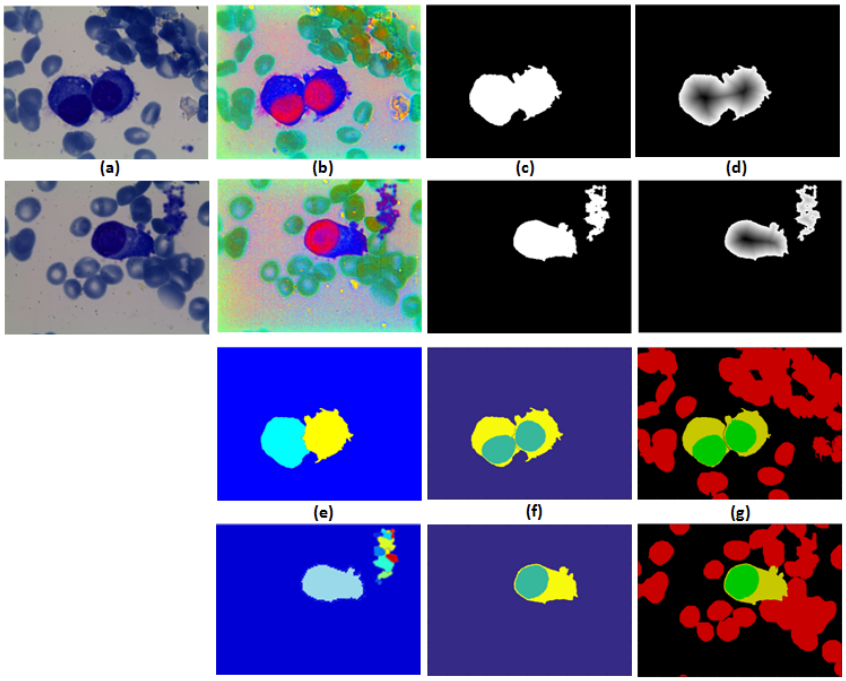


Figure 5. Image segmentation step: (a) Input image, (b) RGB Decorrelation stretch, (c) Binary mask, (d) Distance transform, (e) Watershed, (f) Segmented cell, (g) Ground truth.

This work is made in collaboration with Ellen Van Obberghen eand Georgios Efthymiou (iBV).

Cells of multicellular organisms interact continually with their local environment which is largely determined by the extracellular matrix (ECM). The biochemical, topological and physical properties (stiffness, elasticity) of the ECM regulate many physiological processes (embryonic development and tissue repair) and their dysregulation plays a key role in the evolution of inflammatory, fibrotic and tumoral diseases. Fibronectin (FN) is a major component of the ECM. The biologists at iBV have identified certain molecular mechanisms involved in the assembly of FN into fibrillar arrays (FN fibrillogenesis) on the cell surface. The resulting fibrillar networks display variable densities and organizations that convey specific biological signals to the cells that encounter them (see figure 6).

We have developed a classification scheme that consists in clustering features extracted from the images to define a texture dictionary. The extracellular matrix are then classified with respect to their signature on this dictionary. We have compared two sets of features that are SIFT histograms and the curvelet coefficients. The SIFT approach appears to be more discriminant for classification purposes but the curvelet approach is better suited for modeling the texture. Next step will consists in modeling the extracellular matrix.

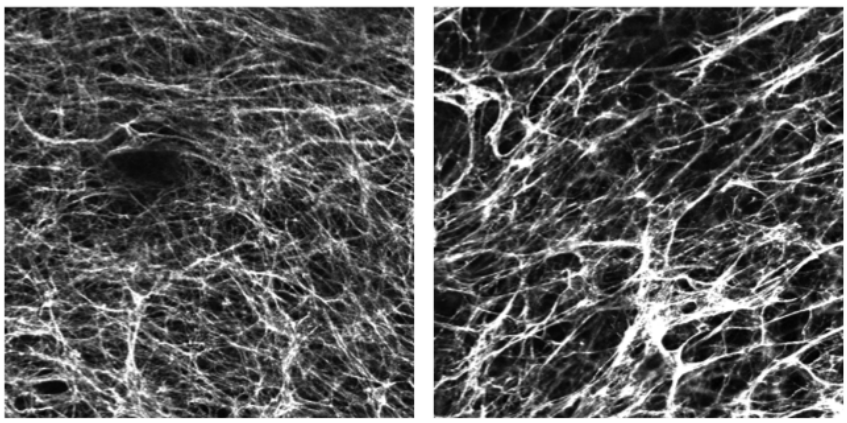


Figure 6. Two examples of extracellular matrices.

5.7. CNuclei/cytoplasm detection and classification in genome-wide RNAi screens

Participants: Eric Debreuve, Djampa Kozlowski, Florence Besse, Xavier Descombes.

This work is made in collaboration with Fabienne De Graeve (iBV).

The work described hereafter is part of the RNAGRIMP ANR project which started in January 2016 and lasts 48 months (see Section 7.2.1). A pilot genome-wide RNAi (Ribonucleic Acid interference) screen on Drosophila cultured cells has been performed with different mutant conditions. The purpose is to study the density and repartition of cytoplasmic RNP (RiboNucleoprotein Particles) granules containing the IMP protein (IGF-II mRNA-binding protein where IGF stands for Insulin-like Growth Factor).

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). A first task that must be accomplished is to detect the nuclei on the DAPI images, and to learn a classification procedure into *living cell* or *dead cell* based on morphologic and

radiometric nuclei properties. A CellProfiler [0] pipeline has been developed to automatically detect the nuclei and compute some properties on them. The detection was based on the following main steps: intensity re-scaling, Kapur-based thresholding, and small object discarding. For each detected nucleus, the computed properties were (non-exhaustive list) average intensity, area, granularity, circularity ...

Then, a learning set has been built where a significant number of nuclei were manually assigned their correct (*living* or *dead*) class by a biologist of the team. This learning set was fed to CellProfiler Analyst [0] in order to learn a decision tree for automatic nuclei (hence, cell) classification (see Fig 7 , left).

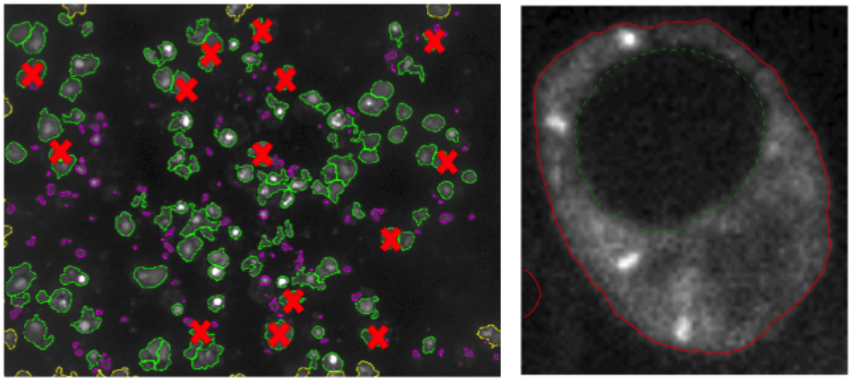


Figure 7. Left: automatic classification of the detected nuclei into living (encircled in green) or dead (with a red cross); objects encircled in yellow are cropped by the field of view, and objects encircled in purple are too small ; they are all discarded. Right: active contour segmentation of the cytoplasm of a cell (previously classified as a living cell); red contour: cytoplasm external boundary; green, dashed contour: nucleus boundary (also cytoplasm internal boundary).

Once the living cell nuclei have been identified on the DAPI images, the next step is to segment (*i.e.*, extract automatically the region of) their cytoplasm 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 are developing an active contour-based segmentation method relying on local image contrast. The current version still has to be robustified in order to be applicable batchwise (see Fig 7 , right).

5.8. Small Particle Detection

Participants: Nicolas Cedilnik, Xavier Descombes, Eric Debreuve, Florence Besse.

This work is made in collaboration with Fabienne De Graeve (iBV).

One task of the RNAGRIMP project is to detect RNA granules from fluorescent images. These granules have their size close to the image resolution, they typically represent very few pixels. At this scale, shape parametric models are only crude approximations of the object geometry and not adapted for a detection task. To overcome this difficulty we have defined a shape dictionary consisting of all the shapes included in a five by five tile and satisfying some properties of regularity and convexity. Then we are mimicking the marked point process framework by defining an energy function on the finite sets of shapes as the sum of a data term, applied on each object, and a non overlapping constraint between neighboring objects. The solution minimizing the energy is approximated by a greedy algorithm. We have compared different data terms and shown better performances than the traditional threshold approaches and the wavelet based approach as provided by the software Icy.

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

5.9. Inter-individual spatio-temporal registration strategies applied to 3D microscopy image sequences of Arabidopsis floral meristems

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Yassin Refahi (Sainsbury Lab., University of Cambridge), Jan Traas (ENS Lyon) and Christophe Godin (Inria Virtual Plants team, Montpellier).

In developmental biology, the study of model organisms such as the plant *Arabidopsis thaliana* aims at understanding genetic mechanisms responsible of morphogenesis. Today, fluorescent confocal microscopy is a means for *in vivo* imaging of organs of interest such as Arabidopsis floral meristems at cell level with a high spatio-temporal resolution. To handle such 3D+t image sequences, adapted computer-assisted methods are highly desirable. Moreover, the inter-individual development variability quantification requires the ability to register spatio-temporal image sequences from a population of individuals.

In the related work, we propose a dedicated tool for the inter-individual spatio-temporal sequence-to-sequence registration applied to developing Arabidopsis flower meristems. We also discuss the different strategies that may be adopted by the user for the method application in order to assist the choice of parameters for the registration method such as:

- the image primitives to be registered;
- the initialization of the image-to-sequence registration optimization process;
- the initialization "propagation" strategy for sequence-to-sequence registration;
- the parameters selection for the optimization step.

Figure 8 shows the result of the temporal registration between three interpolated image sequences of developing Arabidopsis floral meristems. Figure 9 shows an example of spatial registration between two images from different floral meristems at the same developing stage.

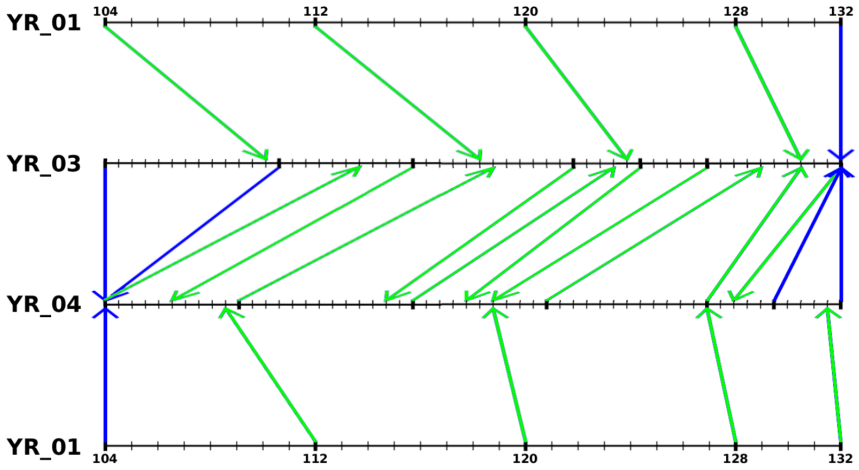


Figure 8. Inter-individual temporal registration result between three floral meristem 3D+t interpolated image sequences. Blue arrows correspond to border registrations.

5.10. Coherent temporal extrapolation of labeled images

Participants: Gaël Michelin, Grégoire Malandain.

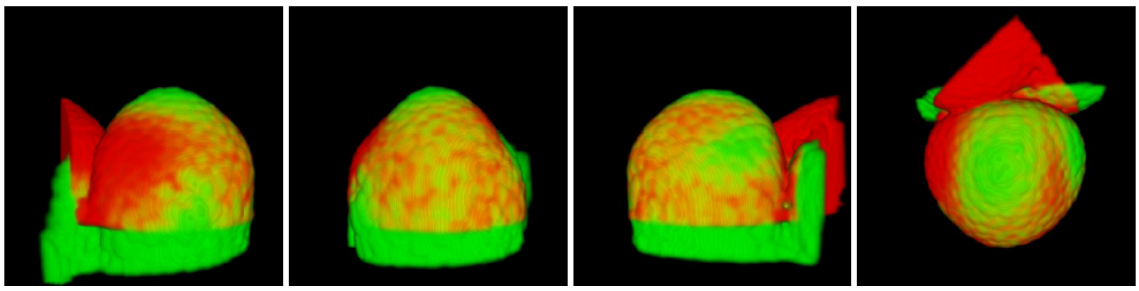


Figure 9. 3D views of the spatial registration of images from two flower meristems (in red and green) at the same developing stage.

In developmental imaging, 3D+t series of microscopic images allows to follow the organism development at the cell level and has become the standard way of imaging the development of living organs. Dedicated tools for cell segmentation in 3D images as well as cell lineage calculation from 3D+t sequences have been proposed to analyze these data. For some applications (such as section 5.9), it may be desirable to interpolate images at intermediary time-points. However, the known methods do not allow to locally handle the topological changes (ie cell. division).

In the present work, we propose an extrapolation method that coherently deformed the images to be interpolated so that to guarantee a topological continuity of borders (see figure 10).

5.11. 3D/2D Coronary Arteries dynamic registration

Participants: Emmanuelle Poulain, Grégoire Malandain.

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

Integrating vessel information, extracted from pre-operative 3D CT angiography images, into a live fluoroscopic 2D sequence can greatly improve the guidance of percutaneous coronary interventions. We are developing a framework aiming at deformed a vessel 3D from the CT so that it moves along the cardiac cycle observed through the 2D angiographic sequence.

The vessel is approximated by a spline which will be deformed thanks to a gradient descent with a length constraint. The length preservation of the vessel allows us to provide a realistic movement, i.e any point will keep its curvilinear abscissa along the spline. This is exemplified by figure 11 where the vessel projection and a remarkable point is tracked at 3 different cardiac phases.

5.12. Modelling axon growth from in vivo data

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

During the first part of this work, we focused at identifying the main morphological features that allow to describe and discriminate genetically different Drosophila Gamma neurons, as well as to automatically assess a quantification of the overall morphological distance between them [8]. The second part, developed this year, approaches the process of neuron growth and morphogenesis in pupal stage. Important advances have already been achieved in identifying the main factors involved in neuron development. The next step that has to be done is concerning how we approach the question.

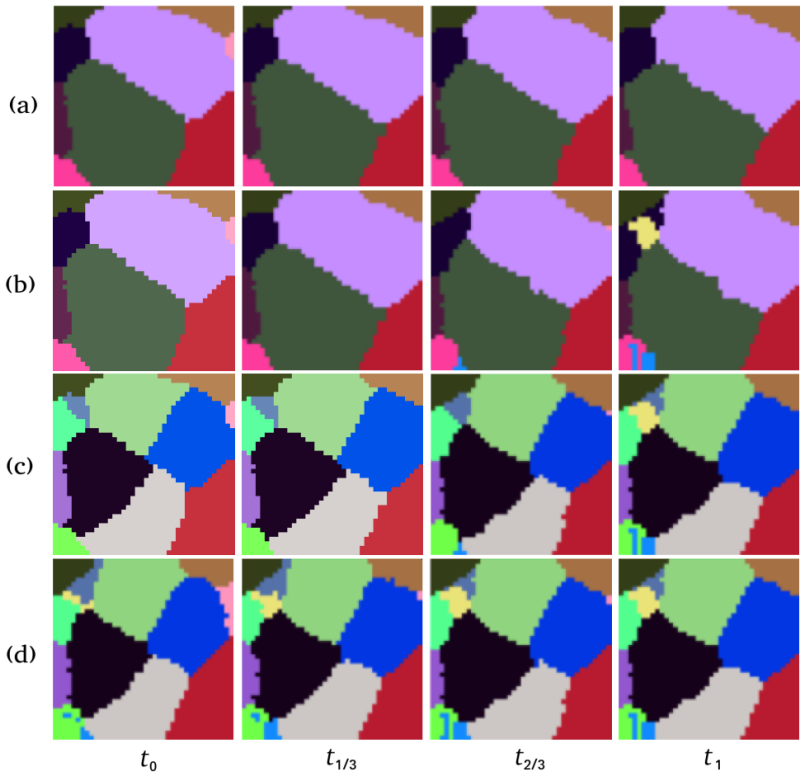


Figure 10. Images compounded of labeled regions at increasing time. (a) and (d) are the original sequences. One can see that the region borders of the corresponding groups of regions of these sequences do not superimpose perfectly. (b) and (c) are the images of transformed regions respectively from the sequences (a) and (d) so that the corresponding region borders superimpose perfectly with the constraint that the image of (a) at t_0 and the image of (d) at t_1 are not modified.

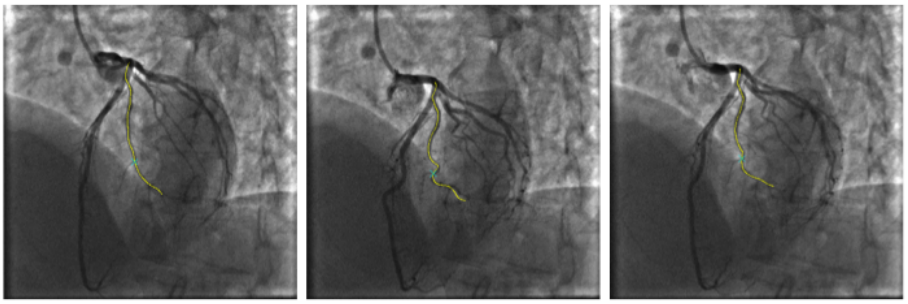


Figure 11. The 3 images show the projection of a 3D vessel in yellow and a remarkable point on the vessel (bifurcation) in blue at 3 different phases of the cardiac cycle.

In this work we intend to close the gap between classic in vitro experimental assumptions and real in vivo situations, where the final neuronal morphology is acquired through a dynamic and environmental-dependent process. In particular, the branch formation process - how or why branches are created - has been belittled or over-simplified by neuron development models. In our opinion, this represents a constraint in the general understanding of neuron development, hierarchy of the neuronal tree and adult functionality.

Our goal is to bring light to the mechanisms of branch formation during development in realistic conditions. We study the particular case of Drosophila Gamma neuron remodeling and analyze, for the first time to our knowledge, the mechanical situation of a whole population of Gamma neurons (650 individuals) growing together in a constraint space (i.e. medial lobe of the Mushroom Body). We hypothesize that one kind of branches are born when the growing tip encounters a mechanical obstacle (i.e. other neurons or the lobe limits), enhancing the probability that at least one neurite reaches the end of the lobe. We model the neurites growth by a Gaussian Markov chain, and the parameters of the model –which account for axon elasticity and guiding cues attractiveness- are estimated from data.

Our database is composed by different sets –wild type and mutations- of confocal images of a single neuron that we treated, segmented and normalized. We show that the proposed mechanistic branch generation process is plausible, and explore unsolved problems concerning the understanding of some particular Gamma neuron mutation phenotypes. This approach allows us also to analyze dynamic aspects of the Gamma neuron collective growth process such as speed and density in function of space and time, which help to explain several characteristics of the Gamma neuron morphology and behavior during development. Figure 12 shows examples of wild type as well as imp mutant neurons of our database and contrasts them with neurons from simulations that are morphologically close.

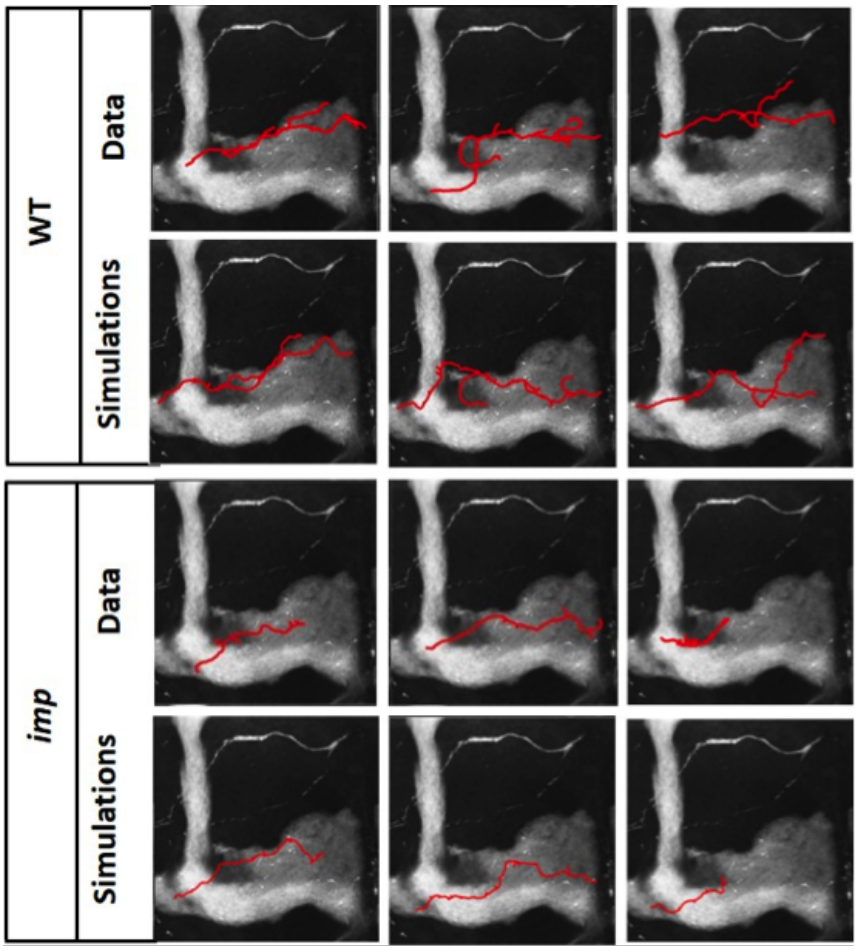


Figure 12. Wild type and imp mutant neurons from the database, contrasted to very similar neurons from our simulations.

PLEIADE Team

7. New Results

7.1. Clavispora lusitaniae genome

Clavispora lusitaniae is an ubiquist environmental 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 onco-haematology patients [24].

So far, two strains have had their genomes sequenced: ATCC 42720, isolated from the blood of a patient with myeloid leukemia [29], and MTCC 1001, a self-fertile strain isolated from citrus [27]. We performed the genome assembly of the *C. lusitaniae* type strain CBS 6936 [37], isolated from citrus peel juice.

Illumina sequences were obtained by our collaborator (T. Noel, UMR 5234 CNRS Université Bordeaux) and we ran the assemblies using several assemblers, e.g. MINIA [21], MIRA [20] and SPAdes [19]. Each assembly gave sequence scaffolds colinear with the already sequence genome of ATCC 42720. However the number of scaffolds varied dramatically in assemblies. SPAdes gave the best results by an order of magnitude (MINIA: 2913, MIRA: 930, SPAdes: 53). This last assembly will serve as a basis for further experiments and SNP detections in mutants strains derived from CBS 6936.

This work is a collaboration between Pleiade team, UMR 5234 CNRS/Université de Bordeaux, and MIAT INRA.

7.2. Elaeis guineensis transcriptome

The mesocarp of the oil palm tree (*Elaeis guineensis* Jacq.) contains oil up to 90% of its dry weight and is the oil richest known vegetal tissue [23]. Our goal is to understand how this tissue achieves this result, in order to increase, by synthetic biology approaches, the yield in oil for crops grown in Europe. In a first milestone, oil palm tree genes relevant for oil synthesis from fatty acids will be transiently expressed in tobacco leaves.

In order to select relevant gene candidates, we performed transcriptome assemblies on high quality Illumina sequences. As an annotated genome is available [34], we performed genome-guided assemblies with TRINITY assembler [26]. The sequence reads (total 300 millions) came from 5 RNA independent isolates from 3 tissues: leaf (2 isolates), kernel (1 isolate) and mesocarp (2 isolates). Transcriptomes coming from duplicate isolates show a good level of overlap as regards the predicted transcripts.

Using expression specificity, abundance and transcript annotation from the genome, we selected genes or transcrpt isoforms candidates, as well as transcription factors. A set of 19 sequences has been retained for expression in tobacco leaves and is under genetic engineering processing.

The 5 transcriptomes were fused into a pan-transcriptome which will be used in another angle of the project. The yield and the composition of oil measured from wild-type palm tree specimens varies dramatically, indicating a high level of bio-diversity. We are currently doing a sampling campaign in Africa, RNA extracts will be sequenced and compared to our pan-genome to serve in variant detection (SNPs) and association genetics studies.

This work is a collaboration between Pleiade team and UMR 5200 CNRS/Université de Bordeaux.

7.3. A Geometric View of Diversity

Diversity may be understood as a set of dissimilarities between objects. The underlying mathematical construction is the notion of distance. Knowing a set of objects, on which pairwise distances can be measured, it is possible to build a Euclidean image of it as a point cloud in a space of relevant dimension. The objects under study are microbial communities, given as a set of reads produced by NGS technologies. Distances between specimen are computed as genetic distances between associated reads (so called amplicon approach). Then, the diversity of a community can be associated with the shape of the point cloud built from such distances. Such an embedding is classically implemented by MDS (Multidimensional Scaling). Such an approach triggers two methodological questions, addressed in 2016:

- the numerical solution is through finding the eigenvectors and eigenvalues of a large, full, symmetric matrix. Current algorithm, parallelized or not, are in complexity $\mathcal{O}(n^3)$ if n is the number of specimen on which to study patterns of diversity, i.e. the size of the matrix. This is not feasible for dataset produced by NGS technologies, which can assemble 10^5 to 10^6 sequences. We have set up a collaboration with a PhD student in HIEPACS Team (Inria Bordeaux SO), Pierre Blanchard, to develop a connection between random projection methods and MDS. Random Projection Methods are methods relying on Johnson-Lindenstrauss Lemma, which states that the likelihood that the distances are very well conserved is very high when projecting a point cloud in a space of very large dimension (say n) to a random space of large dimension (say, proportional to $\mathrm{Log}\, n$) . This permits to compute eigenvectors and eigenvelaues in space of reasonable dimension. The method for MDS has been studied by Pierre Blanchard, under supervision by Olivier Coulaud and Alain Franc, and proved to be surprisingly efficient and precise. This work builds one chapter of the PhD thesis of P. B. to be defended by early 2017. This collaboration has lead to a joint poster at Platform for Advanced Scientific Computing (PASC), June 2016, Lausanne, Switzerland.
- The eigenvalues of the matrix under study can be positive or negative. Positive eigenvalues lead to Euclidean structure behind MDS. Classically, negative eigenvalues are ignored. We have begun a study on the role of negative eigenvalues in the discrepancy between Euclidean distances computed between points in MDS, and genetic distances between reads produced by NGS, which adds to the well understood discrepancy in MDS due to dimension reduction. This has lead to three seminars or presentations:
  - a seminar at MIAT research Unit, Toulouse, on February 19
  - a seminar at LABRI on April, 28
  - a presentation at the days of mathematics and computing sciences organized by MIA INRA division (INRA global meeting), on October, 5

  These three events have permitted to “polish” the analysis of the problem through several exchanges, and to orientate its study towards quadratic embedding, or isometry into pseudo-euclidean spaces.

7.4. Topological Data Analysis

Leyla Mirvakhabova has defended and obtained her BSc memoir on "Distance geometry and biodiversity patterns" at the Department of Mathematics of the National Research University - Higher School of Economics, Moscow. Here is the abstract: In this work, we study the biodiversity of the tree species in French Guiana. We consider the matrix of the pairwise genetic distances between the 1501 species. The distances are measured by using the Smith-Waterman algorithm applied to the trnH regions - the chloroplasts of trees. The aim of the project is to analyze the shape of a point cloud in a Euclidean space built from the pairwise distances. To study the structure of the point cloud built from the given distances, we have considered the following methods: Hierarchical Clustering, Multidimensional Scaling (MDS), Nonlinear Mapping (NLM), t-Distributed Stochastic Neighbor Embedding (t-SNE), and Topological Data Analysis (TDA). For the first four methods, we used the Python package scikit-learn 0.17.1 implementations and have written the program for the TDA algorithm. All of these methods were tested on the given dataset. This work has been performed as part of a collaborative research project of the PLEIADE team in the Inria Bordeaux – Sud-Ouest (supervisor Alain Franc) with

the Laboratory of System Biology and Computational Genetics in the Vavilov Institute of General Genetics (supervisor Ivan Kulakovskiy).

7.5. Bespoke tools for comparative genomics

Large-scale comparison of strains of cell factory species is an indispensible tool for understanding the genetic origin of phenotypic variability, and can considerably optimize the selection and construction of high-performing industrial strains. For example, in oenological applications new strains may be selected based on their influence on aroma, their adaptation to grape musts, or their robustness during fermentation. In oil production applications, new strains may be selected based on their yield, or on the saturation degree of the lipids, or on their growth characteristics. Comparative genomics has proven quite effective in understanding cell factory diversity [1], [6], [5], [36], [31]. A typical project will involve 500 segregants and 50 genomes. Accurate and rapid analysis of the concomitant data volumes requires efficient tool sets that must be adapted to the real use cases of the industrial application.

PLEIADE addresses this problem through the definition of bespoke software systems that associate integrated sets of tools, including its Magus software platform (section 6.1). A key challenge in defining this kind of integrated system is the need to connect the components. We develop configuration formalisms whose solutions are orchestrations of weakly-coupled microservices running in independant containers. These services may be data banks, genome browser and visualization software, workflow tools like Galaxy, machine learning algorithms for classification, or shared workbooks like Jupyter or Zeppelin. By formalizing the connections between services, we can simplify deployment, and also create an opportunity for *continuous deployment*.

SERPICO Project-Team

7. New Results

7.1. Statistical aggregation methods for image denoising and estimation

Participants: Charles Kervrann, Frédéric Lavancier.

In the line of the Non-Local means [43] and ND-SAFIR [9], [10], [5] 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 that our algorithm improves significantly the classical NL-means, and is competitive when compared to the more sophisticated NL-means filters both in terms of PSNR values and visual quality.

Previously, we investigated statistical aggregation methods which optimally combine several estimators to produce a boosted solution [11]. In this range of work, we also introduced in [24] a general method to combine estimators in order to produce a better estimate. From a theoretical point of view, we proved that this method is optimal in some sense. It is illustrated on standard statistical problems in parametric and semi-parametric models where the averaging estimator outperforms the initial estimators in most cases. This method has been subsequently adapted in [39] to models in spatial statistics. As part of an on-going work, we are applying this method to improve patch-based image denoising algorithms.

References: [24] [39]

Collaborators: Qiyu Jin (School of Mathematical Science, Inner Mongolia University, China),
Ion Grama and Quansheng Liu (University of Bretagne-Sud, Vannes),
Paul Rochet (Laboratoire de Mathématiques Jean Leray (LMJL), University of Nantes).

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

Participants: Hoai Nam Nguyen, Giovanni Petrazzuoli, Aminata Diouf, Charles Kervrann.

In fluorescence microscopy, the image quality is limited by out-of-focus blur and high noise. Traditionally, image deconvolution is needed to estimate a good quality version of the observed image. The result of deconvolution depends heavily on the choice of the regularization term and the noise dependent fidelity term. The regularization functional should be designed to remove noise while preserving image discontinuities. Accordingly, we investigated new regularization terms to preserve fine details of underlying structures and we studied appropriate proximal algorithms. The deconvolution method has been especially dedicated to large 2D 20000×60000 images acquired with ISO scan imager (see Fig.3). The images are preliminary pre-processed to compensate non constant pixel displacement during acquisition/scanning (dejittering effect). The method has also been evaluated on 2D Vimentin filament images (UTSW, CytoDI Associated Team) to facilitate filament segmentation. The method is able to process a 512×512 image in 250 ms with a non optimized implementation.

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys company),
Philippe Roudot (UTSW, Dallas, USA).

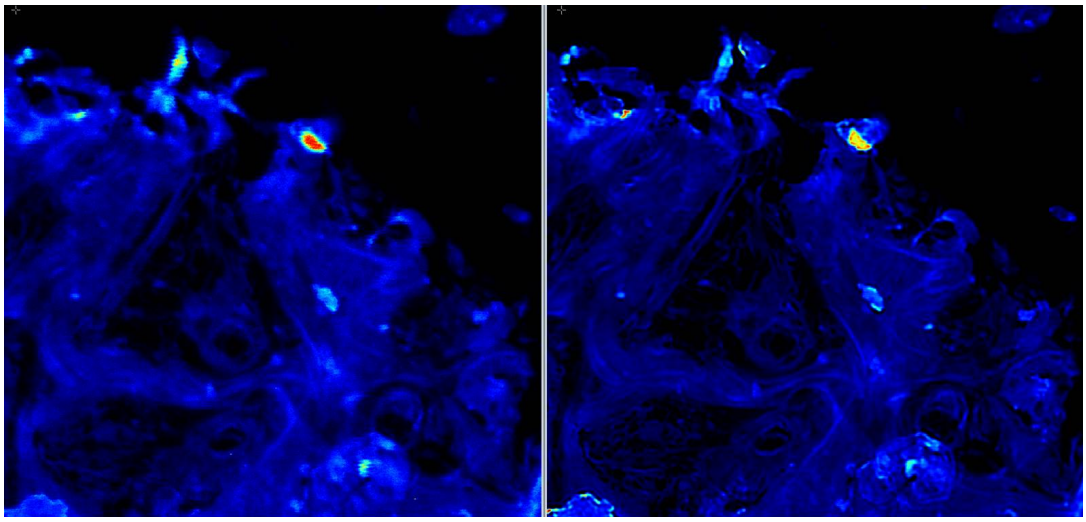


Figure 3. Illustration of image dejittering and deconvolution algorithms applied to a selected fluorescence region extracted from a large TMA image (by courtesy of Innopsys).

7.3. Quantifying the spatial distribution of intracellular events

Participants: Thierry Pécot, Charles Kervrann.

Automated processing of fluorescence microscopy data allows quantifying cell phenotypes in an objective and reproducible way. However, most computational methods are based on the complex combination of heterogeneous features such as statistical, geometrical, morphological and frequency properties, which makes difficult to draw definitive biological conclusions. Additionally, most experimental designs, especially at single cell level, pool together data coming from replicated experiments of a given condition, neglecting the biological variability between individual cells. To address these issues, we developed a generic and non-parametric framework (QuantEv) to study the spatio-temporal distribution of moving Rab6 membranes and the effect of actin disruption on Rab11 trafficking in coordination with cell shape. The main advantage of QuantEv is to process robustly and accurately homogeneous and heterogeneous populations. As demonstration, we compared the results obtained by QuantEv with those from kernel density maps, for Rab6 positive membranes on crossbow- and disk-shaped cells.

Collaborators: Jean Salamero, Jérôme Boulanger and Liu Zengzhen (UMR 144 CNRS-Institut Curie).

7.4. 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.

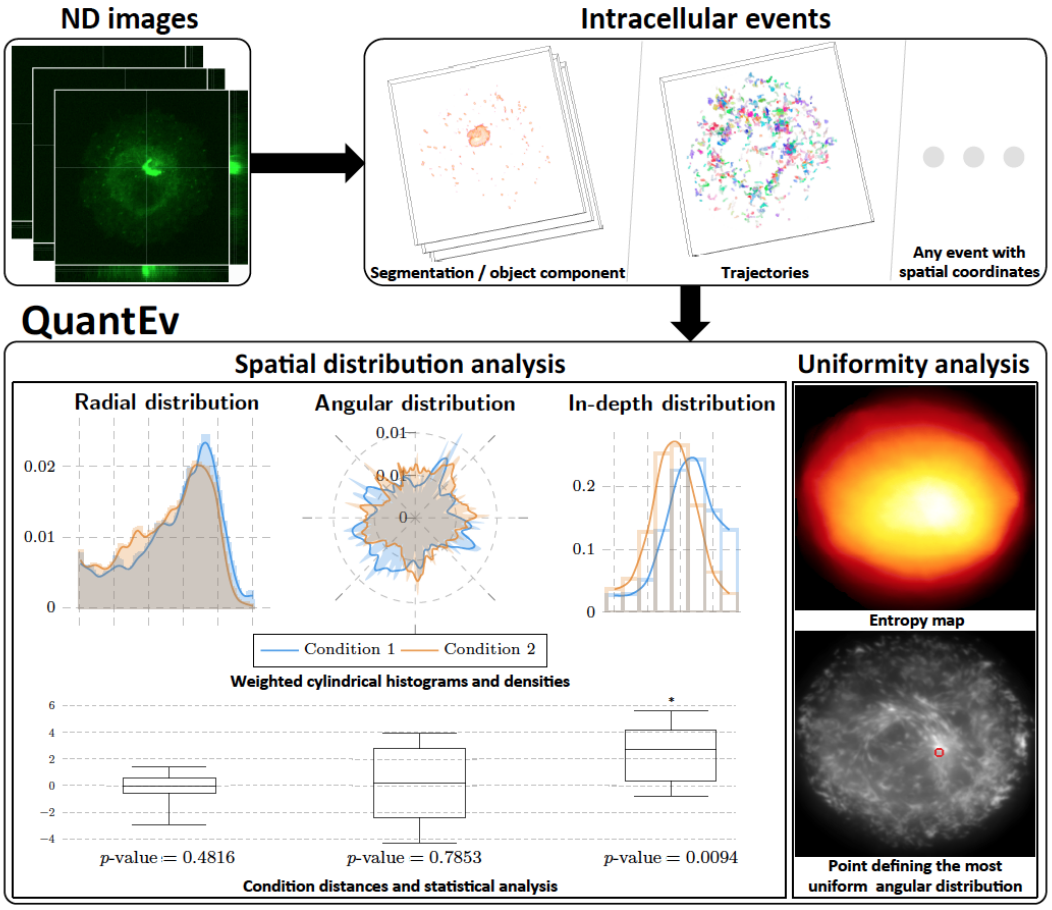


Figure 4. Overview of QuantEv approach.

This year, we have continued our study of correlation-based methods for local diffusion estimation in TIRFM images. Our original method was tested on both synthetic and real images showing an isolated diffusion event, and a robust algorithm was developed to cope with noisy data. Our first model was linear and had only two parameters to estimate. Diffusion coefficient estimation was accurate on synthetic images even with moderate to low signal-to-noise ratio, and within reasonable margins of error on real images with little noise. We have then extended our mathematical model by using a global approach subject to initial local diffusion conditions. Isolated diffusion events are well described, but this new model can also handle the case of noisy images with non-uniform background, and the case of two or more diffusion events in the region of interest. The extended model is non-linear but has few parameters to estimate. An iterative minimization method is used to fit the model parameters to the data points (see Fig. 5). Despite non-linearity, results are accurate on images with pure diffusion events and show robustness to background. The quality of parameters estimation is barely influenced by the length and size of the input TIRFM sequence, which is not the case with standard correlation methods. We have thus developed a correlation-based method that is able to estimate diffusion in a variety of cases in TIRFM images (Fig. 5).

Collaborators: Francois Waharte (UMR 144 CNRS-Institut Curie, PICT-IBiSA),
Perrine Paul-Gilloteaux (UMS 3556, IRS-UN, Nantes).

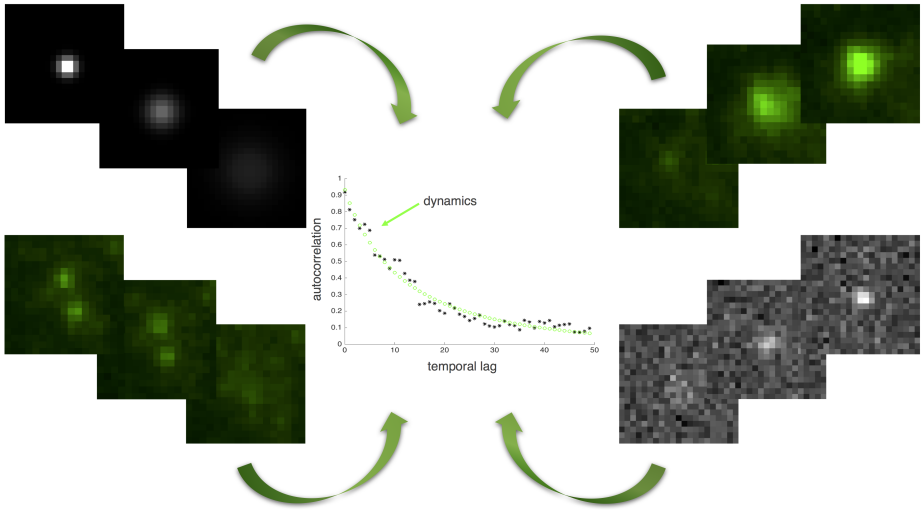


Figure 5. Local diffusion estimation in both synthetic (black and white spots) and real TIRF images (green spots, UMR 144 CNRS-Institut Curie, PICT-IBiSA). Only three frames of each sequence are shown. The computed autocorrelation values (in black) for different temporal lag values, and the fitting function for these values (in green) are given (plot in the middle) for the upper-right sequence.

7.5. Colocalization for live cell and super-resolution imaging

Participants: Frédéric Lavancier, Thierry Pécot, Charles Kervrann.

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. In statistics, this amounts to characterizing the joint spatial repartition and the spatial overlap between different fluorescent labels. Two distinct categories of colocalization approaches are considered to address this issue: intensity-based methods and object-based methods. The popular (intensity-based) Pearson's correlation method, which returns values between -1 and +1, is known to be sensitive to high intensity backgrounds and provides errors if the signal-to-noise ratio (SNR) is typically low. The object-based method, recommended in single molecule imaging, analyses the spatial distribution of the two sets of detected spots by using point process statistics.

Accordingly, we developed an original, fast, robust-to-noise and versatile approach that reconciles intensity-based and object-based methods for both conventional diffraction-limited microscopy and sub-resolved microscopy. The procedure is only controlled by a p-value and tests whether the Pearson correlation between two binary images is significantly positive. This amount to quantifying the interaction strength by the area/volume of the intersection between the two binary images viewed as random distributions of geometrical objects. Under mild assumptions, it turns out that the appropriately normalized Pearson correlation follows a standard normal distribution under the null hypothesis if the number of image pixels is large. Unlike previous methods, the method handles 2D and 3D images, variable SNRs and any kind of cell shapes. It is able to colocalize large regions with small dots, as it is the case in TIRF-PALM experiments and to detect negative colocalization. The typical processing time is two milliseconds per image pair in 2D and a few seconds in 3D, with no dependence on the number of objects per image. Finally, the method provides maps to geocolocalize molecule interactions in specific image regions.

Collaborators: Jean Salamero and Liu Zengzhen (UMR 144 CNRS-Institut Curie).

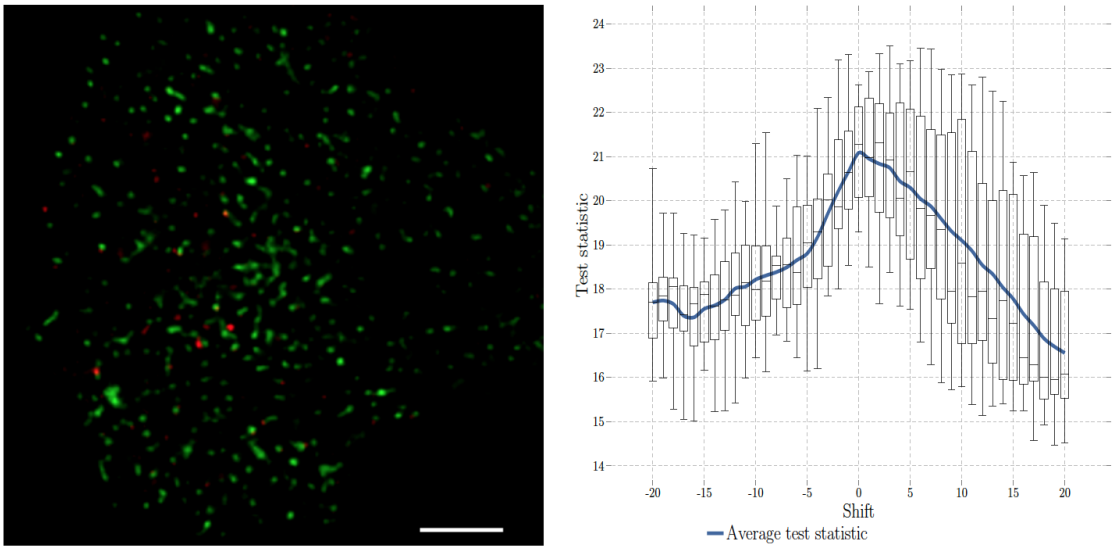


Figure 6. Illustration of statistical spatio-temporal colocalization. a) Average intensity projection of a 3D TIRF acquisition showing m-Cherry Langerin (red channel) and GFP Rab11A proteins (green channel) (courtesy of UMR 144 CNRS-Institut Curie); b) test statistic value with respect to the shift between frames.

7.6. Classification of diffusion dynamics from particle trajectories

Participants: Vincent Briane, Charles Kervrann.

In this study, we are currently interested in describing the dynamics of particles inside live cell. We assume that the motions of particles follow a certain class of random process: the diffusion processes. In 2015, we developed a statistical test to classify the intracellular motions into three groups : free diffusion (*i-e* Brownian motion), subdiffusion and superdiffusion. This method is an alternative to the commonly used Mean Square Displacement (MSD) analysis. This year, we have studied theoretical properties of our procedure. We have shown that it behaves well asymptotically, that is when we observe the particle trajectory for a very long time, for certain parametric models. The models on which we assess our procedure are representative of the three classes aforementioned and extensively used in the literature. Among them we can cite Brownian motion with drift, Ornstein-Uhlenbeck process and fractional Brownian motion. An illustration of the testing procedure is shown in Fig. 7 .

We also extend our method to address two different questions. First, we are interested in testing a large number of trajectories. The first version of our test is a single test procedure. It is known that applying multiple times a test without care leads to a high number of false positives. Then, we modify our initial method to overcome this problem. Secondly, in the case in which we observe very long trajectories, it is likely that the particle motion changes over time. Therefore, we are currently adapting our initial procedure to detect change-point along a single trajectory.

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

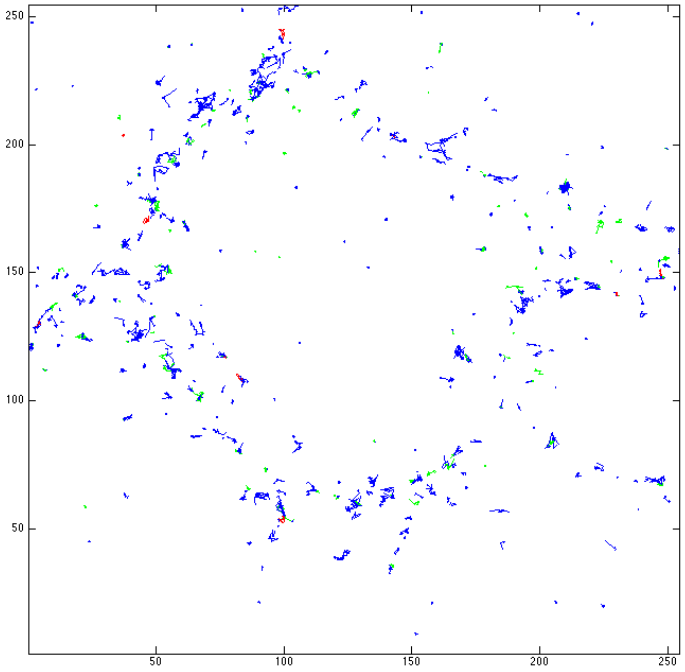


Figure 7. Labelling of the dynamics of trajectories in Single Particle Tracking PALM (Courtesy of Institut Interdisciplinaire de Neurosciences CNRS UMR 5297). The color code is green for subdiffusion, blue for Brownian motion and red for superdiffusion.

7.7. Inference for spatial Gibbs point processes

Participant: Frédéric Lavancier.

Gibbs point processes are popular and widely used models in spatial statistics to describe the repartition of points or geometrical structures in space. They initially arose from statistical physics where they are models for interacting particles. They are now used in as different domains as astronomy, biology, computer science, ecology, forestry, image analysis and materials science.

Assuming a parametric form of the Gibbs interaction, the natural method to estimate the parameters is likelihood inference. Since its first use in the 80's, this method is conjectured to be consistent and efficient. However the theoretical properties of maximum likelihood for Gibbs point processes remain largely unknown. In [17], we partly solved this 30 years old conjecture by proving the consistency of the likelihood procedure for a large class of Gibbs models. As important examples, we deduce the consistency of the maximum likelihood estimator for all parameters of the Strauss model, the hardcore Strauss model, the Lennard-Jones model and the area-interaction model, which are commonly used models in practice.

A practical issue of likelihood estimation yet is that this method depends on an intractable normalizing constant that has to be approximated by simulation. To avoid this problem, other methods of estimation have been introduced, including pseudo-likelihood estimation. The theoretical properties of the pseudo-likelihood method are fairly well known in the case of finite-range Gibbs interactions. However this setting rules out some major Gibbs model as the Lennard-Jones model. In [16], we extend the pseudo-likelihood procedure to infinite range Gibbs interactions and we prove its consistency and its asymptotic normality.

References: [16] [17]

Collaborators: David Dereudre (Laboratoire Paul Painlevé (UMR 8524), University of Lille 1),
Jean-François Coeurjolly (Laboratoire Jean Kutzmann, University of Grenoble).

7.8. Statistical aspects of Determinantal Point Processes

Participant: Frédéric Lavancier.

Determinantal point processes (DPPs) have been introduced in their general form by Macchi (1975) and have been extensively studied from a probabilistic point of view in the 2000's (one of the main reason being their central role in random matrix theory). In a previous work, we have demonstrated that DPPs provide useful models for the description of spatial point pattern datasets where nearby points repel each other.

In [15], we have addressed the question of how repulsive a stationary DPP can be, in order to assess the range of practical situations this promising class of models may model. We determine the most repulsive DPP (in some sense) and we introduce new parametric families of stationary DPPs that can cover a large range of DPPs, from the stationary Poisson process (the case of no interaction) to the most repulsive DPP. Some theoretical aspects of inference for stationary DPPs are tackled in [13] and [14]. In the former study we establish the Brillinger mixing property of stationary DPPs, a first important step toward asymptotic inference. In the latter contribution, we exploit this result to deduce the consistency and asymptotic properties of contrast estimators for stationary DPPs.

References: [13] [15] [14]

Collaborators: Christophe Ange Napoléon Biscio (LMJL, University of Nantes),
Jesper Möller (Department of Mathematical Sciences, Aalborg University, Denmark),
Ege Rubak (Department of Mathematical Sciences, Aalborg University, Denmark).

7.9. Modeling aggregation and regularity in spatial point pattern datasets

Participant: Frédéric Lavancier.

In the spatial point process literature, analysis of spatial point pattern datasets are often classified into three main cases: (i) Regularity (or inhibition or repulsiveness), modelled by Gibbs point processes, hard core processes like Matern hard core models, and determinantal point processes; (ii) Complete spatial randomness, modelled by Poisson point processes; (iii) Aggregation (or clustering), modelled by Poisson cluster processes and Cox processes. For applications the classification (i)-(iii) can be too simplistic, and there is a lack of useful spatial point process models with, loosely speaking, aggregation on the large scale and regularity on the small scale. For instance, we may be interested in such a model for the repartition of the centres of vesicles in a cell, that exhibit some spatial clustering at large scales while having a minimal distance between them.

In [23], we have considered a dependent thinning of a regular point process with the aim of obtaining aggregation on the large scale and regularity on the small scale in the resulting target point process of retained points. Various parametric models for the underlying processes are suggested and the properties of the target point process are studied. Simulation and inference procedures are discussed when a realization of the target point process is observed, depending on whether the thinned points are also observed or not.

Reference: [23]
Collaborator: Jesper Möller (Department of Mathematical Sciences, Aalborg University, Denmark).

7.10. Multi-scale spot segmentation with automatic selection of image scales

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy.

Detecting spot-like objects of different sizes in images is required for many applications. A spot detection framework can be divided in three sub-steps : first, image preprocessing to smooth out noise; second, signal enhancement to highlight spots; third, spot detection by thresholding; the two first ones being often merged in a single operator. However, elements of interest do not all correspond to the same image scale, if the collection includes subgroups of different sizes or if perspective effects occur. Then, the need is not merely the selection of the optimal image scale, but of all the meaningful scales. We dealt with the problem of multi-scale spot detection while automatically selecting the meaningful scales. Our primary interest is to detect particles in microscopy images, but our method can be applied to other types of images as well. We defined an original criterion based on the a contrario approach and the LoG scale-space framework to automatically select the meaningful scales. We designed a coarse-to-fine multi-scale spot segmentation scheme involving a locally adaptive thresholding across scales, to come up with the final map of segmented spots. We carried out experimental results on simulated and real images of different types, and we demonstrated that our method outperforms other existing methods.

Reference: paper accepted, ICASSP'2017.
Collaborator: Antoine Basset (CNES, Toulouse).

7.11. Multi-modal registration for correlative light-electron microscopy

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy, Charles Kervrann.

We pursue our work on correlative light-electron microscopy (CLEM), which combines the strengths of two different imaging modalities, light microscopy (LM) and electron microscopy (EM), to jointly study intracellular dynamics and ultrastructure of a biological sample. CLEM registration is an important and difficult problem given the significant differences between LM and EM images regarding resolution, field of view, image size and appearance. We designed an automated approach for retracing and registering CLEM images, by implementing a patch-based search using a common Laplacian of Gaussian (LoG) image representation of the LM and EM images. We have redefined the geometry of the patch, opting for a disk-shaped patch. The search (or retracing) step uses histogram-based methods as they are invariant to rotation, and it provides a pre-registration by producing the estimate of the translation component. Usually, there is a large disparity on the orientation of EM and LM images. To handle this problem, we have implemented a mutual information-based method to compute the rotation between the EM and LM patches and to refine the registration. We have also tackled the registration issue in both directions (LM to EM, and EM to LM), and compared our approach to a correlation-based method.

We have tested our approach on a larger set of real CLEM images (provided by Institut Curie) presenting a large diversity in content, image size, and appearance, further validating our method (see Fig. 8). We are currently exploring how our automated CLEM registration method could be exploited to guide EM acquisition within a coarse-to-fine framework.

Reference: [35]
Collaborators: Xavier Heiligenstein (UMR 144 CNRS-Institut Curie),
Grégoire Malandain (Inria, Morpheme EPC, Sophia-Antipolis).

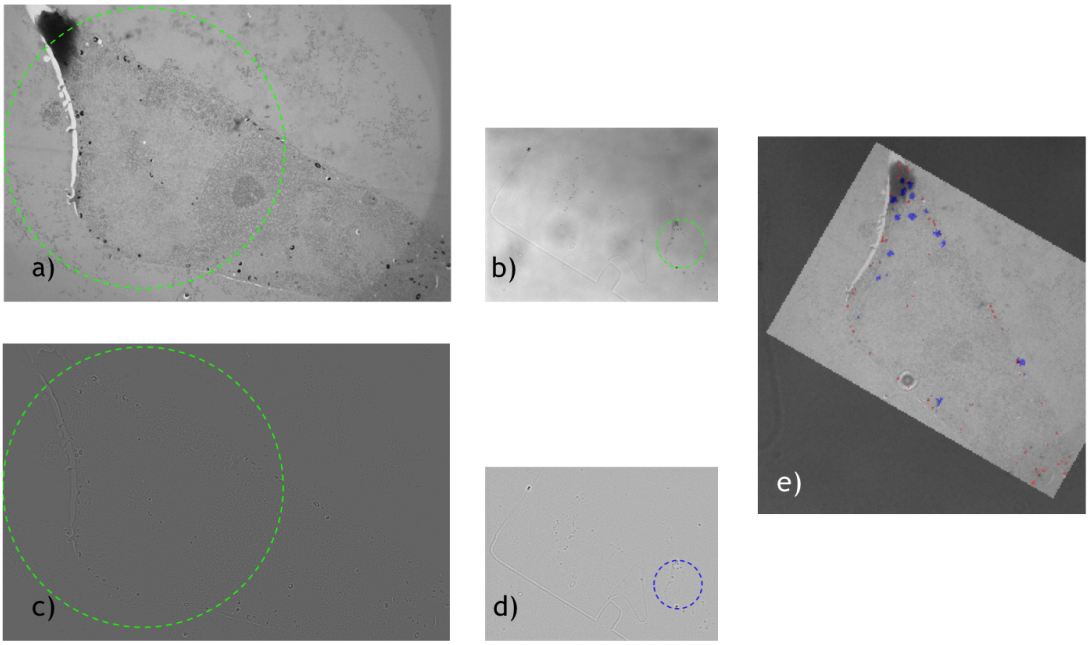


Figure 8. Figure 1. CLEM experiment (images from UMR 144 CNRS-Institut Curie, Xavier Heiligenstein): a) EM image with Region of Interest (ROI) framed in green; b) ground-truth location of the corresponding LM patch framed in green; c) ROI delineated in the Log-EM image, framed in green; d) selected patch (SP) in the LoG-LM image, framed in blue to be comaprd with the green disk in b); f) overlay (after registration).

7.12. Denoising and compensation of the missing wedge in cryo electron tomography

Participants: Emmanuel Moebel, Charles Kervrann.

In this study, we address two important issues in cryo electron tomography (CET) images: reduction of noise and restoration of information in the missing wedge (MW). The MW is responsible for several type 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. 9). 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.

Several aspects of the method have been studied in order to gain a deeper understanding of this strategy: different kinds of noise as well as several denoising algorithms (BM3D, NL-Bayes, NL-means, Total Variation...) have been evaluated. Furthermore, different kinds of transforms have been tested in order to apply the constraint (Fourier transform, Cosine transform, pseudo-polar Fourier transform). Also, a process has been set up in order to evaluate the performance of the proposed method on experimental data. Thus, convincing results on experimental data have been achieved (see Fig. 9) using the Fourier Shell Correlation (FSC) as an evaluation metric. In order to measure the quality of the recovered MW only, we also compute the FSC over the MW support ("constrained FSC").

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

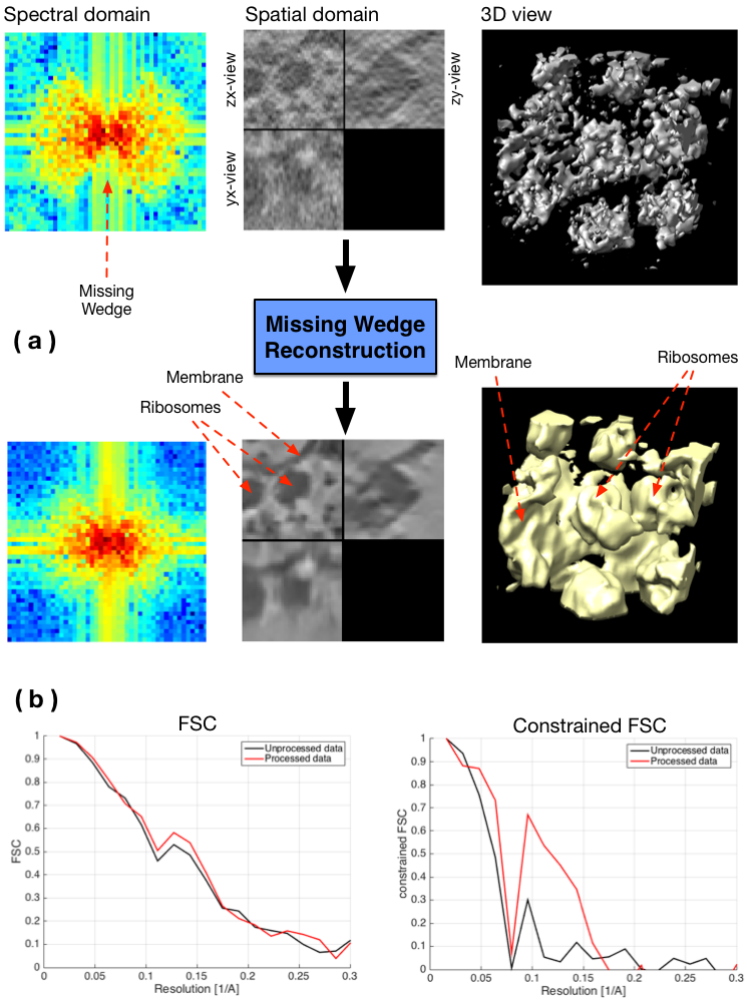


Figure 9. Experimental sub-tomogram containing ribosomes attached to a membrane. (a) Top row: original data in the spectral (left) and spatial (middle) domains and 3D view of the thresholded data (right). Bottom row: denoised data shown as above. (b) FSC and constrained FSC measures of the method input (in black) and output (in red). All measures are wrt the same reference.

7.13. Spatially-variant kernel for optical flow under low signal-to-noise ratios

Participant: Charles Kervrann.

Local and global approaches can be identified as the two main classes of optical flow estimation methods. This year, we have proposed a framework to combine the advantages of these two principles, namely robustness to noise of the local approach and discontinuity preservation of the global approach. The idea is to adapt spatially the local support of the local parametric constraint in the combined local-global model [42]. To this end, we jointly estimate the motion field and the parameters of the spatial support. We apply our approach to the case of Gaussian filtering, and we derive efficient minimization schemes for usual data terms. The estimation of a spatially varying standard deviation map prevents from the smoothing of motion discontinuities, while ensuring robustness to noise. We validated our method for a standard model and demonstrated how a baseline approach with pixel-wise data term can be improved when integrated in our framework. The method has been evaluated on the Middlebury benchmark with ground truth and on real fluorescence microscopy data for which noise is the main limitation for usual optical flow methods.

Collaborator: Denis Fortun (EPFL-BIG, Lausanne, Switzerland)
Noémie Debroux (Laboratory of Mathematics, INSA Rouen, Normandie University)

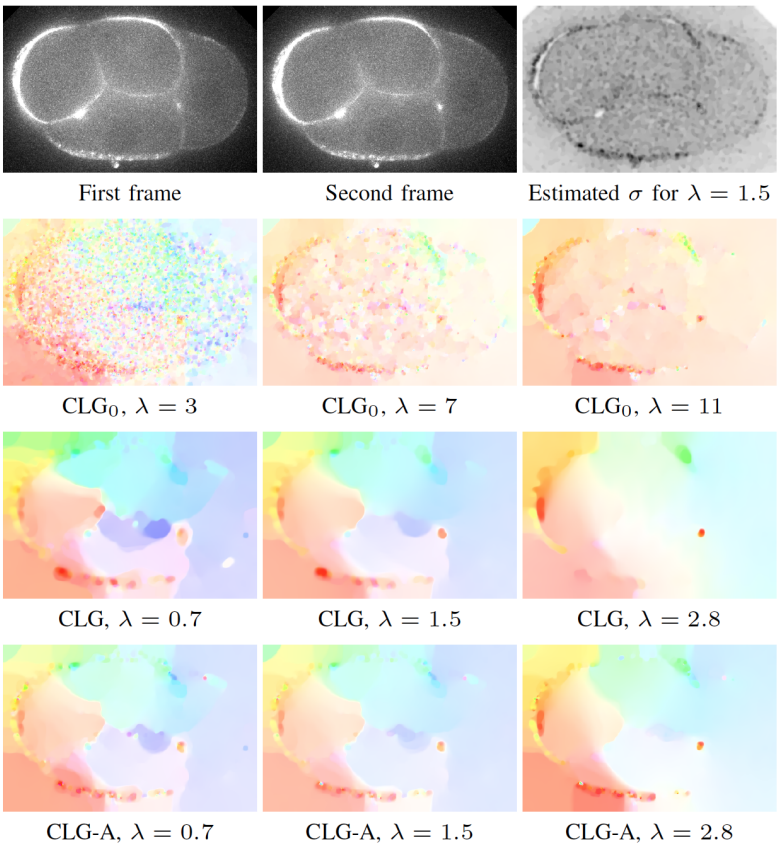


Figure 10. Visual results obtained with variants of CLG ("Combined Local-Global") methods for several values of regularization parameters λ: CLG-A (our adaptive CLG), CLG_0 (pointwise method) and CLG ([42]) on a image pair from a CElegans sequence.

7.14. Frame-based hierarchical motion decomposition and segmentation

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

A number of applications in video analysis rely on a per-frame motion segmentation of the scene as key preprocessing step. Moreover, different settings in video production require extracting segmentation masks of multiple moving objects and object parts in a hierarchical fashion. In order to tackle this problem, we propose to analyze and exploit the compositional structure of scene motion to provide a segmentation which is not purely driven by local image information. Specifically, we leveraged a hierarchical motion-based partition of the scene to capture a mid-level understanding of the dynamic video content. To recover the decomposition tree, we formulated the problem as a per-pixel label selection interleaved with motion models estimation. The labels represent the set of nodes from the initial proposal tree which are selected to explain globally the input correspondence field. We carried out experimental results showing the strengths of this approach in comparison to current video segmentation approaches. Indeed, they demonstrated the superior ability of our method to capture the main moving objects of the scene in the first layer of the tree, and to segment them in moving parts in deeper layers. As such, we believe our segmentation method is closer to the complex needs of video editing than current hierarchical segmentation approaches.

Reference: [34]
Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

7.15. Trajectory-based discovery of motion hierarchies in video sequences

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

The dynamic content of physical scenes is largely compositional, that is, the movements of the objects and of their parts are hierarchically organized and relate through composition along this hierarchy. This structure also prevails in the apparent 2D motion that a video captures. Visual motion in the scene is roughly organized along a tree, with the dominant motion (typically induced by camera motion) at the root, and motion components adding up along the branches. Accessing this visual motion hierarchy is important to get a better understanding of dynamic scenes and is useful for video manipulation. We proposed to capture it through learned, tree-structured sparse coding of point trajectories. We found that dictionary learning and sparse coding provide appealing tools to disentangle this latent hierarchical structure. More precisely, we introduced a new tree-structured dictionary learning method that allows describing each track with a few basis functions, all but one being inherited from its parent in the structure. The sparse codes thus associated to the tracks capture the desired structure and lend themselves naturally to hierarchical clustering of the collection. We showed through experiments on motion capture data that our model is able to extract moving segments along with their organization. We also obtained competitive results on the task of segmenting objects in real video sequences from trajectories.

Reference: [33]
Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

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, Ibrahim Chedaddi, Mathilde Balduzzi, Julien Diener.

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (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 [81], [68]. 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.* (Olivier Simler [AFEF, AGAP], Chakkrit Preuksakarn, 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 [78]. 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 [59].

  This year, the editor has been augmented for better user control over the different step of the reconstruction process. Some first alignment procedures of scans and reconstructions made at different times of the year have been also implemented. An application for the reconstruction of an apple tree core collection has been conducted during the internship of O. Simler in a collaboration with the AFEF Team of UMR AGAP.

- *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 ADEL), 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. This software framework was presented to “BMVA technical meeting: Plants in Computer Vision”.

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

  This research theme is supported by the Agropolis project MecaFruit3D.

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function. To this end, a modeling pipeline has been developed in the OpenAlea platform. It involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex vascular network that is embedded within this mesh using the concept of space colonization [80]. Previous studies demonstrated the possibility to create species-specific models of fruit structure with relatively low effort [63]. We focus now on validating the vascular networks by comparing them to experimental data from the literature. This work has been presented at the ISHS symposium in Montpellier [60] and resulted in a publication [17].

  Using these fruit virtual structures, a mechanical model of fruit growth is also developed (see section 6.3.2) taking into account the distribution of water fluxes in the fruit.

- *Review on morphological plant modelling.* (Christophe Pradal, Mathilde Balduzzi, Alexander Bucksch [Georgia Univ., USA], Daniel H. Chitwood [Donald Danforth Plant Science Center, 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 [45], 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.

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. [54], [70], [71], [67], 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" [75], "age state" [66] or "physiological age" [56]. 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 [66], [56]. 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 [25] 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.

- *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 [64]. 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). Two replications of each genotype were available.

Indices were proposed for early detection of alternation during the juvenile phase. They were based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. This allowed early quantification of alternation from the yearly numbers of inflorescences at tree scale. Some quantitative trait loci (QTL) were found in relation with this indices.

For better interpretation of the relationships of alternation at both scales, new models and indices were developed for sequences of flowering events at axis scale. 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, and other QTL were found using these new indices.

- *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]).

  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. This statistical modeling framework was applied to young mango trees [32].

- *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 [73], [74]. The global aim is to have a crop simulation model to predict fruit yield and quality on mango tree. An overview of this global model based on the coupling of different structural or ecophysiological sub-models has been also presented in the FSPMA conference [44].

  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 perpetual flowering strawberry genotypes, which is of particular importance for better predicting fruit production. We applied multiple change-point models for the synchronous segmentation of the individuals of a given genotype in successive flowering phases [24]. We identified two groups of genotypes that differ by the intensity of the flowering at the end of the flowering period. Using a genetic approach, we identified a locus

Figure 1. Simulation of the development of a mango tree over two cycles [58]. 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.

controlling the flowering intensity at the end of the flowering period that likely explain these two groups of genotypes. A multivariate generalization of the synchronous segmentation approach is developed in the context of Marc Labadie's PhD [50], 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 [6], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG (Directed Acyclic Graph, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree.

  The method proposed in [6] thus compresses a tree in width, but not in height. In a new work, we designed an extension of this compression method in which a tree is compressed in both width and height. The method is based on the detection of so-called *quasi-isomorphic paths* in a tree and on the compression of these paths in height. A paper describing the corresponding algorithms has been recently accepted in the Journal of Theoretical Biology [16].

  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. [42]

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], Bruno Andrieu [Ecosys, INRA], 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. This year, we worked on four application studies that used the framework. In the frame of the PhD of Guillaume Garin, we perform a validation of the wheat septoria model, an analysis of the influence of the wheat architecture on the competition between septoria and brown rust, and a sensitivity analysis of the response of the severity of septoria to architectural traits. In the frame of the Echap project, we use the wheat-septoria model to indentify optimal date of pesticide application. All these studies allows 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] 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 [30], 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), 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 two-scale growth pattern with (i) at coarse scale, roughly stationary growth phases with phase changes likely corresponding to cold seasons 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 specific multiple change-point models (piecewise 1st-order autoregressive models) for characterizing this two-scale growth pattern. The objective will be now to study the modulation of this pattern in response to different water stress and fertilization levels.

- *Quantifying the impact of water deficit on the production and flowering of apple trees* (Jean-Baptiste Durand, Benoit Pallas [AGAP, AFEF team], Evelyne Costes [AGAP, AFEF team])

  Water stress generates a number of physiological and morphological responses in plants that depend on the intensity and duration of stress as well as the plant species and development stage. In perennial plants, WS may affect plant development through cumulative effects that modify plant functions, architecture and production over time. Plant architecture depends on the fate of the terminal and axillary buds that can give rise, in the particular case of apple, to reproductive or vegetative growth units (GUs) of different lengths. In this study, the impact of long-term WS (7 years) on the fate of terminal and axillary buds was investigated in relation to flowering occurrence and production pattern (biennial vs regular) in the "Granny Smith" cultivar. It was observed that water stress decreased the total number of GUs per branch, regardless of their type. Conversely, water stress did not modify the timing of the two successive developmental phases characterized by the production of long and medium GUs and an alternation of floral GUs over time, respectively. The analysis of GU successions over time using a variable-order Markov chain that included both the effects of the previous flowering events and water treatment, revealed that water stress reduced the transition towards long and medium GUs and increased transition probabilities toward floral, short and dead GUs. Water stress also slightly increased the proportion of axillary floral GUs. The higher relative frequency of floral GUs compared with vegetative ones reduced the tendency to biennial bearing under water stress. The accelerated ontogenetic trend observed under water stress suggests lower vegetative growth that could, in turn, be beneficial to floral induction and fruit set [29], [37]. Ongoing work is conducted to determine the role of external (temperature and water stress) and internal (hormonal signalling, C source-sink relationships) factors in floral induction and consequently, in the regular or biennial behaviour in fruiting in apple trees. Particularly, its aim is to determine at which scale within the plant the production patterns are impacted by each factor. To analyse the carbon source-sink relationships from shoot to tree scales, this study is based on a set of genotypes displaying a large variability in flowering and production patterns.

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 [5] 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.

This year, the ALT tracking pipeline has been reformulated by using a generic cell modeling approach (enabling for example more than one cell division), and both stability and robustness were improved. The modeling approach is generic and can be used on other kind of data (nuclei, human cells, ...). Moreover, the architecture of the image processing components has been modified (plugin approach) and integrated with the TissueLab platform. The new segmentation-tracking library is called TimageTK will be released at the beginning of next year.

We also finalize 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).

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.

- *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 [62]. 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 [61].

  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.*

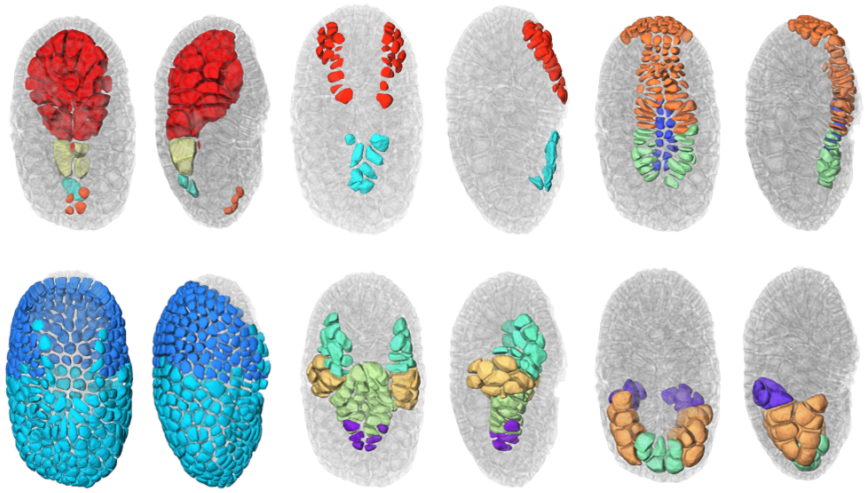


Figure 2. 3D projection of the segmented embryo at the early tailbud stage. The cells are colored by tissue type. The cells are slightly roded to allow their distinction. The other cells of the embryo are in transparent grey. The doral and lateral sides are shown.

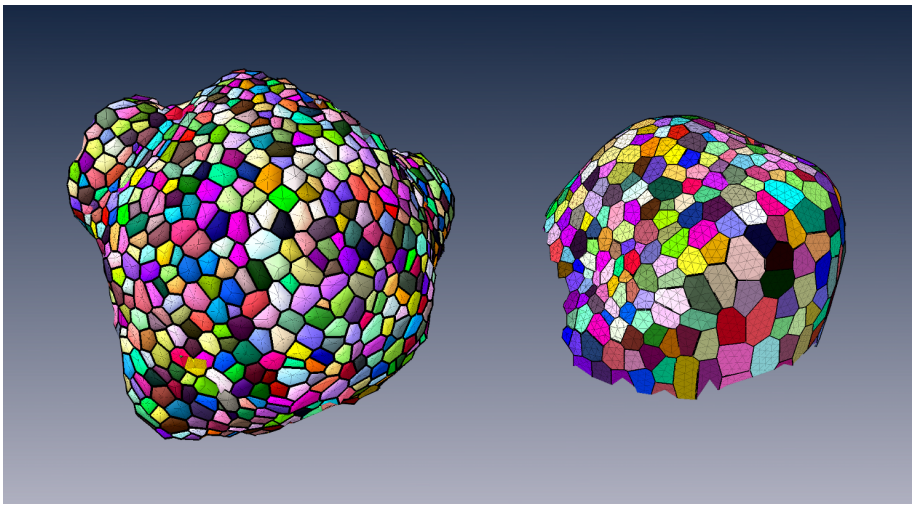


Figure 3. Triangular mesh representations of shoot apical meristem and flower meristem tissues obtained from MARS segmentations

Participants: Sophie Ribes, Yassin Refahi [RDP, ENS, Sainsbury Lab], Guillaume Cerutti, Christophe Godin, Christophe Pradal, Christophe Pradal, Frédéric Boudon, Gregoire Malandain [RDP, ENS], Gaël Michelin [RDP, ENS], Jan Traas [RDP, ENS], Teva Vernoux [RDP, ENS], Patrick Lemaire [CRBM, CNRS].

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 [72]. 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.

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

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

During the previous years, we set up a 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). The aim of this project is to provide a computational framework for simulating growth of multicellular plant tissue. This framework integrates a theoretical description of the major biophysical processes at stake. A first version of the model, based on a static description of the tissues rheological properties, has been published last year [57].

This year, we used this model in close collaboration with biologists to investigate the coupling between growth and cell wall remodeling required in early stages of organogenesis. 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 has been submitted to a high factor Biology journal.

In parallel, we also improved the underlying biophysical theory. One important aspect of the problem is the multiscale interconnections between mechanical forces generated at the scale of the whole tissue and the molecular response to these forces at the subcellular level. To tackle this issue, we established a parcimonous molecular description of the cell wall (one of the main organelle involded in growth) attesting for its biochemical behavior under mechanical loading. This description has been formalized as a unidimensional toy-model. With this toy-model we exposed how large-scale behavior of an expanding cell wall could be controlled by the biochemical behavior of a limited set of molecular actors. This work has been published [14].

Additionally, 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. On the molecular scale, we introduced a tensor formalism to quantify cell polarity, based on the description of its cortical microtubule network. Microtubules being stress-sensitive, we described this feedback loop through the coupling between this polarity tensor and the mechanical stress field. In parallel, through a parcimonous model of microtubule-guided cell wall turnover, we derived an expression of the stiffness tensor as a function of cell polarity. This enabled us to relate subcellular stress-induced dynamics of microtubules to the evolution of large scale rheological properties of the tissue. We also started to work on the numerical implementation of this feedback mechanism. FEM-based simulations have been carried out on simple structures as proof of concept. By doing so we assessed the numerical valididy of our resolution scheme along with the relevance of our biophysical description.

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. Modelling the influence of dimerisation sequence dissimilarities on the auxin signalling network

Participants: Jonathan Legrand, Yann Guédon, Jean-Benoist Léger [INRA, MIA, Paris], Stéphane Robin [INRA, MIA, Paris], Teva Vernoux [ENS-Lyon].

Auxin is a major phytohormone involved in many developmental processes by controlling gene expression through a network of transcriptional regulators. In *Arabidopsis thaliana*, the auxin signalling network is made of 52 potentially interacting transcriptional regulators, activating or repressing gene expression. All the possible interactions were tested in two-way yeast-2-hybrid experiments. Our objective was to characterise this auxin signalling network and to quantify the influence of the dimerisation sequence dissimilarities on the interaction between transcriptional regulators [20]. We applied model-based graph clustering methods relying on connectivity profiles between transcriptional regulators. Incorporating dimerisation sequence dissimilarities as explanatory variables, we modelled their influence on the auxin network topology using mixture of linear models for random graphs. Our results provide evidence that the network can be simplified into four groups, three of them being closely related to biological groups. We found that these groups behave differently, depending on their dimerisation sequence dissimilarities, and that the two dimerisation sub-domains might play different roles. We propose here the first pipeline of statistical methods combining yeast-2-hybrid data and protein sequence dissimilarities for analysing protein-protein interactions. We unveil using this pipeline of analysis the transcriptional regulator interaction modes.

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.

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. 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 (see section 4) 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.

6.3.3. Analyzing root growth and branching

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

This research theme is supported by two PhD programmes.

New 2D and 3D root phenotyping plateforms are emerging with associated image analysis toolbox (e.g. SmartRoot, 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 [35].
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.

This pipeline of analysis methods was applied to different species (maize, Pearl millet [23]) with contrasting biological objectives (study of genetic diversity for Pearl millet and of metabolic and hormonal controls of morphogenesis for maize).

6.3.4. Analyzing shoot and leaf elongation

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

This research theme is supported by the labex Agro project "Integrated model of plant organ growth".

This study is based on the observation that there is a lack of methods enabling the integrated analysis of the processes controling the vegetative development in *Arabidopsis thaliana*.

The change in leaf size and shape during ontogeny associated with heteroblastic development is a composite trait for which extensive spatio-temporal data can be acquired using phenotyping platforms. However, only part of the information contained in such data is exploited, and developmental phases are usually defined using a selected organ trait. We introduced new methods for identifying developmental phases in the *Arabidopsis* rosette using various traits and minimum a priori assumptions [21]. A first pipeline of analysis was developed combining image analysis and statistical models to integrate morphological, shape, dimensional and expansion dynamics traits for the successive leaves of the *Arabidopsis* rosette. Dedicated segmentation models called semi-Markov switching models were built for selected genotypes in order to identify rosette developmental phases. Four successive developmental phases referred to as seedling, juvenile, transition and adult were identified for the different genotypes. We show that the degree of covering of the leaf abaxial surface with trichomes is insufficient to define these developmental phases. Using our pipeline of analysis, we were able to identify the supplementary seedling phase and to uncover the structuring role of various leaf traits. This enabled us to compare on a more objective basis the vegetative development of *Arabidopsis* mutants.

We developed a second pipeline of analysis methods combining a semi-automatic method for segmenting leaf epidermis images based on the ilastik software, and the analysis of the obtained cell areas using a gamma or inverse Gaussian mixture models whose component parameters are tied by a scaling rule. These mixture models allowed us to estimate the distribution of the number of endocycles. We highlighted in this way that the mean number of endocycles changes drastically with leaf rank. We extended the inference approach to take into account not only complete cell areas but also censored cell areas (corresponding to cells that intercept the edges of the images). We also investigated possible temporal interpretations of endoreduplication using stochastic processes.

6.3.5. A stochastic model of phyllotaxis

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

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

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

Recently, our colleagues from ENS-Lyon observed intriguing perturbation in *arabidopsis* mutants. These perturbations were also present, to a lesser extent in the wild type. In a series of works [79], [69], [1], we could show that these perturbations patterns in both wild-type and mutant plants could be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs. After closer inspection, we realized that the mutated gene encodes a protein diffusing from the organs and creating a field around the organs that regulates the plastochron. We could demonstrate that in the mutant, the absence of this field leads to co-initiations and subsequently to the observed permutations.

To proceed further and find a mechanistic interpretation of this phenomenon, we developed a stochastic extension of the standard model of phyllotaxis. We first analyzed the properties of the inhibitory fields created by the existing primordia on the initiation of new primordia, and concluded that the angular positions of organs are very robust to perturbations while plastochrons may be dramatically affected. This suggested that there exists a strong decoupling between space and time in the patterning process. To account for this observation, we modeled the perception of the initiation signal by cells using stochastic processes coupled with the intensity of inhibitory fields and showed that the observed permutation patterns emerge spontaneously from this purely local processes. This model recapitulates accurately the classical phyllotactic patterns and, in addition, produces realistic pattern disorders at higher organization levels as a result of stochasticity in signal perception. We show that these subtle disorders surprisingly reveal key information on the functioning of the developmental system and can therefore be regarded as *biological watermarks* of the system. In genetically or environmentally modified plants, these biological watermarks inform us on the molecular mechanisms that have been affected in the experiment. Our theoretical analysis allows us to predict the specific pattern variations that would arise from perturbations of the signaling pathways involved in lateral inhibition signaling at the shoot apex [27].

6.3.6. 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

In the context of our research work on biological questions, we develop concepts and tools in mathematics, statistics and computer science. This paragraph is intended to put emphasis on the most important results obtained by the team during the current year in these disciplins, independently of their biological application.

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].

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 [26] 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 Scientific workflows

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

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 8.2.5.4 , we study how scientific workflows can help to improve the reproducibility of computational experiment in the domain of life science. 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, 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 [18] and [25].

6.4.4. Lossy compression of tree structures

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

In in [6], 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 [55] 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 [6] 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 [42].

6.4.5. Version climber

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

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.

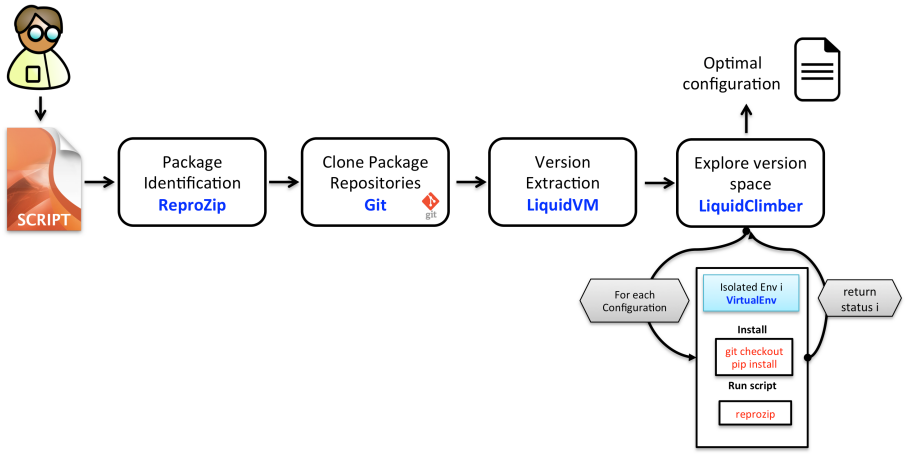


Figure 4. The steps of the operational subsystem: capture the execution of the initial configuration, liquify, fetch versions from git/svn etc., then deploy as directed by VersionClimber.

ARAMIS Project-Team

7. New Results

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

Participants: Pietro Gori [Correspondant], Olivier Colliot, Linda Marrakchi-Kacem, Yulia Worbe, Alexandre Routier, 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 (Figure 1). 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.

More details in [15].

7.2. Parsimonious Approximation of Streamline Trajectories in White Matter Fiber Bundles

Participants: Pietro Gori [Correspondant], Olivier Colliot, Linda Marrakchi-Kacem, Fabrizio de Vico Fallani, Mario Chavez, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman.

Fiber bundles stemming from tractography algorithms contain many streamlines. They require therefore a great amount of computer memory and computational resources to be stored, visualised and processed. We propose an approximation scheme for fiber bundles which results in a parsimonious representation of weighted prototypes. Prototypes are chosen among the streamlines and they represent groups of similar streamlines. Their weight is related to the number of approximated streamlines. Both streamlines and prototypes are modelled as weighted currents. This computational model does not need point-to-point correspondences and two streamlines are considered similar if their endpoints are close to each other and if their pathways follow similar trajectories. Moreover, the space of weighted currents is a vector space with a closed-form metric. This permits easy computation of the approximation error and the selection of the prototypes is based on the minimisation of this error. We propose an iterative algorithm which approximates independently and simultaneously all the fascicles of the bundle in a fast and accurate way. We show that the resulting representation preserves the shape of the bundle and it can be used to accurately reconstruct the original structural connectivity (Figure 2). We evaluate our algorithm on bundles obtained from both deterministic and probabilistic tractography algorithms. The resulting approximations use on average only 2% of the original streamlines as prototypes. This drastically reduces the computational burden of the processes where the geometry of the streamlines is considered. We demonstrate its effectiveness using as example the registration between two fiber bundles.

More details in [14].

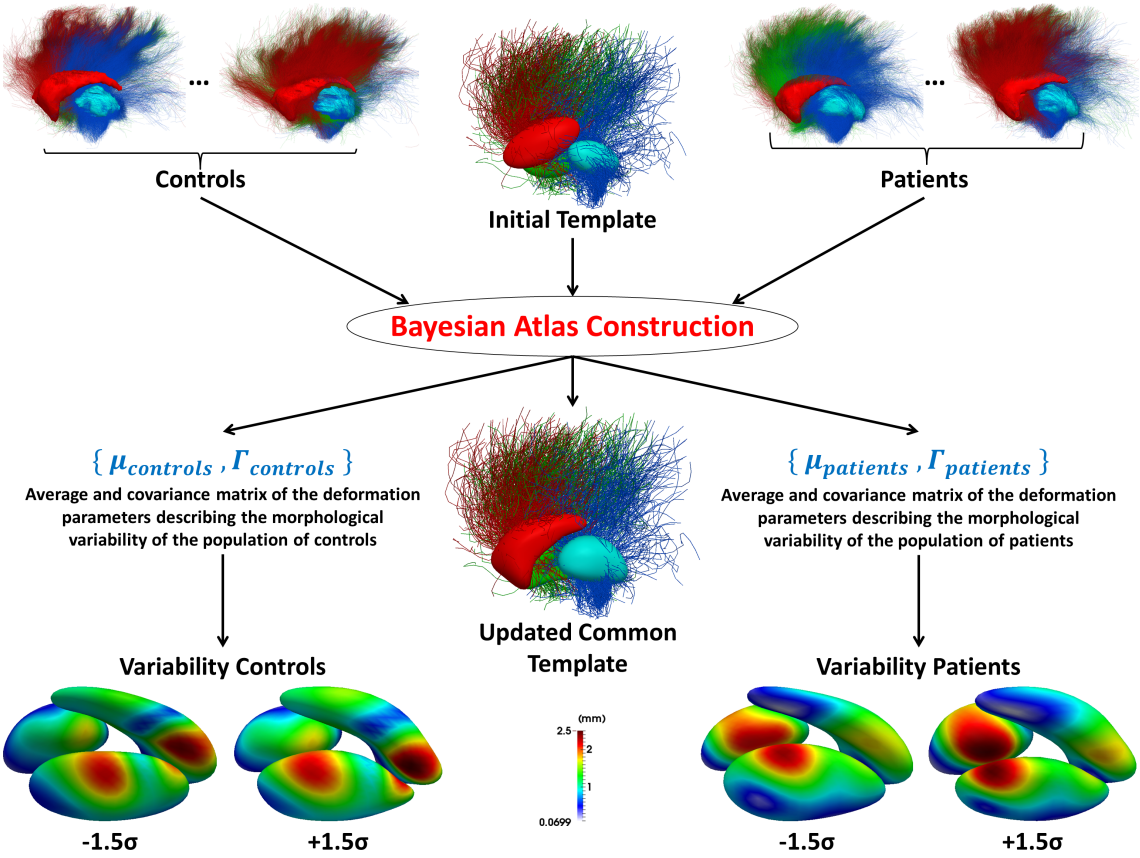


Figure 1. Bayesian framework for atlas construction of multi-object shape complexes comprised of both surface and curve meshes.

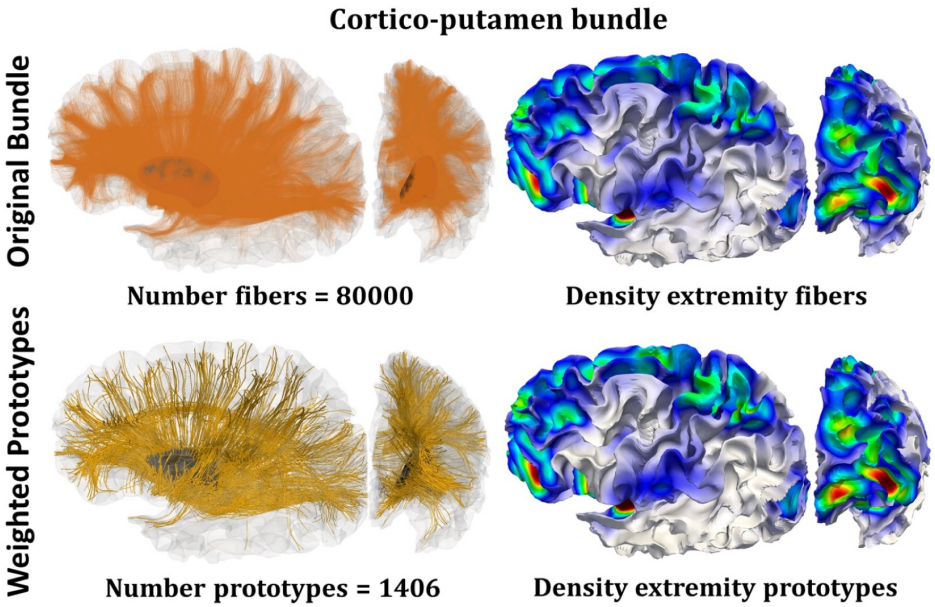


Figure 2. Weighted prototype approximations of a fiber bundle. As it is possible to notice, our approximation alters neither the global shape of the bundle nor the densities of the endpoints onto the cortical surface.

7.3. White matter lesions in FTLD: distinct phenotypes characterize GRN and C9ORF72 mutations

Participants: Fatima Ameur, Olivier Colliot, Didier Dormont, Alexis Brice, Isabelle Le Ber, Anne Bertrand [Correspondant].

Frontotemporal lobar degeneration (FTLD) has a high frequency of genetic forms; the 2 most common are GRN (progranulin) and C9ORF72 mutations. Recently, our group reported extensive white matter (WM) lesions in 4 patients with FTLD caused by GRN mutation, in the absence of noteworthy cardiovascular risk factors in line with other studies in GRN mutation carriers. Here we compared the characteristics of frontal WM lesions in patients with behavioral variant of FTLD (bv-FTLD) caused by GRN and C9ORF72 mutations. We found that WM lesions were more frequent and more atypical on both sides in the GRN group than in the control group and the C9ORF72 group.

More details in [3].

7.4. Riemannian geometry applied to detection of respiratory states from EEG signals: the basis for a brain-ventilator interface

Participants: Xavier Navarro-Sune, Anna Hudson, Fabrizio de Vico Fallani, Jacques Martinerie, Adrien Witon, Pierre Pouget, Mathieu Raux, Thomas Similowski, Mario Chavez [Correspondant].

During mechanical ventilation, patient-ventilator disharmony is frequently observed and may result in increased breathing effort, compromising the patient's comfort and recovery. This circumstance requires clinical intervention and becomes challenging when patients are sedated or verbal communication is difficult. In

this work, we propose a brain computer interface (BCI) to automatically and non-invasively detect patient-ventilator disharmony from electroencephalographic (EEG) signals: a brain-ventilator interface. Our framework exploits the cortical activation provoked by the inspiratory compensation when the subject and the ventilator are desynchronized (Figure 3). Use of a one-class approach and Riemannian geometry of EEG covariance matrices allows effective classification of respiratory states. The BVI is validated on nine healthy subjects that performed different respiratory tasks that mimic a patient-ventilator disharmony. Results evidence that classification performances, in terms of areas under ROC curves, are significantly improved using EEG signals compared to detection based on air flow. Reduction in the number of electrodes that can achieve discrimination can often be desirable (e.g. for portable BCI systems). By using an iterative channel selection technique, the Common Highest Order Ranking (CHOrRa), we find that a reduced set of electrodes (n=6) can slightly improve for an intra-subject configuration, and it still provides fairly good performances for a general inter-subject setting. Results support the discriminant capacity of our approach to identify anomalous respiratory states, by learning from a single training set containing only normal respiratory epochs. The proposed framework opens the door to brain-ventilator interfaces for monitoring patient's breathing comfort and adapting ventilator parameters to patient respiratory needs.

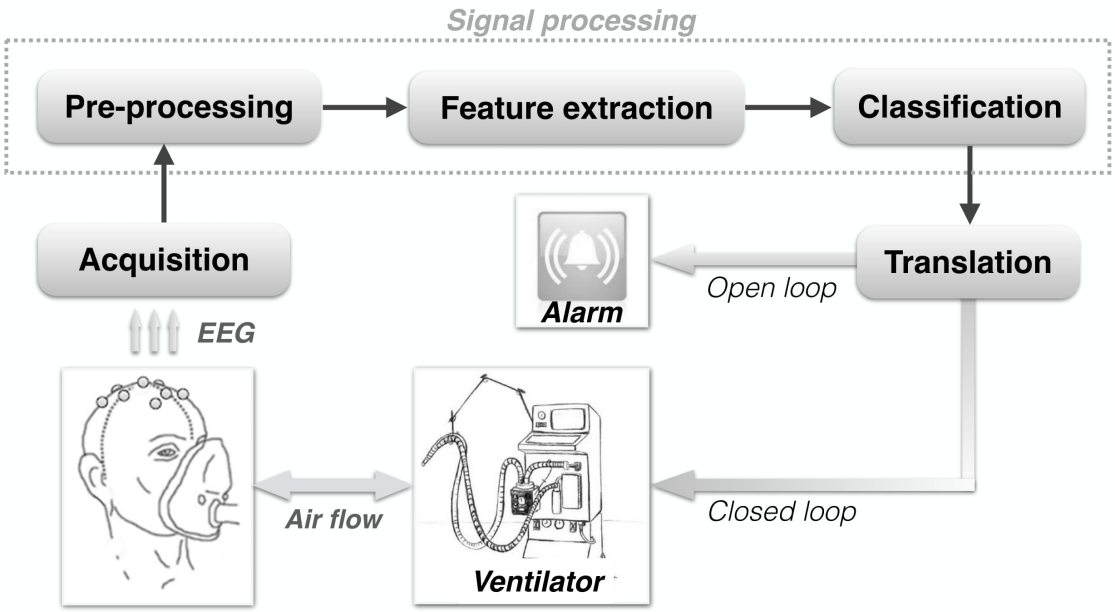


Figure 3. Scheme of a brain-ventilator interface.

More details in [25].

7.5. Interhemispheric Connectivity Characterizes Cortical Reorganization in Motor-Related Networks After Cerebellar Lesions

Participants: Fabrizio de Vico Fallani, Silvia Clausi, Maria Leggio, Mario Chavez, Miguel Valencia, Anton Giulio Maglione, Fabio Babiloni, Febo Cincotti, Donatella Mattia, Marco Molinari [Correspondant].

Although cerebellar-cortical interactions have been studied extensively in animal models and humans using modern neuroimaging techniques, the effects of cerebellar stroke and focal lesions on cerebral cortical

processing remain unknown. In the present study, we analyzed the large-scale functional connectivity at the cortical level by combining high-density electroencephalography (EEG) and source imaging techniques to evaluate and quantify the compensatory reorganization of brain networks after cerebellar damage. The experimental protocol comprised a repetitive finger extension task by 10 patients with unilateral focal cerebellar lesions and 10 matched healthy controls. A graph theoretical approach was used to investigate the functional reorganization of cortical networks. Our patients, compared with controls, exhibited significant differences at global and local topological level of their brain networks. An abnormal rise in small-world network efficiency was observed in the gamma band (30-40 Hz) during execution of the task, paralleled by increased long-range connectivity between cortical hemispheres (Figure 4). Our findings show that a pervasive reorganization of the brain network is associated with cerebellar focal damage and support the idea that the cerebellum boosts or refines cortical functions. Clinically, these results suggest that cortical changes after cerebellar damage are achieved through an increase in the interactions between remote cortical areas and that rehabilitation should aim to reshape functional activation patterns. Future studies should determine whether these hypotheses are limited to motor tasks or if they also apply to cerebro-cerebellar dysfunction in general.

More details in [11].

7.6. 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 an unrealistic 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 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 [12].

7.7. Robust imaging of hippocampal inner structure at 7T: in vivo acquisition protocol and methodological choices

Participants: Linda Marrakchi-Kacem [Correspondant], Alexandre Vignaud, Julien Sein, Johanne Germain, Thomas Henry, Cyril Poupon, Lucie Hertz-Pannier, Stephane Lehericy, Olivier Colliot, Pierre-François Van de Moortele, Marie Chupin.

Motion is a crucial issue for ultra-high resolution imaging, such as can be achieved with 7T MRI. An acquisition protocol was designed for imaging hippocampal inner structure at 7T. It relies on a compromise between anatomical details visibility and robustness to motion. In order to reduce acquisition time and motion artifacts, the full slab covering the hippocampus was split into separate slabs with lower acquisition time. A robust registration approach was implemented to combine the acquired slabs within a final 3D-consistent high-resolution slab covering the whole hippocampus. Evaluation was performed on 50 subjects overall, made of three groups of subjects acquired using three acquisition settings; it focused on three issues: visibility of hippocampal inner structure, robustness to motion artifacts and registration procedure performance. Overall,

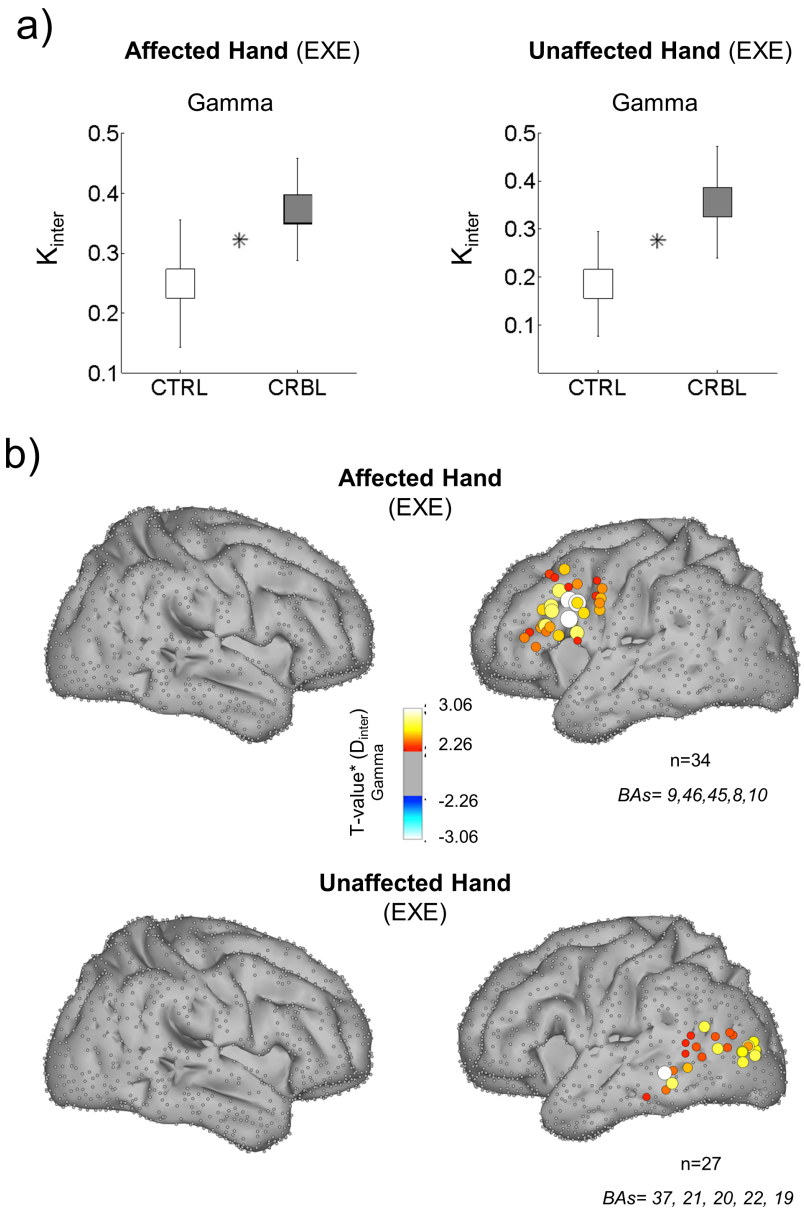


Figure 4. Gamma-band inter-hemispheric density (Kinter) and statistical contrasts of node degrees for brain networks during movement execution Panel a) Averaged Kinter values in the Gamma band for the CTRL and CRBL groups for the affected and unaffected hand conditions in the EXE phase. Panel b) T-value maps of the between-groups contrasts for the node-degree values over lateral views of the MNI cortical model in the Tailarach space in the affected (upper part) and unaffected (bottom part) hand conditions in the EXE phase.

T2-weighted acquisitions with interleaved slabs proved robust. Multi-slab registration yielded high quality datasets in 96% of the subjects, thus compatible with further analyses of hippocampal inner structure. Multi-slab acquisition and registration setting is efficient for reducing acquisition time and consequently motion artifacts for ultra-high resolution imaging of the inner structure of the hippocampus.

More details in [22].

7.8. Improved cerebral microbleeds detection using their magnetic signature on T2*-phase-contrast: a comparison study in a clinical setting

Participants: Takoua Kaaouana [Correspondant], Anne Bertrand, Fatma Ouamer, Bruno Law-Ye, Nadya Pyatigorskaya, Ali Bouyahia, Nathalie Thiery, Carole Dufouil, Christine Delmaire, Didier Dormont, Ludovic de Rochefort, Marie Chupin.

In vivo detection of cerebral microbleeds (CMBs) from T2* gradient recalled echo (GRE) magnitude image suffers from low specificity, modest inter-rater reproducibility and is biased by its sensitivity to acquisition parameters. New methods were proposed for improving this identification, but they mostly rely on 3D acquisitions, not always feasible in clinical practice. A fast 2D phase processing technique for computing internal field maps (IFM) has been shown to make it possible to characterize CMBs through their magnetic signature in routine clinical setting, based on 2D multi-slice acquisitions. However, its clinical interest for CMBs identification with respect to more common images remained to be assessed. To do so, systematic experiments were undertaken to compare the ratings obtained by trained observers with several image types, T2* magnitude, Susceptibility Weighted Imaging reconstructions (SWI) and IFM built from the same T2*-weighted acquisition. 15 participants from the MEMENTO multi-center cohort were selected: six subjects with numerous CMBs (20+/-6 CMBs), five subjects with a few CMBs (2 +/-1 CMBs) and four subjects without CMB. 2D multi-slice T2* GRE sequences were acquired on Philips and Siemens 3T systems. After pilot experiments, T2* magnitude, Susceptibility Weighted Imaging (SWI) minimum intensity projection (mIP) on three slices and IFM were considered for the rating experiments. A graphical user interface (GUI) was designed in order to consistently display images in random order. Six raters of various background and expertise independently selected “definite” or “possible” CMBs. Rating results were compared with respect to a specific consensus reference, on both lesion and subject type points Results: IFM yielded increased sensitivity and decreased false positives rate (FPR) for CMBs identification compared to T2* magnitude and SWI-mIP images. Inter-rater variability was decreased with IFM when identifying subjects with numerous lesions, with only a limited increase in rating time. IFM thus appears as an interesting candidate to improve CMBs identification in clinical setting.

More details in [19].

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, 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). Part of this work is also funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images). 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 modeling, Anatomical variability, Cochlear implant, Temporal bone.

- We evaluated the optimal electrode diameter in relation to the cochlear shape [20].
- We proposed a novel framework for estimating the insertion depth and its uncertainty from segmented CT images based on a new parametric shape model [37].
- We provided a proof of concept for the estimation of postoperative cochlear implant electrode-array position from clinical CT [44] (Fig. 1).

6.1.2. Infarct Localization and Uncertainty Quantification from Myocardial Deformation

Participants: Nicolas Duchateau [Correspondant], Maxime Sermesant.

This work received the partial support from the European Union 7th Framework Programme (VP2HF FP7-2013-611823) and the European Research Council (MedYMA ERC-AdG-2011-291080).

Myocardial infarct, Computer-aided diagnosis, Dimensionality reduction, Biomechanical modeling.

- We build upon preliminary work for the automatic localization of myocardial infarct from local wall deformation, which has potential for risk stratification from routine examination such as 3D echocardiography. Non-linear dimensionality reduction serves to estimate the Euclidean space of coordinates encoding deformation patterns (training phase), and is combined with multi-scale kernel regressions to link the deformation patterns, the low-dimensional coordinates and the infarct location for new cases (testing phase).
- We extend this approach by considering the different components of myocardial strain considered in clinical practice, and by taking advantage of the space of low-dimensional coordinates to model uncertainty in the infarct localization [18].
- These concepts were tested on 500 synthetic cases with infarcts of random extent, shape, and location, generated from a realistic electromechanical model, and 108 pairs of 3D echocardiographic sequences and delayed-enhancement magnetic resonance images from real cases. Infarct prediction is made at a spatial resolution more than 10 times smaller than the current diagnosis, made regionally. Our method is accurate, and significantly outperforms the clinically-used thresholding of the deformation patterns. Uncertainty adds value to refine the diagnosis and eventually re-examine suspicious cases.

6.1.3. Longitudinal Analysis and Modeling of Brain Development

Participants: Mehdi Hadj-Hamou [Correspondant], Xavier Pennec, Nicholas Ayache, Hervé Lemaitre [Inserm U1000], Jean-Luc Martinot [Inserm U1000].

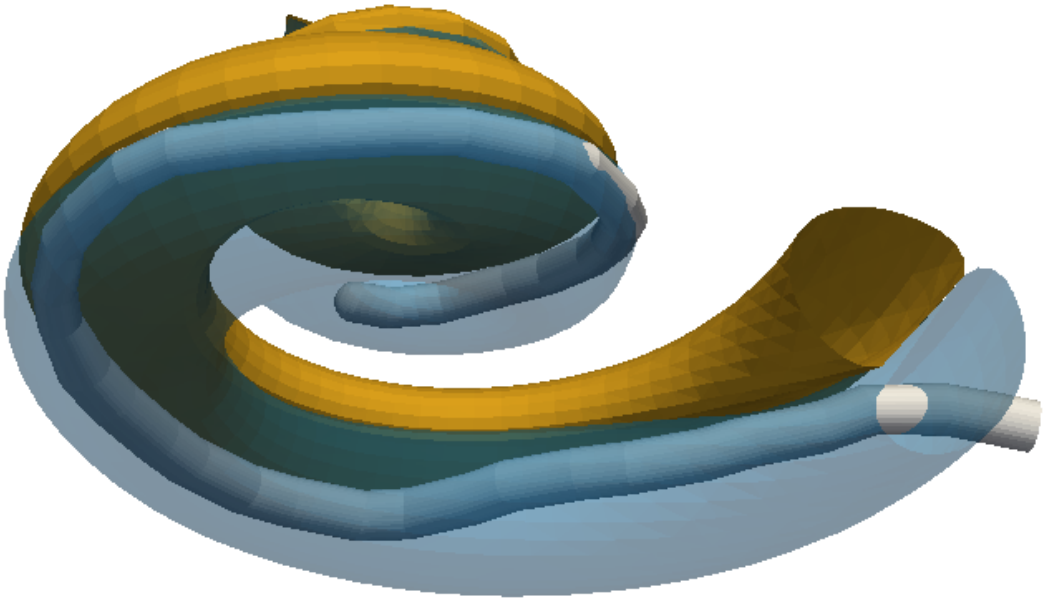


Figure 1. Cochlear implant electrode-array position (white) with respect to scala tympani (blue) and scala vestibuli (orange).

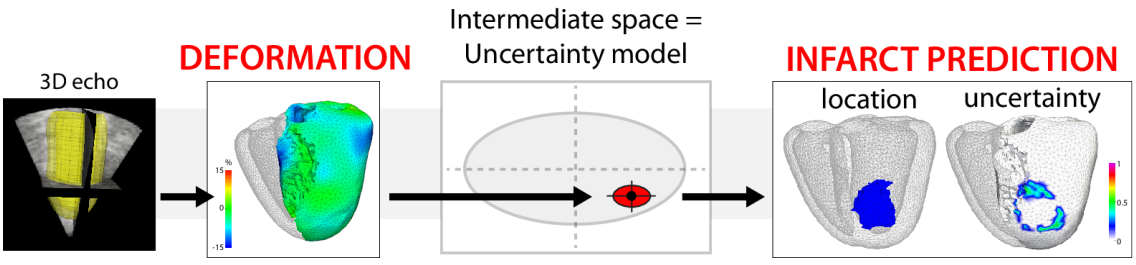


Figure 2. Overview of the proposed localization of myocardial infarct and uncertainty quantification from local wall deformation, extracted from 3D echocardiographic sequences.

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

Processing pipeline, Brain development, Adolescence, Longitudinal analysis, Non-rigid registration algorithm, Extrapolation.

1. We propose and detail a deformation-based morphometry computational framework, called Longitudinal Log-Demons Framework (LLDF), to estimate the longitudinal brain deformations from image data series, transport them in a common space and perform statistical group-wise analyses. It is based on freely available software and tools, and consists of three main steps (cf. Fig. 3):
   - Pre-processing;
   - Position correction; and
   - Non-linear deformation analysis.

   It is based on the LCC log-Demons non-linear symmetric diffeomorphic registration algorithm with an additional modulation of the similarity term using a confidence mask to increase the robustness with respect to brain boundary intensity artifacts.
2. This work led to a published journal publication [23].
3. The LLDF pipeline is exemplified on the longitudinal Open Access Series of Imaging Studies (OASIS) database and is applied to the study of longitudinal trajectories during adolescence, for which little is known. The aim of this project is to provide models of brain development during adolescence based on diffeomorphic registration parametrised by SVFs. We particularly focused our study on the link between sexual dimorphism and the longitudinal evolution of the brain. This work was done in collaboration with J.L. Martinot et H. Lemaître (Inserm U1000).

6.1.4. Left Atrial Wall Segmentation and Thickness Measurement using Region Growing and Marker-Controlled Geodesic Active Contour

Participants: Shuman Jia [Correspondant], Loïc Cadour, Hubert Cochet [IHU Liryc, Bordeaux], Maxime Sermesant.

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

Atrial fibrillation, Left atrial wall thickness, Image segmentation, Cardiac computed tomography (CT), Region growing, Geodesic active contour.

We proposed a method to segment the left atrial (LA) wall and measure the wall thickness from cardiac computed tomography images, making use of patient-specific intensity value information and surrounding environment (see Fig. 4).

We partially implemented the method in the MUSIC software and tested our pipeline on 10 datasets. The results achieved a good match of wall thickness with manual segmentation. We received a Best Paper Award for this work [39] at the 2016 STACOM Workshop in Athens, Greece.

6.1.5. Weakly Supervised Learning for Tumor Segmentation

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

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

Deep Learning, Segmentation, Classification, Tumor.

- The goal of this work is to develop new machine learning methods for the localization and segmentation of tumors, without relying on the ground truth provided by experts. In particular, we study Deep Learning methods for classification and weakly supervised localization (Figure 5).
- We proposed two methods of synthesis of brain 3D MR images, in order to use them during the training of Neural Nets. The proposed methods showed to improve our performance in localization of brain tumors.

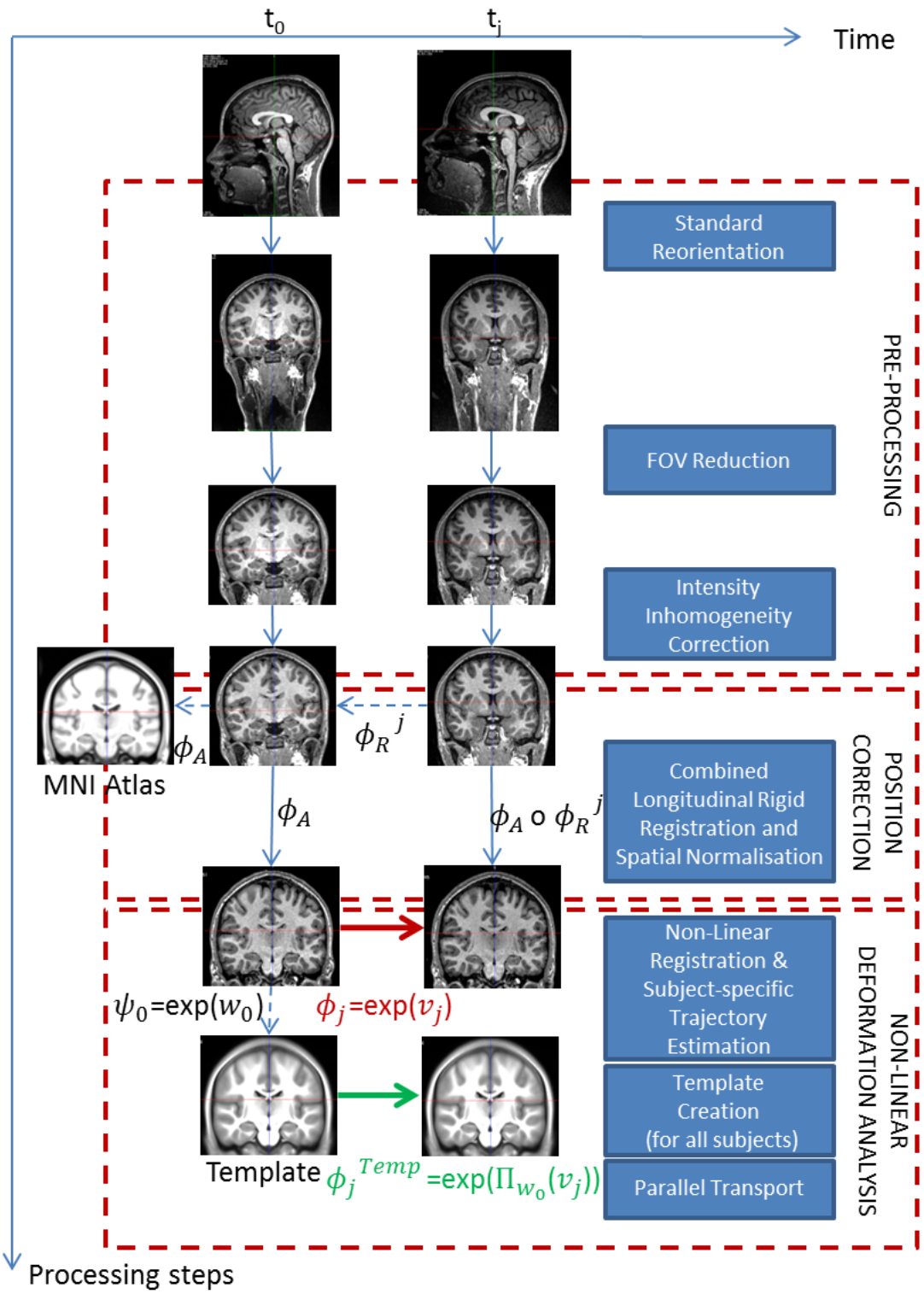


Figure 3. Processing pipeline for longitudinal analysis.

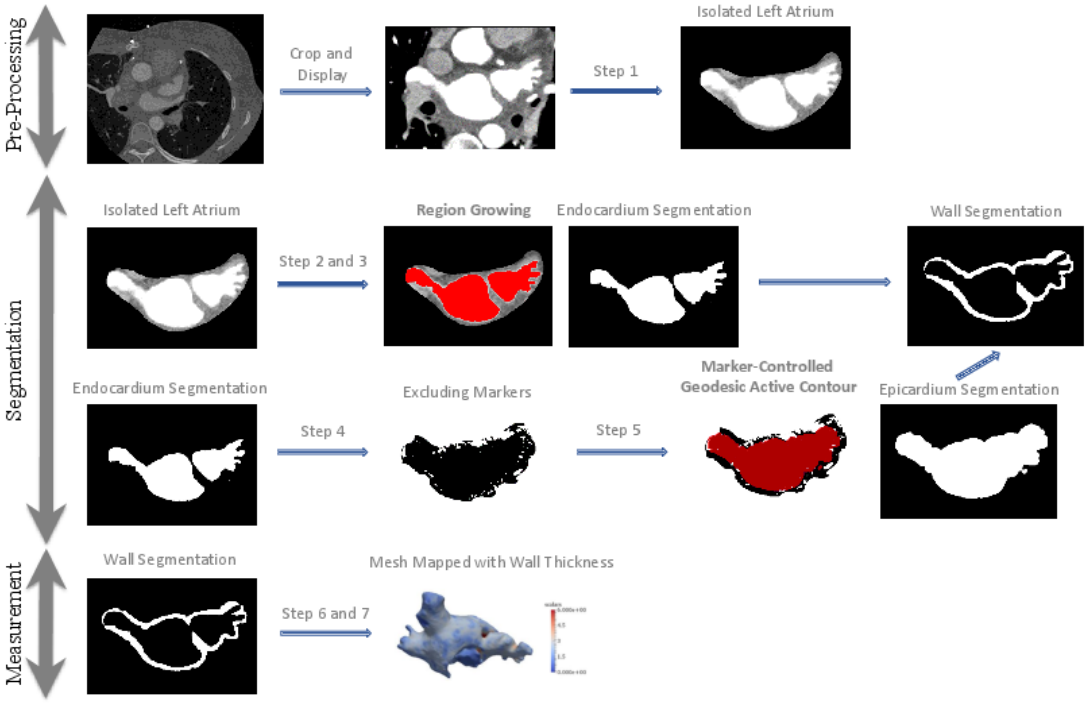


Figure 4. Flowchart of the method.

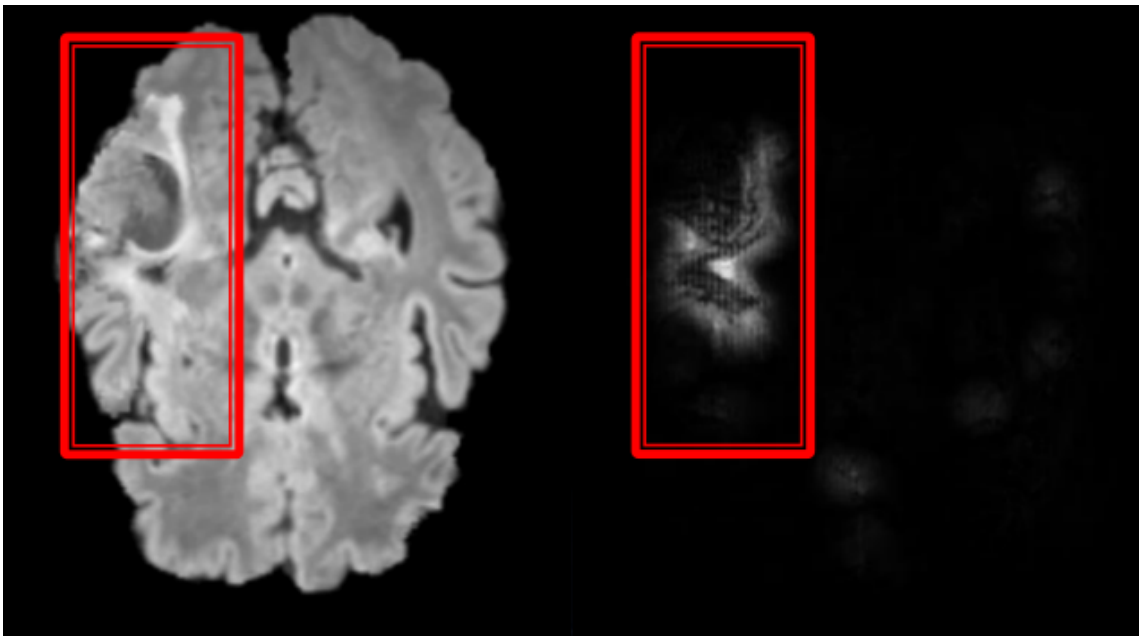


Figure 5. Left: axial slice of a brain MR image containing a tumor. Right: test of a weakly supervised (no ground truth provided during the training) Deep Learning method for an approximate localization of the tumor.

6.1.6. Inter-Operative Relocalization in Flexible Endoscopy

Participants: Anant Vemuri [Correspondant], Stéphane Nicolau, Luc Soler, Nicholas Ayache.

This work has been performed in collaboration with IHU Strasbourg and IRCAD, France.

Computer-assisted intervention, Barrett's esophagus, Biopsy relocalization, Electromagnetic tracking.

Oesophageal adenocarcinoma arises from Barrett's oesophagus, which is the most serious complication of gastro-oesophageal reflux disease. Strategies for screening involve periodic surveillance and tissue biopsies. A major challenge in such regular examinations is to record and track the disease evolution and relocalization of biopsied sites to provide targeted treatments.

In support of this work, the thesis [6] was defended before a committee of medical experts and scientific reviewers, on April 26th 2016.

References:

- Vemuri, Nicolau, Sportes, Marescaux, Soler, Ayache. Inter-Operative Biopsy Site Relocalization in Endoluminal Surgery [35].
- Nicolau, Vemuri, Soler, Marescaux. Anatomical site relocalisation using dual data synchronisation (patent) [49].

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

Participants: Wen Wei, Nicholas Ayache, Stanley Durrleman [ARAMIS], Olivier Colliot [ARAMIS].

Multiple sclerosis, Neuroimageing.

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.

6.1.8. Deep Learning for Cardiac Image Analysis

Participants: Qiao Zheng [Correspondant], Hervé Delingette, Nicholas Ayache.

Deep learning, Artificial neural network, Cardiac image.

Deep learning has proven to be very successful in computer vision and image understanding. However, its potential for medical image analysis has yet to be explored. We apply deep learning on cardiac images in order to learn cardiac image processing and anomaly detection. Our work includes data collection and preprocessing, software engineering, learning process design, etc.

6.2. Computational Anatomy

6.2.1. Inconsistency of the Estimation of the Template in Quotient Spaces

Participants: Loïc Devilliers [Correspondant], Stéphanie Allassonnière [Ecole Polytechnique], Alain Trouvé [ENS Cachan], Xavier Pennec.

Atlas computation, Template estimation, Fréchet mean, Quotient spaces, Inconsistency.

One issue in computational anatomy is to compute a template (a prototype of our data) in presence of two effects: an unknown deformation on data and the noise due to error in measurement in the ambient space considered here as an infinite dimensional linear space. The template computation can be done by minimizing an energy function (or variance) in the quotient space. In [50], we show that this method can lead to inconsistency that we quantify (see Fig. 6). This paves the way to a better understanding of the geometric and statistic foundation of the template estimation.

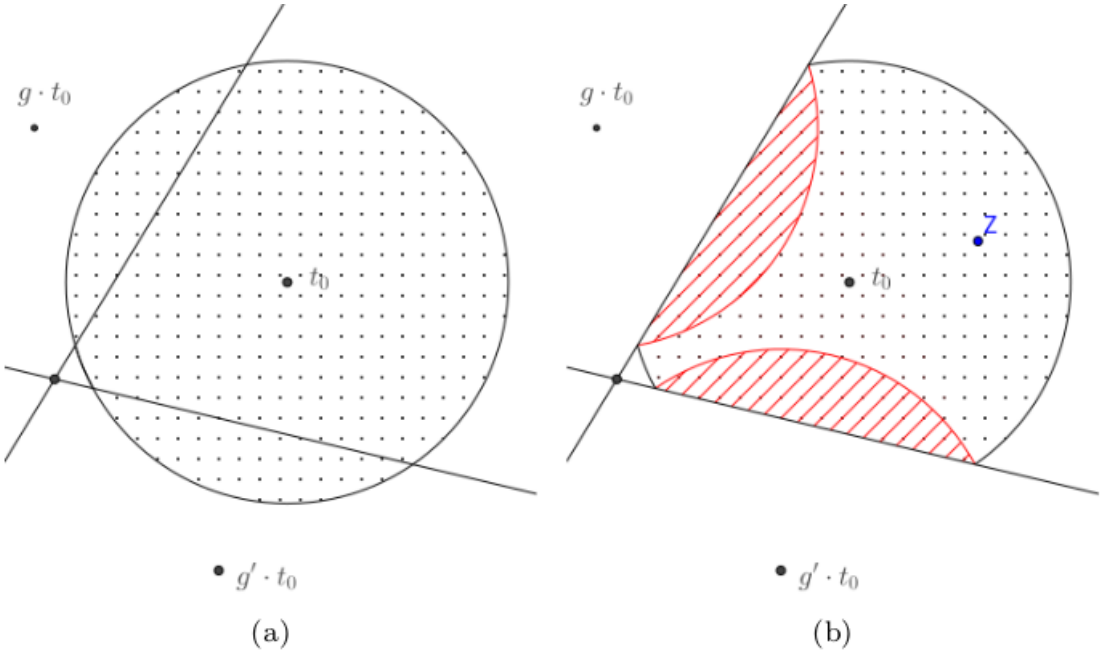


Figure 6. Geometric origin of the inconsistency: the distribution in the ambient space (panel (a)) is folded by the quotient (symmetries around the two lines, panel (b)), which biases the distribution in the quotient space.

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.

- First, we have shown in [52] that the usual algorithm of template organ shape estimation is biased (see Fig. 7). We proposed two bootstrap procedures that quantify the bias and correct it.
- In [53], we unified the template estimation problem with a manifold learning problem. We showed how the Bayesian framework enables correction in cases that are pathological if only the Maximum Likelihood estimator is used.

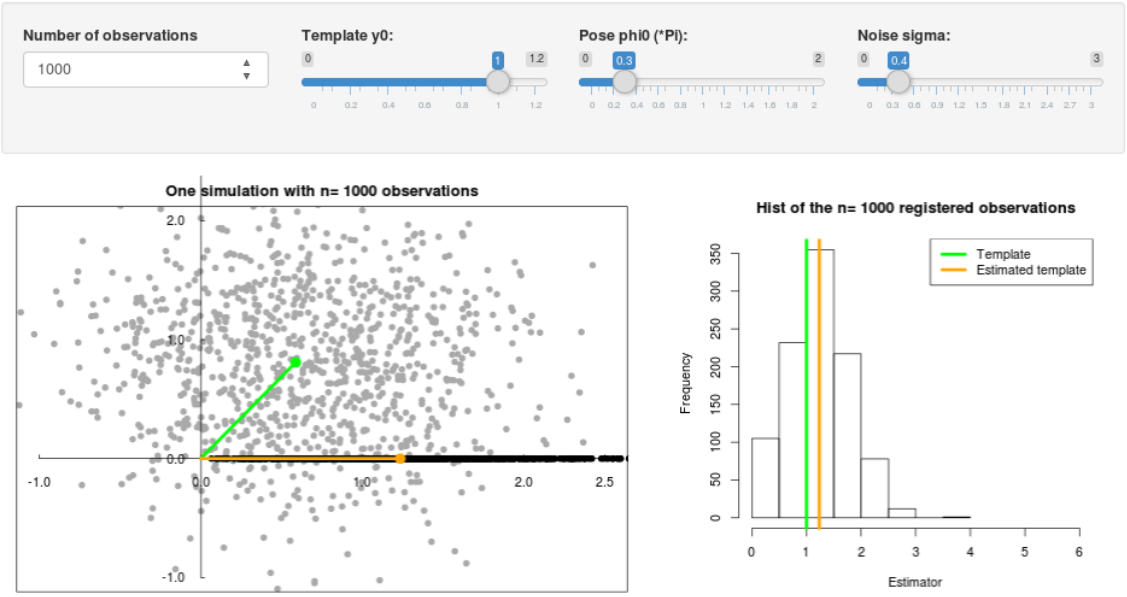


Figure 7. In this toy example, the template estimate (in orange) is biased with respect to the parameter (in green).

6.2.3. Barycentric Subspace Analysis: a new Symmetric Group-Wise Paradigm for Cardiac Motion Tracking

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

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

Low-dimensional analysis, Cardiac motion, Registration, Image synthesis.

We propose a novel approach to study cardiac motion in 4D image sequences using low-dimensional subspace analysis [43]. Instead of building subspaces relying on a mean value we use a novel type of subspaces called Barycentric Subspaces, which are are implicitly defined based on $k+1$ reference images instead of being defined with respect to one reference image. This allows:

- First: to build low-dimensional representation of the cardiac motion signature which actually separates perfectly two different populations.

- Second: to build a better prior for the cardiac motion tracking, which improves the registration accuracy at end-systole by 30%.
- Third: to reconstruct the sequence of images with better accuracy than traditional single reference methods.

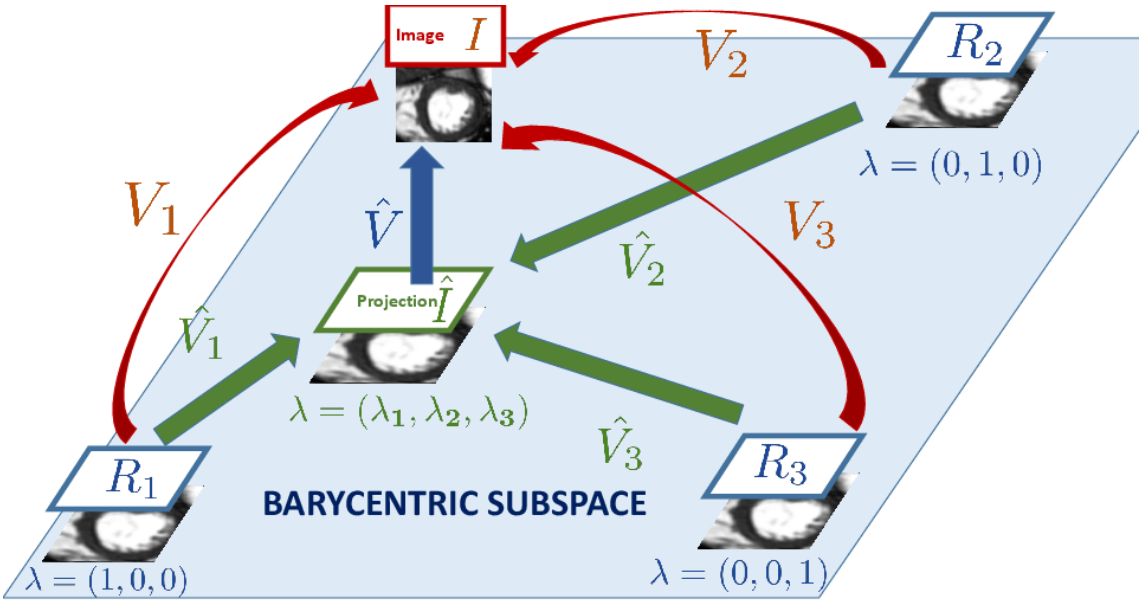


Figure 8.

6.2.4. Compact Representation of Longitudinal Deformations

Participants: Raphaël Sivera [Correspondant], Hervé Delingette, Xavier Pennec, Nicholas Ayache.

Longitudinal modeling, Learning in manifolds, Structured sparsity.

The use of a comprehensive and meaningful decomposition of a set of structural transformations would be useful to describe evolutions and to enhance diagnosis. In this context, we aim to model the brain anatomical evolution which goes along the Alzheimer's neurodegenerative disease. Based on the Stationary Velocity Fields representation of diffeomorphisms, we proposed a description of deformations in both space and time. The objective is to go beyond simple discriminative approaches (see Fig. 9) to propose a synthetic description of the disease evolution, population and subject-wise.

6.3. Computational Physiology

6.3.1. Computational Modeling of Radiofrequency Ablation for the Planning and Guidance of Abdominal Tumor Treatment

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

This work is carried out between the Asclepios research group, Inria Sophia Antipolis, France and the Medical Imaging Technologies, Healthcare Technology Center, Siemens Medical Solutions USA, Princeton, NJ.

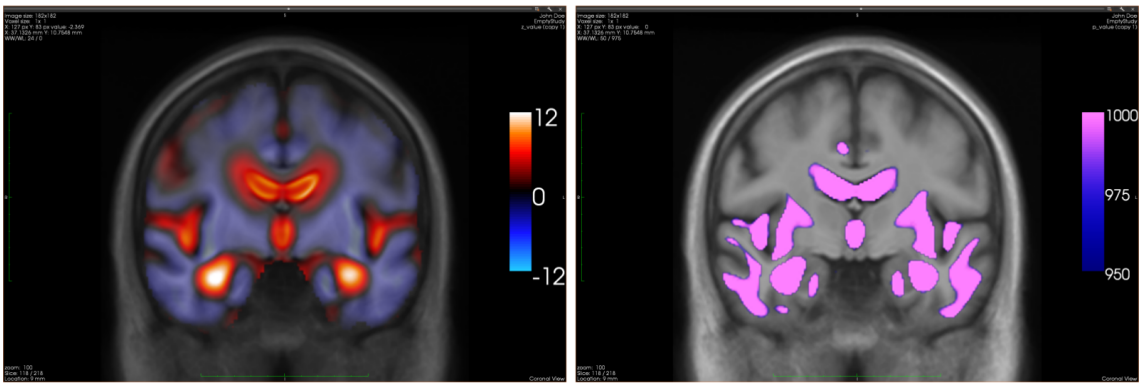


Figure 9. Left: z-values associated with group-wise differences between controls and Alzheimer-diagnosed subjects. Right: Areas of statistically significant differences.

Radio frequency ablation modeling, Patient-specific simulation, Lattice Boltzmann method, Computer model, Computational fluid dynamics, Heat transfer, Cellular necrosis, Parameter estimation, Therapy planning, Liver, Pre-clinical study, Medical imaging.

RFA is a minimally invasive therapy appropriated for liver tumor ablation. However, a patient-specific predictive tool to plan and guide the treatment is needed. We developed a computational framework for patient-specific planning of RFA with the following contributions:

- A detailed computational model of the biophysical mechanisms (heat transfer, cellular necrosis, hepatic blood flow) involved in RFA of abdominal tumors based on patient images.
- A new implementation of the bio-heat equations coupled with a cellular necrosis model using the Lattice Boltzmann Method (LBM) on Graphics Processing Units (GPU), which allows near real-time computation.
- A CFD and porous media solver using LBM algorithm to compute the patient-specific blood flow in the hepatic circulatory system and the blood flow distribution inside the parenchyma.
- A complete patient-specific geometry including hepatic venous and arterial circulation system.
- The automatic estimation of the main parameters of the model. Two personalization strategies tested and evaluated on clinical and pre-clinical data.
- The evaluation of the proposed model on a clinical dataset of ten patients.
- The evaluation on a preclinical dataset of five swines from a comprehensive experimental set-up specially designed for RFA model validation.

6.3.2. Cardiac Electrophysiology Simulation for Arrhythmia Treatment Guidance

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

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Cardiac electrophysiology modeling, Intracardiac electrogram modeling, Radiofrequency ablation planning, Electroanatomical mapping.

1. We developed silico patient-specific models constructed from 3D delayed-enhanced MRI to simulate intracardiac electrograms (EGM), including abnormal EGM as they are potential radiofrequency ablation targets (see Fig. 11) [14].
2. We derived a cardiac model using personalized electro-anatomical parameters and imaging data to define the underlying ventricular tachycardia (VT) substrate and predict re-entrant VT circuits [16].

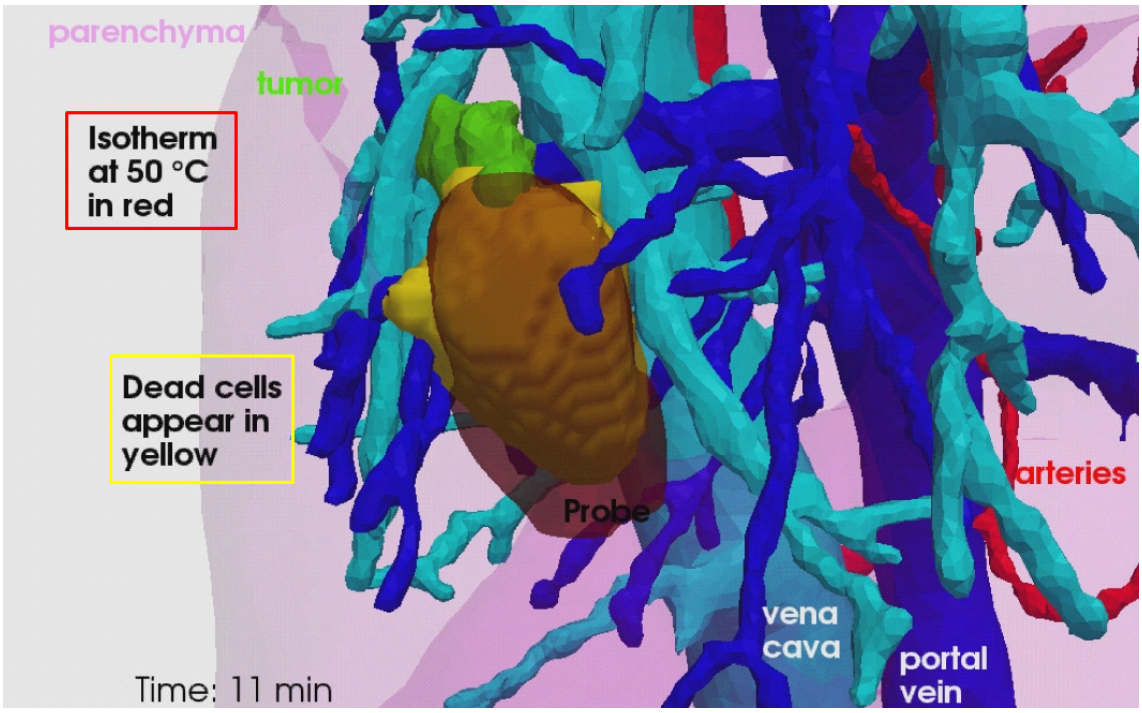


Figure 10. Computed isotherm at 50°C and computed necrosis appears in a subject-specific geometry.

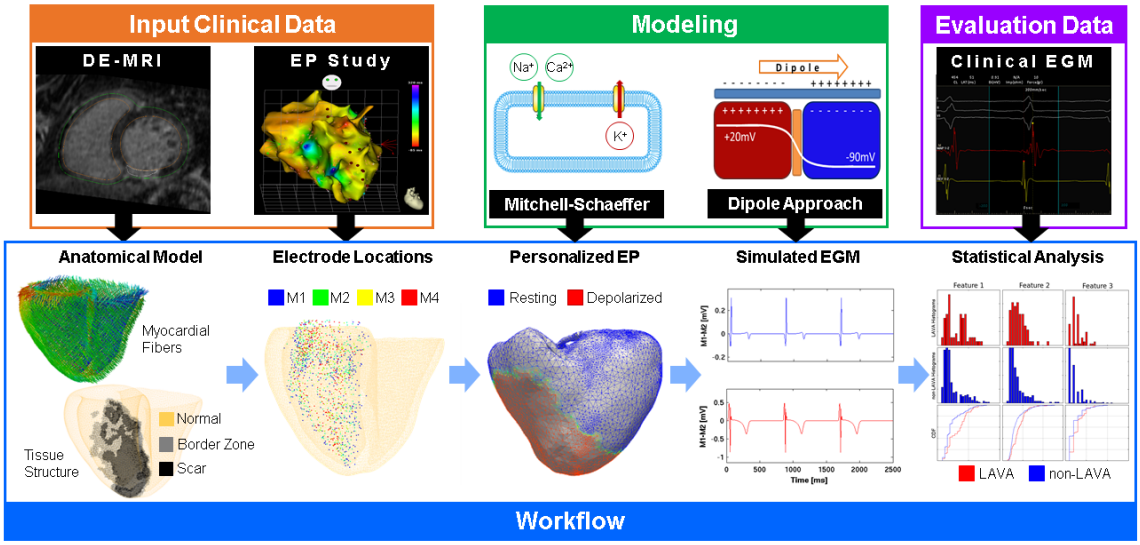


Figure 11. Pipeline developed to simulate intracardiac electrogram using patient-specific models.

6.3.3. Non-Invasive Personalisation of a Cardiac Electrophysiology Model from Body Surface Potential Mapping

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

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

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

Within the VP2HF project, non-invasive cardiac electrical data has been acquired at the St Thomas' Hospital, London. It consists of Body Surface Potential Mapping (BSPM), which are recordings of the electrical potential on several locations on the surface of the torso. In [19], we use non-invasive data (BSPM) to personalise the main parameters of a cardiac electrophysiological (EP) model for predicting different pacing conditions (see Fig. 12). This is an encouraging first step towards a pre-operative prediction of different pacing conditions to assist clinicians for CRT decision and procedure. We have also worked on ECG data that are more commonly used in practice. In [38], we estimated the purkinje activation from 12-lead ECG using an intermittent left bundle brand block patient dataset.

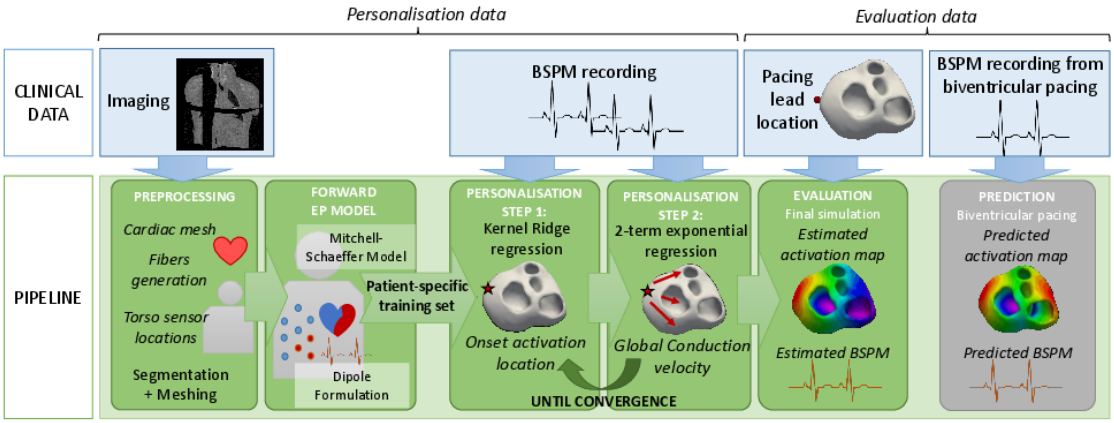


Figure 12. Personalisation framework.

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

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

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

Alzheimer's Disease (AD), Modeling brain deformation, Biophysical model, Simulation.

- We completed a simulation tool that can simulate large databases of virtual realistic longitudinal MRIs with known volume changes[51]. This was based on our biophysical model of brain deformation due to atrophy in Alzheimer's Disease (AD)[25].
- We have released our simulation software, named simul@trophy, as an open source software https://inria-asclepios.github.io/simul-atrophy/.

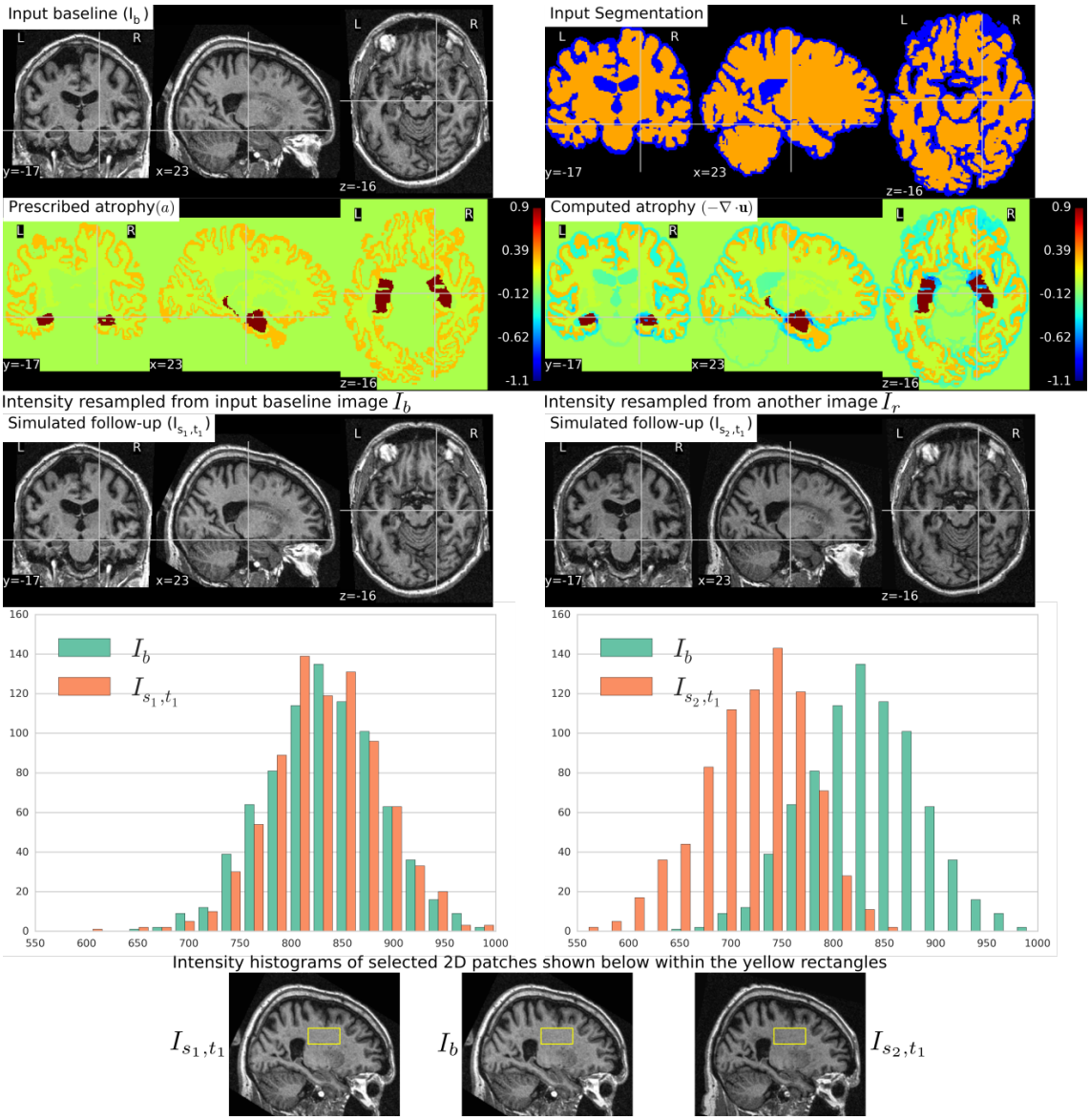


Figure 13. ***1st row:*** *(left) input baseline image I_b; (right) its input segmentation image.* ***2nd row:*** *(left) prescribed atrophy; (right) the atrophy computed from the simulated deformation.* ***3rd row:*** *(left) first time-point simulated follow-up image I_{s_1,t_1} where the intensity is resampled from the input baseline image I_b; (right) first time-point simulated follow-up image I_{s_2,t_2} where the intensity is resampled from a MRI taken at a different time-point than I_b, but of the same patient.* ***4th row:*** *intensity histogram comparison of the two simulated images in the third row.* ***5th row:*** *a relatively uniform region of which the histogram is shown.*

6.3.5. Brain Tumor Growth Personalization and Segmentation Uncertainty

Participants: Matthieu Lê [Correspondant], Hervé Delingette, Jan Unkelbach, Nicholas Ayache.

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

Tumor growth, Radiotherapy, Modeling, Personalization, Segmentation, Uncertainty, Bayesian.

- We elaborated a method for the synthesis of magnetic resonance images (MRIs) presenting glioblastoma [17].
- We elaborated a method for the sampling of several plausible segmentations, based on a single clinical one. This allows the uncertainty quantification of the radiotherapy plan based on several sample clinical target volumes [30].
- We elaborated a method for the Bayesian personalization of a brain tumor growth model based on clinical MRIs [28].
- We combined the segmentation sampling method with the tumor growth model personalization to personalize radiotherapy planning (see Fig. 14).

6.3.6. A Multiscale Cardiac Model for Fast Personalisation and Exploitation

Participants: Roch Philippe Molléro [Correspondant], Xavier Pennec, Hervé Delingette, 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 modeling, Reduced model, Multi-fidelity modeling, Parameter estimation, Finite element mechanical modeling.

We developped a multi-fidelity 0D/3D cardiac model that allows us to get reliable (and extremely fast) approximations of the global behaviour of the 3D model with 0D simulations.

By making geometrical assumptions of symmetry, we first built a reduced 0D model of the heart which is very fast (15 beats/seconds). Then, we developped an original coupling method between the parameters of the 3D model and those of the 0D model. We used this multi-fidelity of the heart (in 0D and 3D) to speed-up an efficient optimization algorithm (the genetic algorithm CMA-ES) for the 3D model. As a result, we now have a fast personalisation method for the 3D model (see 15).

This methodology lead to a publication and poster presentation at the MICCAI Conference 2016 [41].

We applied this methodology in particuar to the cohort of 34 different heart geometries and data from the project MD-PAEDIGREE.

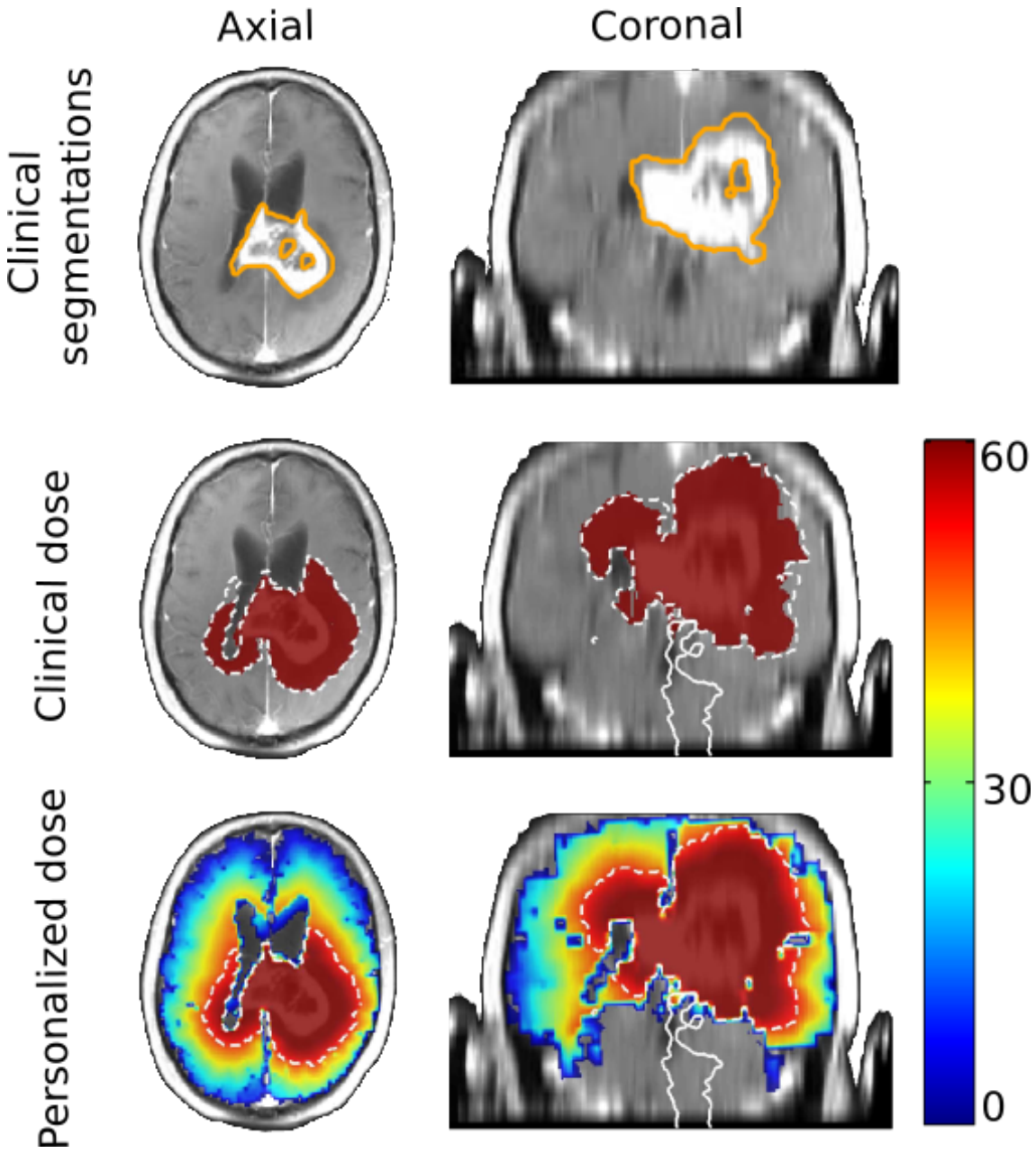


Figure 14. The clinical segmentation of the T1Gd abnormality (top, orange line) is used to define the clinical target volume (CTV, white dashed line) as a 2 cm expansion of the segmentation. In clinical settings, 60 Gy is prescribed to the CTV. We propose to personalize the prescription dose (bottom) to account for tumor infiltration and segmentation uncertainty.

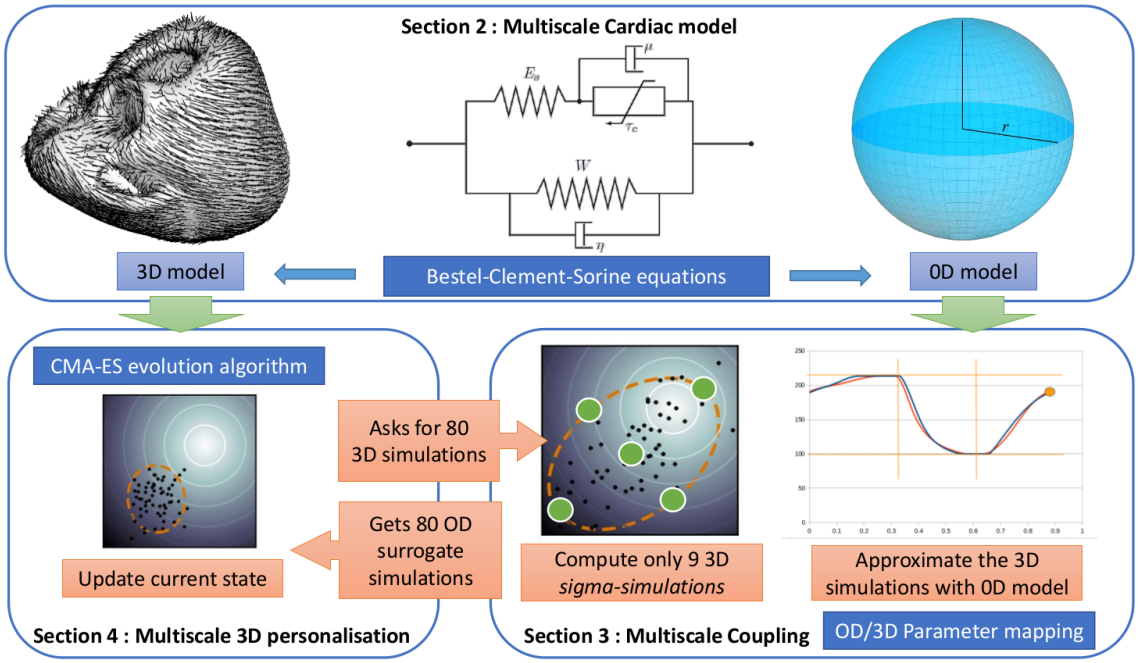


Figure 15. Multi-fidelity model and personalisation pipeline.

ATHENA Project-Team

7. New Results

7.1. Modeling in Diffusion MRI

7.1.1. Computational brain connectivity mapping: A core health and scientific challenge

Participant: Rachid Deriche.

One third of the burden of all the diseases in Europe is due to problems caused by diseases affecting brain. Although exceptional progress have been obtained for exploring the brain during the past decades, it is still terra-incognita and calls for specific efforts in research to better understand its architecture and functioning. To take up this great challenge of modern science and to solve the limited view of the brain provided just by one imaging modality, this article advocates the idea developed we develop in my research group of a global approach involving new generation of models for brain connectivity mapping and strong interactions between structural and functional connectivities. Capitalizing on the strengths of integrated and complementary non invasive imaging modalities such as diffusion Magnetic Resonance Imaging (dMRI) and Electro and Magneto-Encephalography (EEG & MEG) will contribute to achieve new frontiers for identifying and characterizing structural and functional brain connectivities and to provide a detailed mapping of the brain connectivity, both in space and time. Thus leading to an added clinical value for high impact diseases with new perspectives in computational neuro-imaging and cognitive neuroscience.

The work leading to the objectives listed in this article has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Advanced Grant agreement No 694665 started on Sept. 1st, 2016).

This article has been published in [15].

7.1.2. A survey of current trends in diffusion MRI for structural brain connectivity

Participants: Aurobrata Ghosh [University College London, UK], Rachid Deriche.

In this work, we review the state of the art in diffusion magnetic resonance imaging (dMRI) and we present current trends in modelling the brain's tissue microstructure and the human connectome. dMRI is today the only tool that can probe the brain's axonal architecture in vivo and non-invasively, and has grown in leaps and bounds in the last two decades since its conception. A plethora of models with increasing complexity and better accuracy have been proposed to characterise the integrity of the cerebral tissue, to understand its microstructure and to infer its connectivity. Here, we discuss a wide range of the most popular, important and well-established local microstructure models and biomarkers that have been proposed from these models. Finally, we briefly present the state of the art in tractography techniques that allow us to understand the architecture of the brain's connectivity.

This work has been published in [17].

7.1.3. Multi-Spherical Diffusion MRI: Exploring Diffusion Time Using Signal Sparsity

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 t - 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 qt-space, which we call qt-dMRI. To drastically reduce the number of diffusionweighted images (DWIs) we need to represent the qt-space, we regularize the fitting of qt-dMRI by imposing both signal smoothness and sparsity. As the main contribution, qt-dMRI provides the framework for estimating time-dependent q-space indices (qt-indices), providing new means for studying subdiffusion 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 excellent reproducibility of estimated qt-index values and trends. In the hopes of opening up new t-dependent venues of studying nervous tissues, qt-dMRI is the first of its kind in being specifically designed to provide open interpretation of the qt-diffusion signal.

This work has been partly published in [31]. The test-retest study has been submitted to ISMRM'17 and an extended version has been submitted to Neuroimage.

7.1.4. Noise Floor Removal via Phase Correction of Complex Diffusion-Weighted Images: Influence on DTI and q-space Metrics

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

The non-Gaussian noise distribution in magnitude Diffusion-Weighted Images (DWIs) can severely affect the estimation and reconstruction of the true diffusion signal. As a consequence, also the estimated diffusion metrics can be biased. In this work, we study the effect of phase correction, a procedure that re-establishes the Gaussianity of the noise distribution in DWIs by taking into account the corresponding phase images. We quantify the debiasing effects of phase correction in terms of diffusion signal estimation and calculated metrics. We perform in silico experiments based on a MGH Human Connectome Project dataset and on a digital phantom, accounting for different acquisition schemes, diffusion-weightings, signal to noise ratios, and for metrics based on Diffusion Tensor Imaging and on Mean Apparent Propagator Magnetic Resonance Imaging, i.e. q-space metrics. We show that phase correction is still a challenge, but also an effective tool to debias the estimation of diffusion signal and metrics from DWIs, especially at high b-values.

This work has been published in [39] and [60].

7.2. Tissue Microstructures features recovery & applications

7.2.1. MAPL: Tissue microstructure estimation using Laplacian-regularized MAP-MRI and its application to HCP data

Participants: Rutger Fick, Demian Wassermann, Emanuel Caruyer, Rachid Deriche.

The recovery of microstructure-related features of the brain's white matter is a current challenge in diffusion MRI. To robustly estimate these important features from diffusion MRI data, we propose to analytically regularize MAP-MRI's coefficient estimation using the norm of the Laplacian of the reconstructed signal. We first compare our approach, which we call MAPL, with competing state-of-the-art functional basis approaches. We show that it outperforms the original MAP-MRI implementation and the recently proposed modified Spherical Polar Fourier (mSPF) basis with respect to signal fitting, EAP and ODF reconstruction in noisy, sparsely sampled data of a physical phantom with reference gold standard data. Then, to reduce the variance of parameter estimation using multi-compartment tissue models, we propose to use MAPL's signal fitting and extrapolation as a preprocessing step. We study the effect of MAPL on the estimation of axon diameter using a simplified Axcaliber model and axonal dispersion using the Neurite Orientation Dispersion and Density Imaging (NODDI) model. We show the positive effect of using it as a preprocessing step in estimating and reducing the variances of these parameters in the Corpus Callosum of six different subjects of the MGH Human Connectome Project. Finally we correlate the estimated axon diameter, dispersion and restricted volume fractions with Fractional Anisotropy (FA) and clearly show that changes in FA significantly correlate with changes with all estimated parameters. Overall, we illustrate the potential of using a well-regularized functional basis together with multi-compartment approaches to recover important microstructure

tissue parameters with much less variability, thus contributing to the challenge of better understanding microstructure-related features of the brain's white matter.

This work has been published in [16]

7.2.2. Quantifying White Matter Microstructure with a Unified Spatio-Temporal Diffusion Weighted MRI Continuous Representation

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

A current problem Diffusion MRI (dMRI) based microscopy under the narrow pulse approximation is how to best exploit the 4D (q-space + diffusion time) nature of the signal. Assaf et al. showed that exploring the dMRI attenuation at different diffusion times provides information on the distribution of axonal diameters within a voxel in their seminal work: AxCaliber. However, AxCaliber requires knowing beforehand the predominant orientation of the axons within the analyzed volume to adjust the q-space sampling accordingly. In this work, we show that our novel sparse representation of the 3D+t dMRI signal, enables the recovery of axonal diameter distribution parameters without the need to know the axonal direction at acquisition time.

This work has been published in [61]

7.2.3. A sensitivity analysis of Q-space Indices with respect to changes in axonal diameter, dispersion and tissue composition

Participants: Rutger Fick, Marco Pizzolato, Demian Wassermann, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

In Diffusion MRI, q-space indices are scalar quantities that describe properties of the ensemble average propagator (EAP). Their values are often linked to the axonal diameter – assuming that the diffusion signal originates from inside an ensemble of parallel cylinders. However, histological studies show that these assumptions are incorrect, and axonal tissue is often dispersed with various tissue compositions. Direct interpretation of these q-space indices in terms of tissue change is therefore impossible, and we must treat them as as scalars that only give non-specific contrast – just as DTI indices. In this work, we analyze the sensitivity of q-space indices to tissue structure changes by simulating axonal tissue with changing axonal diameter, dispersion and tissue compositions. Using human connectome project data we then predict which indices are most sensitive to tissue changes in the brain. We show that, in both multi-shell and single-shell (DTI) data, q-space indices have higher sensitivity to tissue changes than DTI indices in large parts of the brain. Based on these results, it may be interesting to revisit older DTI studies using q-space indices as a marker for pathology.

This work has been published in [32]

7.2.4. Assessing the feasibility of estimating axon diameter using diffusion models and machine learning

Participants: Rutger Fick, Neda Sepasian [Eindhoven University of Technology, The Netherlands], Marco Pizzolato, Andrada Ianus [Centre for Medical Image Computing, Dept. of Computer Science, UCL, London, UK], Rachid Deriche.

Axon diameter estimation has been a focus of the diffusion MRI community for the past decade. The main argument has been that while diffusion models always overestimate the true axon diameter, their estimation still correlates with changes in true value. Until now, this remains more as a discussion point. The aim of this paper is to clarify this hypothesis using a recently acquired cat spinal cord data set, where the diffusion MRI signal of both a multi-shell and Ax- Caliber acquisition have been registered with the underlying histology values. We find that the axon diameter as estimated by signal models and AxCaliber does not correlate with their true sizes for axon diameters smaller than 3 microns. On the other hand, we also train a random forest

machine learning algorithm to map signal-based features to histology values of axon diameter and volume fraction. The results show that, in this dataset, this approach leads to a more reliable estimation of physically relevant axon diameters than using sophisticated diffusion models.

This work has been submitted to ISBI'2017.

7.2.5. Rotational Invariants of Ternary Quartics

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

This work has been developed in the framework of an "Action Transverse" with the AROMATH team (see section 9.1.1). It aims at creating building blocks for biomarkers for the case of a representation of the diffusion information (acquired using HARDI sequences) as a ternary quartic. Previous work in the team had some drawbacks such as instabilities in the non-polynomial formulae [99] or missing guarantees of the polynomial results [85] (e.g. unknown completeness or impossibility to establish the redundancy of the obtained expressions). This work proposes an alternative construction based on rational expressions and shares some of the best characteristics of the two previous approaches: the set is complete and generative – and thus also generates polynomial invariants –, the number of generators is close to minimal (13 instead of 12 and the expression relating these 13 formulae is known), and has an improved stability compared to the non-polynomial approach. The obtained formulae are furthermore nested making their computation much more effective than previous approaches. Furthermore, the method is generic and can in theory be expanded to higher polynomial degrees.

7.3. Towards microstructural based tractography

7.3.1. White matter tractography guided by anatomical and microstructural priors

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

Diffusion-weighted magnetic resonance imaging is a unique imaging modality sensitive to the microscopic movement of water molecules in biological tissues. By characterizing the movement of water molecules, it is possible to infer the macroscopic neuronal pathways of the brain. The technique, so-called tractography, had become the tool of choice to study non-invasively the human brain's white matter in vivo. For instance, it has been used in neurosurgical intervention planning and in neurodegenerative diseases monitoring. In this thesis, we report biases from current tractography reconstruction and suggest methods to reduce them. We first use anatomical priors, derived from a high resolution T1-weighted image, to guide tractography. We show that knowledge of the nature of biological tissue helps tractography to reconstruct anatomically valid neuronal pathways, and reduces biases in the estimation of complex white matter regions. We then use microstructural priors, derived from the state-of-the-art diffusionweighted magnetic resonance imaging protocol, in the tractography reconstruction process. This allows tractography to follow the movement of water molecules not only along neuronal pathways, but also in a microstructurally specific environment. Thus, the tractography distinguishes more accurately neuronal pathways and reduces reconstruction errors. Moreover, it provides the mean to study white matter microstructure characteristics along neuronal pathways. Altogether, we show that anatomical and microstructural priors used during the tractography process improve brain's white matter reconstruction

This work has been published in [12].

7.3.2. Microstructure driven tractography in the human brain

Participants: Gabriel Girard [SCIL, Sherbrooke University, CA], Alessandro Daducci [SP Lab - Laboratoire de Traitement du signal, EPFL], Kevin Whittingstall [SCIL, Sherbrooke University, CA], Rachid Deriche, Maxime Descoteaux [SCIL, Sherbrooke University, CA], Demian Wassermann.

Diffusion-weighted (DW) magnetic resonance imaging (MRI) tractography has become the tool of choice to probe the human brain's white matter (WM) in vivo. However, the relationship between the resulting streamlines and underlying WM microstructure characteristics, such as axon diameter, remains poorly understood. In this work, we reconstruct human brain fascicles using a new approach to trace WM fascicles while simultaneously characterizing the apparent distribution of axon diameters within the fascicle. This provides the mean to estimate the microstructure characteristics of fascicles while improving their reconstruction in complex tissue configurations.

This work has been published in [24].

7.3.3. Reducing Invalid Connections with Microstructure Driven Tractography

Participants: Gabriel Girard [SCIL, Sherbrooke University, CA], Kevin Whittingstall [SCIL, Sherbrooke University, CA], Alessandro Daducci [SP Lab - Laboratoire de Traitement du signal, EPFL], Jean-Philippe Thiran [SP Lab - Laboratoire de Traitement du signal, EPFL], Laurent Petit [GIN - IMN UMR 5293 CNRS CEA Université de Bordeaux], Rachid Deriche, Demian Wassermann, Maxime Descoteaux [SCIL, Sherbrooke University, CA].

Diffusion-weighted imaging (DWI) tractography has become the tool of choice to probe the human brain's white matter (WM) in vivo. However, tractography algorithms produce a large number erroneous/invalid streamlines largely due to complex ambiguous local fiber configurations (e.g. crossing, kissing or fanning). Moreover, the relationship between the resulting streamlines and the underlying WM microstructure characteristics, such as axon diameter, remains poorly understood. The distinctive aspect of our tractography algorithm from previous methods is the active use of microstructure information about fascicles during the tracking. This enables us to solve areas of complex tissue configuration and separate parallel fascicles with different microstructure characteristics, hence improving the overall tractography process.

This work has been published in [35]

7.3.4. Quantitative evaluation of Fiber Orientations Extractions

Participants: Thinhinane Megherbi [LRPE, USTHB, Alger], Gabriel Girard [SCIL, Sherbrooke University, CA], Maxime Descoteaux [SCIL, Sherbrooke University, CA], Fatima Oulebsir Boumghar [LRPE, USTHB, Alger], Rachid Deriche.

Recovering the fiber orientations in each voxel constitutes an important step for the fiber tracking algorithms. In fact, the reliability of the resulted connectivity depends on how well the local fiber orientations were extracted. Based on the tractography results we evaluated and compared different methods of fiber orientations extraction. Thus, we analyzed quantitatively the resulted connectivity by using the Tractometer tool. This later allows by measuring a number of metrics to quantify the connections reliability and the tractography performance. All the methods of fiber orientations extraction were evaluated on two types of tractography algorithms, deterministic and probabilistic algorithms. Furthermore, all of these methods have been executed on two types of data, high angular resolution data acquired with 60 gradient directions and low angular resolution data, acquired with 30 gradient directions. These two types of data were corrupted with a Ricien noise of ratio SNR=20, 10. In this work, we present the results obtained by our validation and comparison work.

This work has been published in [37]

7.4. Perfusion MRI & PLI

7.4.1. Unveiling the dispersion kernel in DSC-MRI by means of dispersion-compliant bases and control point interpolation techniques

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

In DSC-MRI the presence of dispersion affects the estimation, via deconvolution, of the residue function that characterizes the perfusion in each voxel. Dispersion is descibed by a Vascular Transport Function (VTF) which knolewdge is essential to recover a dispersion-free residue function. State-of-the-art techniques aim at characterizing the VTF but assume a specific shape for it, which in reality is unknown. We propose to estimate the residue function without assumptions by means of Dispersion-Compliant Bases (DCB). We use these results to find which VTF model better describes the in-vivo data for each tissue type by means of control point interpolation approaches.

This work has been published in [57].

7.4.2. Elucidating dispersion effects in perfusion MRI by means of dispersion-compliant bases

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

Dispersion effects in perfusion MRI data have a relevant influence on the residue function computed from deconvolution of the measured arterial and tissular concentration time-curves. Their characterization allows reliable estimation of hemodynamic parameters and can reveal pathological tissue conditions. However the time-delay between the measured concentration time-curves is a confounding factor. We perform deconvolution by means of dispersion-compliant bases, separating dispersion from delay effects. In order to characterize dispersion we introduce shape parameters, such as the dispersion time and index. We propose a new formulation for the dispersed residue function and perform in-silico experiments that validate the reliability of our approach against the block-circulant Singular Value Decomposition. We successfully apply the approach to stroke MRI data and show that the calculated parameters are coherent with physiological considerations, highlighting the importance of dispersion as an effect to be measured rather than discarded.

This work has been published in [38].

7.4.3. Improved vascular transport function characterization in DSC-MRI via deconvolution with dispersion-compliant bases

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

Bolus dispersion phenomena affect the residue function computed via deconvolution of DSC-MRI data. Indeed the obtained effective residue function can be expressed as the convolution of the true one with a Vascular Transport Function (VTF) that characterizes the dispersion. The state-of-the-art technique CPI+VTF allows to estimate the actual residue function by assuming a model for the VTF. We propose to perform deconvolution representing the effective residue function with Dispersion-Compliant Bases (DCB) without assumptions on the VTF, and then apply the CPI+VTF on DCB results. We show that DCB improve robustness to noise and allow to better characterize the VTF.

This work has been published in [60].

7.4.4. Perfusion Deconvolution in DSC-MRI with Dispersion-Compliant Bases

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

Perfusion imaging of the brain via Dynamic Susceptibility Contrast MRI (DSC-MRI) allows characterization of tissue perfusion by recovering the tissue impulse response function and scalar parameters such as the cerebral blood flow (CBF), blood volume (CBV) and mean transit time (MTT). However, the presence of bolus dispersion causes the data to reflect macrovascular properties, in addition to tissue perfusion. In this case, when performing deconvolution of the measured arterial and tissue concentration time-curves it is only possible to recover the effective, i.e. dispersed, response function and parameters. We introduce Dispersion-Compliant Bases (DCB) to represent the response function in the presence and absence of dispersion. We perform in silico and in vivo experiments, and show that DCB deconvolution outperforms oSVD and the state-of-the-art CPI+VTF techniques in the estimation of effective perfusion parameters, regardless of the presence and amount of dispersion. We also show that DCB deconvolution can be used as a pre-processing step to improve the estimation of dispersion-free parameters computed with CPI+VTF, which employs a model of the vascular transport function to characterize dispersion. Indeed, in silico results show a reduction of relative errors up to 50% for dispersion-free CBF and MTT. Moreover, the DCB method recovers effective response functions that

comply with healthy and pathological scenarios, and offers the advantage of making no assumptions about the presence, amount, and nature of dispersion.

This work has been submitted for publication in Medical Image Analysis.

7.4.5. Solving the inclination sign ambiguity in three dimensional polarized light imaging with a PDE-based method

Participants: Abib Olushola Yessouffou 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 postmortem human brain or heart at a sub-millimeter 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 submitted to ISBI'2017.

7.5. Structural Connectivity Network

7.5.1. Extracting the Core Structural Connectivity Network: Guaranteeing Network Connectedness Through a Graph-Theoretical Approach

Participants: Demian Wassermann, Dorian Mazauric [ABS Project Team], Guillermo Gallardo Diez, Rachid Deriche.

In this work, we present a graph-theoretical algorithm to extract the connected core structural connectivity network of a subject population. Extracting this core common network across subjects is a main problem in current neuroscience. Such network facilitates cognitive and clinical analyses by reducing the number of connections that need to be explored. Furthermore, insights into the human brain structure can be gained by comparing core networks of different populations. We show that our novel algorithm has theoretical and practical advantages. First, contrary to the current approach our algorithm guarantees that the extracted core subnetwork is connected Girardagreeing with current evidence that the core structural network is tightly connected. Second, our algorithm shows enhanced performance when used as feature selection approach for connectivity analysis on populations.

This work has been published in [26].

7.5.2. Groupwise Structural Parcellation of the Cortex: A Sound Approach Based on Logistic Models

Participants: Guillermo Gallardo Diez, Rutger Fick, William Wells, Rachid Deriche, Demian Wassermann.

Current theories hold that brain function is highly related with 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. 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 [33] and an extended version has been submitted to Neuroimage.

7.5.3. Efficient Population-Representative Whole-Cortex Parcellation Based on Tractography

Participants: Guillermo Gallardo Diez, Rachid Deriche, Demian Wassermann.

The human brain is arranged in areas based on criteria such as cytoarchitecture or extrinsic connectivity. Current hypotheses attribute specialized functions to several areas of this patchwork. Hence, parcellating the cortex into such areas and characterizing their interaction is key to understanding brain function. Diffusion MRI enables the exploration of physical connections through axonal bundles, namely extrinsic connectivity. Current theories hold that brain function is determined by extrinsic connectivity. However, obtaining a population-representative parcellation based on extrinsic connectivity remains challenging (Jbabdi 2013). Particularly, whole-cortex parcellation methods (Moreno-Dominguez 2014; Parisot 2015) are computationally expensive and need tuning of several parameters. Our main contribution is an efficient technique to create single-subject and population-representative parcellations based on tractography. Our method creates a dendrogram using only one parameter: the minimum size of each parcel. Then, by choosing cutting criteria, we can explore different parcellation granularities without recomputing the dendrogram. Experiments show that our parcellations are consistent within subjects with anatomical (Desikan 2006) and functional (Barch 2013) parcellations existent in the literature.

This work has been published in [34].

7.6. Clinical and Neurocognitive Applications of Diffusion MRI

7.6.1. 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], Rutger Fick, Demian Wassermann, Romain Valabregue [ICM, CENIR, Paris], Richard Levy [ICM, CENIR, Paris], Isabelle Arnulf [ICM, CENIR, Paris], Stéphane Lehericy [ICM, CENIR, Paris], 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. Here we used mean apparent propagator to investigate white matter changes in KLS and correlated diffusion changes with apathy scores. Results showed that the corpus callosum was involved in KLS during episodes and mean RTAP measures in the corpus callosum correlated with apathy scores. Results were in accordance with known motivation-based circuits involving the orbitomedial frontal cortex.

This work has been submitted to ISMRM'2017.

7.6.2. Comparison of Biomarkers in Transgenic Alzheimer Rats Using Multi-shell Diffusion MRI

Participants: Rutger Fick, Madelaine Daianu [Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, USC, USA], Marco Pizzolato, Demian Wassermann, Russel Jacobs [Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, USC, USA], Paul Thompson [Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, USC, USA], Terrence Town [Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, USC, USA], Rachid Deriche.

In this study, we assessed the evolution of diffusion MRI (dMRI) derived markers from different white matter models as progressive neurodegeneration occurs in transgenic Alzheimer rats (TgF344-AD) at 10, 15 and 24 months. We compared biomarkers reconstructed from Diffusion Tensor Imaging (DTI), Neurite Orientation Dispersion and Density Imaging (NODDI) and Mean Apparent Propagator (MAP)-MRI in the hippocampus, cingulate cortex and corpus callosum using multi-shell dMRI. We found that NODDI's dispersion and MAP-MRI's anisotropy markers consistently changed over time, possibly indicating that these measures are sensitive to age-dependent neuronal demise due to amyloid accumulation. Conversely, we found that DTI's mean

diffusivity, NODDI's isotropic volume fraction and MAP-MRI's restriction-related metrics all followed a two-step progression from 10 to 15 months, and from 15 to 24 months. This two-step pattern might be linked with a neuroinflammatory response that may be occuring prior to, or during microstructural breakdown. Using our approach, we are able to provide for the first time-preliminary and valuable insight on relevant biomarkers that may directly describe the underlying pathophysiology in Alzheimer's disease.

This work has been published in [30].

7.7. MEEG and Diffusion MRI

7.7.1. Cortical surface parcellation via dMRI using mutual nearest neighbor condition

Participants: Brahim Belaoucha, Maureen Clerc, Théodore Papadopoulo.

This work aims at parcellating the cortical surface from individual anatomy. With respact to previous works, it works for the whole brain and produces connected patches. The parcellation is obtained using the Mutual Nearest Neighbor (MNN) criterion to obtain regions with similar structural connectivity. The structural connectivity is obtained by applying a probabilistic tractography on the diffusion MRI (dMRI). Several similarity measures of connectivity are compared. The results of our method are compared to some of the atlases that can be found in the literature. We show that these atlases have lower similarity of structural connectivity than the proposed algorithm implying that the regions of the atlases may have lower functional homogeneity.

This work has been published in [27].

7.7.2. Iterative estimation of focal sources and their interactions constrained by dMRI

Participants: Brahim Belaoucha, Mouloud Kachouane, Théodore Papadopoulo.

This work aims at further exploiting the dMRI constraints: not only sources are constrained anatomically by patches (extracted by the method of the previous paragraph) but their dynamical behaviour is constrained by a brain network extracted from an individual dMRI. The framework reconstructs spatially localized sources from Magnetoencephalogra-phy (MEG)/Electroencephalography (EEG) using spatiotemporal constraints extracted from dMRI. The spatial reconstruction is based on our previous work on patch reconstruction [71]. The source dynamics are represented by a Multivariate Autoregressive (MAR) model whose matrix elements are constrained by the anatomical connectivity obtained from dMRI. The framework assumes that the whole brain dynamic follows a constant MAR model in a time window of interest. The source activations and the MAR model parameters are estimated iteratively. The proposed framework outperforms the two-stage approaches which have traditionally been used to estimate source interactions. Such approaches first reconstruct sources and then compute the MAR model for the localized sources. They showed good results when working in high signal-to-noise ratio (SNR) settings, but fail in detecting the true interactions when working in low SNR. Our framework iteratively refines both the reconstruction and the MAR model in two steps: sources activations are first estimated for a given MAR model and then, the MAR model is estimated for a given source reconstruction. These two steps are repeated until a stopping criterion is achieved. The work is exploratory in nature and for now focuses on simulations made with real MR data. We could confirm that accurate reconstructions and MAR models can be obtained with our method in both high and low noise levels.

This work has been published in [28], [21].

7.8. MEG, EEG

7.8.1. ECoG

Participants: Maureen Clerc, Analisa Pascarella [CNR-IAC Roma], Michele Piana [University of Genova].

Electrocorticography (ECoG) measures the distribution of the electrical potentials on the cortex produced by the neural currents. A full interpretation of ECoG data requires solving the ill-posed inverse problem of reconstructing the spatio-temporal distribution of the neural currents. This study addresses the ECoG source modeling developing a beamformer method. New method: We computed the lead-field matrix by using a novel routine provided by the OpenMEEG software. We performed an analysis of the numerical stability of the ECoG inverse problem by computing the condition number of the lead-field matrix for different configurations of the electrodes grid. We applied a Linear Constraint Minimum Variance (LCMV) beamformer to both synthetic data and a set of real measurements recorded during a rapid visual categorization task. For all considered grids the condition number indicates that the ECoG inverse problem is mildly ill-conditioned. For realistic SNR we found a good performance of the LCMV algorithm for both localization and waveforms reconstruction. Comparison with existing method: The flow of information reconstructed by analyzing real data seems consistent with both invasive monkey electrophysiology studies and non-invasive (MEG and fMRI) human studies. Despite a growing interest from the neuroscientific community, solving the ECoG inverse problem has not quite yet reached the level of systematicity found for EEG and MEG. Starting from an analysis of the numerical stability of the problem we considered the most widely utilized method for modeling neurophysiological data based on the beamformer method in the hope to establish benchmarks for future studies.

This work has been published in [18].

7.8.2. Conductivity estimation

Participants: Maureen Clerc, Christos Papageorgakis, Juliette Leblond [APICS project-team], Jean-Paul Marmorat [CMA Ecole des Mines Paritech].

Considering a geometry made of three concentric spherical nested layers, each with constant homogeneous conductivity, we establish a uniqueness result in inverse conductivity estimation, from partial boundary data in presence of a known source term. We make use of spherical harmonics and linear algebra computations, that also provide us with stability results and a robust reconstruction algorithm. As an application to electroencephalography (EEG), in a spherical 3-layer head model (brain, skull, scalp), we numerically estimate the skull conductivity from available data (electrical potential at electrodes locations on the scalp, vanishing current flux) and given pointwise dipolar sources in the brain. This work was supported by the Région Provence-Alpes-Côte d'Azur, France, and BESA GmbH, Germany.

This work has been published in [14] and [29].

7.8.3. Efficient lead field computation a la Reduced Basis Methods

Participants: Kostiantyn Maksymenko, Maureen Clerc, Théodore Papadopoulo.

Bioelectric source analysis in the human brain from scalp electroencephalography (EEG) signals is sensitive to geometry and conductivity properties of the different head tissues. These conductivities can vary a lot across subjects so non-invasive methods of conductivity estimation are required. To achieve this, we should have a possibility to compute a forward EEG problem solution for a large number of conductivity configurations. We propose a method of approximation of these solutions using a relatively small number of basis points, which will allow us to dramatically decrease required computing time.

7.9. Modeling electrical stimulation

7.9.1. Cochlear implants

Participants: Kai Dang, Maureen Clerc, Pierre Stahl [Oticon Medical], Dan Gnansia [Oticon Medical], Clair Vandersteen [Nice University Hospital], Nicolas Guevara [Nice University Hospital].

In cochlear implants, the hybrid common ground is a combination of a classic monopolar stimulation with a standard common ground. This new stimulation montage allows the current to return from both the non-stimulating electrodes on the electrode array and the reference electrode placed between the skull and scalp. In theory, this lead to reach a compromise between the current focality and the efficiency of the stimulation. Even if this stimulation type has already been adopted by some implant manufacturers, the 3D geometry of its current pathways remains to be studied. The study is two-fold. First, an in-vitro experiment aimed to measure the electrical field produced by the hybrid common ground stimulation. An electrode array of an XP implant (Oticon Medical, Neurelec) was placed in saline solution and the electrical field was recorded by a probe that moves along the programmed grid. During the stimulation, the current waveforms on all the grounding electrodes were also recorded. Second, an in-situ measurement was conducted. Another XP device was implanted into a human specimen. The same procedure as in the in-vitro measurement was performed to record, this time, the current waveforms only. The recorded two groups of current data were compared with each other to investigate how the current path is modified by the geometry of a human cochlea. The potential distribution measured during the in-vitro experiment was also compared with other stimulation types such as monopolar. The energy consumption of the stimulation was also computed from the collected data.

This work has been published in [58]. We thank the GRAPHDECO project-team for lending us THE 3D printer which was used in the in-vitro experiment.

7.10. Brain-Computer Interfaces

7.10.1. A new reference book in Brain-Computer Interfaces

Participants: Maureen Clerc, Laurent Bougrain [Neurosys project-team], Fabien Lotte [POTIOC project-team], All Ipl Bci-Lif Members.

Most of the results in Brain-Computer Interfaces are conducted in the context of the Inria Project-Lab BCI-LIFT (see contracts section). Researchers of this Inria Project-Lab (Maureen Clerc, Laurent Bougrain and Fabien Lotte) have edited a reference book in 2016 on the topic of Brain-Computer Intefaces. It consists of two volumes, “Foundations and Methods” (in French [54] and in English [1]) and “Technology and Applications” (in French [55] and in English [53]).

7.10.2. A Separability Marker Based on High-Dimensional Statistics for Classification Confidence Assessment

Participants: Nathalie Gayraud, Maureen Clerc, Nathanaël Foy, Alain Rakotomamonjy [University of Rouen].

This work provides a theoretical analysis framework for features that belong to the high dimensional Riemannian manifold of symmetric positive definite matrices. In non-invasive EEG-based Brain Computer Interfaces, such as the P300 speller, these are sample covariance matrices of the epoched EEG signal that are classified into two classes. An analysis of the class shape on the manifold is performed, and the separability level of the two classes is evaluated. The main contribution is the Separability Marker (SM)-confidence method, a method that appends a confidence marker to the prediction of a binary classifier whose decision function is based on the comparison of Riemannian distances.

This work has been published in [23].

7.10.3. Comparison of Hierarchical and Non-Hierarchical Classification for Motor Imagery Based BCI Systems

Participants: Nathalie Gayraud, Maureen Clerc, Cecilia Lindig-León [Neurosys project-team], Laurent Bougrain [Neurosys project-team].

Motor imagery (MI) based BCI systems record and analyze the brain activity to determine users' intentions while imagining moving some parts of their body. In order to build systems that are able to detect several commands, multiclass schemes need to be applied. Hierarchical methods allow solving multiclass problems by using a tree of binary classifiers, whose root discriminates between two groups, each one containing a half of the classes. Each succeeding node includes again only one half of the classes from the selected group, and the process is recursively repeated until each node contains a single class, from which the final decision can be inferred. In this study we compare a series of multiclass approaches to assert the benefits of hierarchical classification. The compared methods are based on two effective techniques for MI-discrimination, namely, Common Spatial Patterns (CSP) and Riemannian geometry, for which the hierarchical and non-hierarchical approaches have been considered. We include the CSP by Joint Diagonalization method (CSP by JAD), which corresponds with a non-hierarchical approach; and its hierarchical counterpart, namely, Binary CSP. In addition, the non-hierarchical Minimum Distance to Riemannian Mean method (MDRM) [4] is also evaluated, together with its analogous hierarchical approach; a contribution of the present work called Hierarchical MDRM algorithm (HMDRM). All these methods have been applied on dataset 2a of the BCI competition IV to facilitate their comparison. The highest accuracies were reached by the BCSP and HMDRM methods, confirming the effectiveness of hierarchical algorithms.

This work has been published in [36].

7.10.4. Motor imagery learning using Functional Electrical Stimulation

Participants: Maureen Clerc, Saugat Bhattacharyya [CAMIN project-team], Mitsuhiro Hayashibe [CAMIN project-team].

Functional Electrical Stimulation (FES) stimulates the affected region of the human body thus providing a neuroprosthetic interface to non-recovered muscle groups. FES in combination with Brain-computer interfacing (BCI) has a wide scope in rehabilitation because this system can directly link the cerebral motor intention of the users with its corresponding peripheral mucle activations. Such a rehabilitative system would contribute to improve the cortical and peripheral learning and thus, improve the recovery time of the patients. In this paper, we examine the effect of electrical stimulation by FES on the electroencephalography (EEG) during learning of a motor imagery task. The subjects are asked to perform four motor imagery tasks over six sessions and the features from the EEG are extracted using common spatial algorithm and decoded using linear discriminant analysis classifier. Feedback is provided in form of a visual medium and electrical stimulation representing the distance of the features from the hyperplane. Results suggest a significant improvement in the classification accuracy when the subject was induced with electrical stimulation along with visual feedback as compared to the standard visual one.

This work has been published in [13] and [22].

7.10.5. Brain training with neurofeedback

Participants: Maureen Clerc, Lorraine Perronnet [Hybrid project-team], Anatole Lécuyer [Hybrid project-team], Fabien Lotte [Potioc project-team], Christian Barillot [Visages project-team].

Neurofeedback is a promising tool for brain rehabilitation and peak performance training. Neurofeedback approaches usually rely on a single brain imaging modality such as EEG or fMRI. Combining these modalities for neurofeedback training could allow to provide richer information to the subject and could thus enable him/her to achieve faster or more specific self-regulation. Yet unimodal and multimodal neurofeedback have never been compared before. In the present work, we introduce a simultaneous EEG-fMRI experimental protocol in which participants performed a motor-imagery task in unimodal and bimodal NF conditions. With this protocol we were able to compare for the first time the effects of unimodal EEG-neurofeedback and fMRI-neurofeedback versus bimodal EEG-fMRI-neurofeedback by looking both at EEG and fMRI activations. We also propose 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). Such a feedback is intended to relieve the cognitive load of the subject by presenting the bimodal neurofeedback task as a single regulation task instead of two. Additionally, this integrated feedback metaphor gives flexibility on defining

a bimodal neurofeedback target. Participants were able to regulate activity in their motor regions in all NF conditions. Moreover, motor activations as revealed by offline fMRI analysis were stronger during EEG-fMRI-neurofeedback than during EEG-neurofeedback. This result suggests that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-neurofeedback. Our results also suggest that during EEG-fMRI-neurofeedback, participants tended to regulate more the modality that was harder to control. Taken together our results shed light on the specific mechanisms of bimodal EEG-fMRI-neurofeedback and on its added-value as compared to unimodal EEG-neurofeedback and fMRI-neurofeedback.

This work has been published in [51] and [50].

7.10.6. Out-of-the-lab P300 speller

Participants: Maureen Clerc, Théodore Papadopoulo, Nathanaël Foy, Federica Turi, Étienne Guerlais.

New developments have been performed in the context of ADT OpenViBE-X to robustify the P300 speller system, correcting some timing issues (in OpenViBE), and making the system easier to use and install. This has been validated by the use of our system out-of-the-lab, by a patient in Chambéry (see article [104]).

This work has been published in [48]and [49].

7.10.7. Clinical study with the CoAdapt P300 speller

Participants: Maureen Clerc, Théodore Papadopoulo, Marie-Hélène Soriani [Nice University Hospital], Claude Desnuelle [Nice University Hospital], Violaine Guy [Nice University Hospital], Mariane Bruno [Nice University Hospital].

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease which, a few years after onset, restricts patients' communication capacity. In this study, the usability by disabled patients with ALS of a visual Brain-Computer Interface (BCI), the P300-speller, is evaluated as a viable alternative to existing assistive communication tools. After clinical evaluation of their physical, cognitive and language capacities, 20 patients with ALS were included. The study consisted of two 3-block sessions, at 2 to 4-week interval, using a P300-speller BCI in several modes of operation in view of evaluating its usability in its 3 components: effectiveness, efficiency and satisfaction. Over all participants, the system is effective (100% of participants successfully achieved copy spelling and free spelling tasks). It is also efficient (over 95% of correct symbols were selected by 65% of participants). The average number of correct symbols/min ranged from 3.6 (without word prediction) to 5.04 (with word prediction). Participants expressed satisfaction through an average of 8.7/10 measuring comfort, ease of use and utility. Patients quickly learned how to operate this system and were able to use it with good performance without much learning effort. Word prediction and other algorithmic optimizations improve the information transfer rate and make such systems competitive with existing solutions of alternative communication such as eye trackers.

This work was published in [25].

BIOVISION Team

6. New Results

6.1. High tech vision aid systems for low vision patients

This is a new axis in the team that we started this year. We do not have results yet available but one project has started to allow real-time enhancement of environments in Virtual Reality (equipement: Samsung S6 and Samsung VR headset). This is the internship work of Alberto Patino (grant: CONACYT) who is co-supervised by Pierre Kornprobst and Fabio Solari (University of Genoa, Italy). We plan to submit an abstract to Vision 2017 conference, the 12th International Conference by the International Society for Low Vision Research and Rehabilitation.

Another project is in preparation, involving Fabio Solari (University of Genoa, Italy) and other colleagues from Université Cote d'Azur. New results are expected in 2017.

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

6.2.1. Cells characterization from their spike response

6.2.1.1. A new nonconvex variational approach for sensory neurons receptive field estimation

Participants: Audric Drogoul, Gilles Aubert [UCA, Laboratoire Jean Alexandre Dieudonné, Nice, France], Bruno Cessac, Pierre Kornprobst.

Determining the receptive field of a visual sensory neuron is a first but crucial step to- wards the characterization of neurons response to local spatio-temporal stimuli. Existing methods are based on convex optimization methods neglecting biophysical constraints of neurons (bounded firing rate), and they are relatively poor in terms of accuracy and running time. We propose a new method to estimate receptive fields by a nonconvex variational approach, thus relaxing the simplifying and unrealistic assumption of convexity made by standard approaches. The method consists in studying a relaxed discrete energy minimized by a proximal alternating minimization algorithm. We compare our approach with the classical spike-triggered-average technique on simulated data, considering a typical retinal ganglion cell. Results show a high improvement in terms of accuracy and convergence with respect to the duration of the experiment.

This work was presented in [29], [21] and has been submitted, see [24] .

6.2.1.2. Pan-retinal characterization of Light Responses from Ganglion Cells in the Developing Mouse Retina

Participants: Gerrit Hilgen [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Sarah Pirmoradian [ANC - Institute for Adaptive and Neural Computation, Edimburgh, UK], Daniela Pamplona, Pierre Kornprobst, Bruno Cessac, Matthias Hennig Pirmoradian [ANC - Institute for Adaptive and Neural Computation, Edimburgh, UK], Evelyne Sernagor [Institute of Neuroscience, Medical School, Newcastle University, 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 behavior, 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 is under revision, submitted to EScience [25]

6.2.2. Understanding the role of spatio-temporal correlations in visual scene encoding

6.2.2.1. Spike train analysis and Gibbs distributions

Participants: Bruno Cessac, Rodrigo Cofré [Département de Physique Théorique, Université de Genève].

Spikes in sensory neurons are conveyed collectively to the cortex using correlated binary patterns (in space and time) which constitute “the neural code”. Since patterns occur irregularly it is appropriate to characterize them using probabilistic descriptions or statistical models. Two major approaches attempt to characterize the spike train statistics: The Maximum Entropy Principle (MaxEnt) and Neuronal Network modeling (N.N). Remarkably, both approaches are related via the concept of Gibbs distributions. MaxEnt models are restricted to time-invariant Gibbs distributions, via the underlying assumption of stationarity, but this concept extends to non-stationary statistics (not defined via entropy), allowing to handle as well statistics of N.N models and GLM with non-stationary dynamics. We show in this poster that, stationary N.N, GLMmodels and MaxEnt models are equivalent via an explicit mapping. This allows us, in particular, to interpret the so-called "effective interactions" of MaxEnt models in terms of “real connections” models.

This work was presented in the Bernstein Conference 2016 [28] and will be soon submitted to Journal of Statistical Physics.

6.2.2.2. Dimensionality Reduction in spatio-temporal MaxEnt models and analysis of Retinal Ganglion Cell Spiking Activity in experiments

Participants: Rubén Herzog [CINV - Centro Interdisciplinario de Neurociencia de Valparaíso], Maria-Jose Escobar [Univ Tecnico Federico Santa María], Adrian Palacios [CINV - Centro Interdisciplinario de Neurociencia de Valparaíso], 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 model 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 [1]. To resolve this issue we have considered spatio-temporal MaxEnt model formerly developed e.g. by Vasquez et al. [2]. 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 handle 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 presented in the Bernstein 2016 conference [31] and has been submitted to Plos Comp Bio.

6.2.2.3. On the mathematical consequences of binning spike trains

Participants: Bruno Cessac, Arnaud Le Ny [LAMA - Laboratoire d’Analyse et de Mathématiques Appliquées], Eva Loecherbach [AGM - Laboratoire d’Analyse, Géométrie et Modélisation and Département de Mathématiques, [Cergy-Pontoise].

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 Neural Computation, Massachusetts Institute of Technology Press (MIT Press), 2016 [16].

6.2.3. Retinal waves

6.2.3.1. Mathematical and experimental studies on retinal waves

Participants: Dora Karvouniari, Lionel Gil [INLN -Institut Non Linéaire de Nice Sophia-Antipolis], Olivier Marre [Institut de la Vision], Serge Picaud [Institut de la Vision], Bruno Cessac.

We reproduce the spontaneous intrinsic cell-autonomous rhythmic bursting in Starbust Amacrine Cells (SACs) and the slow After Hyperpolarisation Current (sAHP), which modulates the refractory process inbetween two consecutive bursts, observed experimentally in [85]. We describe the dynamical influence of cholinergic synapses, ensuring the level of SAC synchrony necessary for the emergence of waves. We obtain: a) a plausible generic mechanism generating spontaneous retinal waves in development, without any need for external stimulation as opposed to existing models and b) a mathematical characterization of retinal waves. Especially, a biophysical parameter controls the wave arousal and the corresponding shape. The model is accurate enough to reproduce existing experiments, but also to propose new ones.

This work has been presented in the workshop "Modelling the early visual system" [32], 2nd International Conference on Mathematical Neuroscience (ICMNS) [22], the AREADNE conference [34], the Bernstein conference [33]. Two papers are in preparation.

6.2.4. Trajectory anticipation, from retina to V1

This work is just starting. The main work has been done by Selma Souihel in her Master II instership supervised by Bruno Cessac [36]. The aim of the internship is to use and update the software VirtualRetina and Enas in order to reproduce the activity of the retina in response to the stimulus of a moving bar, observed By Mr Berry & al. A form of anticipation of the movement has been demonstrated experimentally by its authors in salamander, rabbit and goldfish retinas. This anticipation can be explained, in the case of a simple trajectory, by the gain control mechanism specific to the ganglion cells, implemented by Virtual-Retina-Enas.

6.2.5. Simulating and analysing retina's response to visual stimuli

6.2.5.1. ENAS: A new software for spike train analysis and simulation

Participants: Bruno Cessac, Pierre Kornprobst, Selim Kraria, Hassan Nasser, Daniela Pamplona, Geoffrey Portelli, Thierry Vieville [Mnemosyne - Mnemonic Synergy LaBRI - Laboratoire Bordelais de Recherche en Informatique, IMN - Institut des Maladies Neurodégénératives, [Bordeaux].

This work, presenting the Enas-Virtual Retina platform has been presented in [27] and submitted to Frontiers in Neuroinformatics [3].

6.2.5.2. Rank order coding: a retinal information decoding strategy revealed by large-scale multielectrode array retinal recordings

Participants: Geoffrey Portelli, John M. Barrett [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Gerrit Hilgen [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Timothée Masquelier [CERCO, Toulouse, France], Alessandro Maccione [NetS3 Lab - Neuro-Engineering & bio-arTificial Synergic SystemS Laboratory, Genova, Italy], Stefano Di Marco [NetS3 Lab - NeuroEngineering & bio-arTificial Synergic SystemS Laboratory, Genova, Italy], Luca Berdondini [NetS3 Lab - NeuroEngineering & bio-arTificial Synergic SystemS Laboratory, Genova, Italy], Pierre Kornprobst, Evelyne Sernagor [Institute of Neuroscience, Medical School, Newcastle University, Newcastle, UK].

How a population of retinal ganglion cells (RGCs) encodes the visual scene remains an open question. Going beyond individual RGC coding strategies, results in salamander suggest that the relative latencies of an RGC pair encodes spatial information. Thus a population code based on this concerted spiking could be a powerful mechanism to transmit visual information rapidly and efficiently. Here, we tested this hypothesis in mouse by recording simultaneous light-evoked responses from hundreds of RGCs, at pan-retinal level, using a new generation of large-scale, high density multielectrode array consisting of 4096 electrodes. Interestingly, we did not find any RGCs exhibiting a clear latency tuning to the stimuli, suggesting that in mouse, individual RGC pairs may not provide sufficient information. We show that a significant amount of information is encoded synergistically in the concerted spiking of large RGC populations. Thus, the RGC population response described with relative activities, or ranks, provides more relevant information than classical independent spike count- or latency- based codes. In particular, we report for the first time that when considering the relative activities across the whole population, the wave of first stimulus-evoked spikes (WFS) is an accurate indicator of stimulus content. We show that this coding strategy co-exists with classical neural codes, and that it is more efficient and faster. Overall, these novel observations suggest that already at the level of the retina, concerted spiking provides a reliable and fast strategy to rapidly transmit new visual scenes.

This work has been published in eNeuro [20].

6.2.5.3. Microsaccades enable efficient synchrony-based coding in the retina: a simulation study.

Participants: Timothée Masquelier [CERCO, Toulouse, France], Geoffrey Portelli, Pierre Kornprobst.

It is now reasonably well established that microsaccades (MS) enhance visual perception, although the underlying neuronal mechanisms are unclear. Here, using numerical simulations, we show that MSs enable efficient synchrony-based coding among the primate retinal ganglion cells (RGC). First, using a jerking contrast edge as stimulus, we demonstrate a qualitative change in the RGC responses: synchronous firing, with a precision in the 10 ms range, only occurs at high speed and high contrast. MSs appear to be sufficiently fast to be able reach the synchronous regime. Conversely, the other kinds of fixational eye movements known as tremor and drift both hardly synchronize RGCs because of a too weak amplitude and a too slow speed respectively. Then, under natural image stimulation, we find that each MS causes certain RGCs to fire synchronously, namely those whose receptive fields contain contrast edges after the MS. The emitted synchronous spike volley thus rapidly transmits the most salient edges of the stimulus, which often constitute the most crucial information. We demonstrate that the readout could be done rapidly by simple coincidence-detector neurons without knowledge of the MS landing time, and that the required connectivity could emerge spontaneously with spike timing-dependent plasticity.

This work has been published in Scientific Reports [17].

6.2.6. Mean-Field models in neuroscience

6.2.6.1. Perspectives on Multi-Level Dynamics

Participants: Fatihcan Atay [MPI-MIS - Max Planck Institute for Mathematics in the Sciences], Sven Banisch [MPI-MIS - Max Planck Institute for Mathematics in the Sciences], Philippe Blanchard [University of Bielefeld-Departement of physics], Bruno Cessac, Eckehard Olbrich [MPI-MIS - Max Planck Institute for Mathematics in the Sciences], Dimitri Volchenkov [University of Bielefeld, Departement of physics].

As Physics did in previous centuries, there is currently a common dream of extracting generic laws of nature in economics, sociology, neuroscience, by focalising the description of phenomena to a minimal set of variables and parameters, linked together by causal equations of evolution whose structure may reveal hidden principles. This requires a huge reduction of dimensionality (number of degrees of freedom) and a change in the level of description. Beyond the mere necessity of developing accurate techniques affording this reduction, there is the question of the correspondence between the initial system and the reduced one. In this paper, we offer a perspective towards a common framework for discussing and understanding multi-level systems exhibiting structures at various spatial and temporal levels. We propose a common foundation and illustrate it with examples from different fields. We also point out the difficulties in constructing such a general setting and its limitations.

This work has been published in The interdisciplinary journal of Discontinuity, Nonlinearity, and Complexity, 2016, 5 [15].

6.2.7. Motion perception

6.2.7.1. The relative contribution of noise and adaptation to competition during tri-stable motion perception

Participants: Andrew Isaac Meso [Institut de Neurosciences de la Timone, Team InVibe, France], James Rankin [Center for Neural Science, New York UniversityNew York, NY], Pierre Kornprobst, Olivier Faugeras [Université Côte d'Azur, Inria, MathNeuro team, France], Guillaume S. Masson [Institut de Neurosciences de la Timone, Team InVibe, France].

Animals exploit antagonistic interactions for sensory processing and these can cause oscillations between competing states. Ambiguous sensory inputs yield such perceptual multistability. Despite numerous empirical studies using binocular rivalry or plaid pattern motion, the driving mechanisms behind the spontaneous transitions between alternatives remain unclear. In the current work, we used a tristable barber pole motion stimulus combining empirical and modeling approaches to elucidate the contributions of noise and adaptation to underlying competition. We first robustly characterized the coupling between perceptual reports of transitions and continuously recorded eye direction, identifying a critical window of 480 ms before button presses, within which both measures were most strongly correlated. Second, we identified a novel nonmonotonic relationship between stimulus contrast and average perceptual switching rate with an initially rising rate before a gentle reduction at higher contrasts. A neural fields model of the underlying dynamics introduced in previous theoretical work and incorporating noise and adaptation mechanisms was adapted, extended, and empirically validated. Noise and adaptation contributions were confirmed to dominate at the lower and higher contrasts, respectively. Model simulations, with two free parameters controlling adaptation dynamics and direction thresholds, captured the measured mean transition rates for participants. We verified the shift from noise-dominated toward adaptation-driven in both the eye direction distributions and intertransition duration statistics. This work combines modeling and empirical evidence to demonstrate the signal-strength–dependent interplay between noise and adaptation during tristability. We propose that the findings generalize beyond the barber pole stimulus case to ambiguous perception in continuous feature spaces.

This work is a a continuation of former paper [72], [12] and has been published in Journal of Vision [19].

6.2.7.2. Understanding the impact of recurrent interactions on MT population tuning: a simulation study.

Participants: Kartheek Medathati, Andrew Isaac Meso [Institut de Neurosciences de la Timone, Team InVibe, France], Guillaume S. Masson [Institut de Neurosciences de la Timone, Team InVibe, France], Pierre Kornprobst, James Rankin [Center for Neural Science, New York University, USA].

In sensory systems, different computational rules are often evident in different neuronal subpopulations. Most previous models of motion estimation by MT cells explain their specific tuning functions by having multiple feedforward inputs, largely ignoring the role of recurrent connectivity, a hallmark of cortical circuits. Therefore they fail to explain the dynamics of these tuning functions and the fact that different behaviour can be achieved by a single subpopulation when varying the spatiotemporal properties of the input. Here, using numerical simulations, we focus on a ring network that models visual motion processing at the level of MT cells. We show how excitatory and inhibitory recurrent connections shape motion direction tuning, thus resulting in different computational rules such as vector averaging, winner-take-all or bimodal representations.

In particular, depending on the inhibition regime the ring network can switch from motion integration to motion segmentation, being able to compute either a single pattern motion or to superpose multiple inputs as in motion transparency. Such feature space centre-surround recurrent mechanisms may be widely applicable to explain context-modulation of sensory processing.

This work has been presented at AREADNE conference [35] and a paper is in preparation.

6.2.8. Bio-Inspired Computer Vision

6.2.8.1. Bio-Inspired Computer Vision: Towards a Synergistic Approach of Artificial and Biological Vision

Participants: Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone, Team InVibe], Kartheek Medathati [correspondent], Heiko Neumann [Ulm University, Germany].

Studies in biological vision have always been a great source of inspiration for design of computer vision algorithms. In the past, several successful methods were designed with varying degrees of correspondence with biological vision studies, ranging from purely functional inspiration to methods that utilise models that were primarily developed for explaining biological observations. Even though it seems well recognised that computational models of biological vision can help in design of computer vision algorithms, it is a non-trivial exercise for a computer vision researcher to mine relevant information from biological vision literature as very few studies in biology are organised at a task level.

In [26], we aim to bridge this gap by providing a computer vision task centric presentation of models primarily originating in biological vision studies. Not only we revisit some of the main features of biological vision and discuss the foundations of existing computational studies modelling biological vision, but also consider three classical computer vision tasks from a biological perspective: image sensing, segmentation and optical flow. Using this task-centric approach, we discuss well-known biological functional principles and compare them with approaches taken by computer vision. Based on this comparative analysis of computer and biological vision, we present some recent models in biological vision and highlight a few models that we think are promising for future investigations in computer vision. To this extent, this paper provides new insights and a starting point for investigators interested in the design of biology-based computer vision algorithms and pave a way for much needed interaction between the two communities leading to the development of synergistic models of artificial and biological vision.

This work has been published in Computer Vision and Image Understanding Journal (CVIU) [9].

6.2.8.2. Retina-inspired tone mapping

Participants: Marco Benzi, Maria-Jose Escobar [Universidad Técnica Federico Santa María, Valparaíso, Chile], Adrien Bousseau [Inria, GraphDeco project-team], Pierre Kornprobst [correspondent].

Real-world radiance values span several orders of magnitudes which have to be processed by biological and artificial systems in order to maintain high visual sensitivity.

In biological systems, process starts at the retina level, where adaptation is absolutely crucial since retinas must maintain high contrast sensitivity over a very broad range of luminance, from starlight to direct sunlight. Adaptation is both global through neuromodulatory feedback loops and local through adaptive gain control mechanisms so that retinal networks can be adapted to the whole scene luminance level while maintaining high contrast sensitivity in different regions of the image, despite their considerable differences in luminance. Adaptation is present at different levels, e.g., at the photoreceptor level where sensitivity is a function of the recent mean intensity, and at the bipolar level where slow and fast contrast adaptation mechanisms are found. These multiple adaptational mechanisms act together, with lighting conditions dictating which mechanisms dominate.

In artificial systems, the process of compressing the range of intensities in High-Dynamic Range (HDR) images is know as tone mapping. It is a necessary step to properly visualize captured natural scenes as common displays are Low-Dynamic Range, spanning up to two orders of magnitude. There is a large body of literature in this area on static images, with approaches which combine luminance adaptation (using empirical laws such as the Naka-Rushton equation) and local contrast enhancement sometimes closely inspired from retinal principles [43], [61]. Recent developments concern video-tone mapping where a few approaches have been developed [49].

In this work, we investigate if the Virtual Retina simulator [14] could serve as a goof basis to develop a new tone mapping operator for videos. One strength of this simulator is its model of fast contrast gain control which has been validated on experimental data. However this model was not designed to deal with color and HDR images. This requires some pre- and post-processing but also changes in the Virtual Retina to account for other adaptation phenomena. Preliminary encouraging results have been obtained and we plan to continue that project in 2017.

CAMIN Team

7. New Results

7.1. Movement analysis and interpretation

7.1.1. Inertial Sensor based Analysis of Gait for Children with Cerebral Palsy

Participants: Christine Azevedo Coste, Benoît Sijobert, Jessica Rose [Stanford University].

Analysis of walking abnormalities is important for clinical diagnosis, to guide treatments, and to assess treatment outcomes for gait disorders particularly in children with cerebral palsy (CP). Motion capture, the current gold standard, enables practitioners to perform gait analyses with high accuracy in the laboratory. However, the motion capture technology used is constrained to a small space, the clinical environment may not be relevant to community mobility. This research collaboration investigated the development of a mobile systems using light-weight inertial measurement units (IMU). These sensor-based systems have potential to provide a more efficient, mobile alternative for movement analysis and can offer real-time feedback to patients for more effective rehabilitation. This interdisciplinary collaboration with Professor Jessica Rose, from the Department of Orthopedic Surgery at Stanford University aims to quantitatively assess walking problems associated with CP and related neurological conditions. Despite their small size, ease-of-use, robust design and low-cost, there are numerous recognized technical issues that make the use of IMUs relatively complex moreover in children. Through a series of experiments we leveraged our complementary skills to propose an IMU sensor system and software to extract meaningful gait parameters for rehabilitation of children with CP. A feasibility study was achieved at the Lucile Packard Children's Hospital Motion & Gait Lab in order to solve technical issues and refine calculations validated based on walking patterns recorded by Laboratory-based 3D motion capture data.

7.1.2. Automatic Human Movement Assessment with Switching Linear Dynamic System: Motion Segmentation and Motor Performance

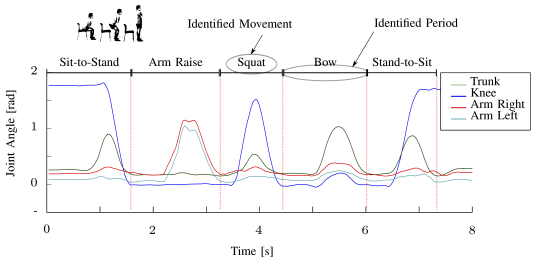
Participants: Baptista Roberto [Universidade de Brasilia, Brasil], Bo Antonio P.l. [Universidade de Brasilia, Brasil], Mitsuhiro Hayashibe.

Performance assessment of human movement is critical in diagnosis and motor-control rehabilitation. Recent developments in portable sensor technology enable clinicians to measure spatiotemporal aspects to aid in the neurological assessment. However, the extraction of quantitative information from such measurements is usually done manually through visual inspection.

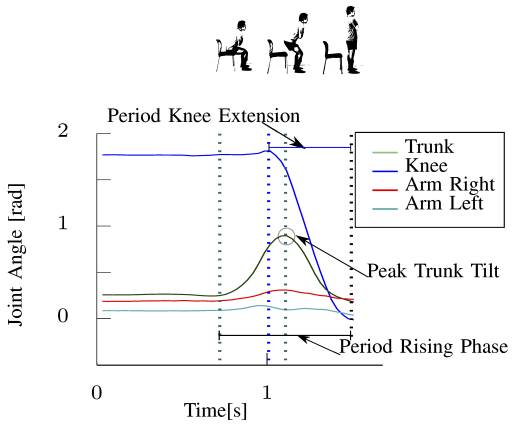
This work presents a novel framework for automatic human movement assessment that executes segmentation and motor performance parameter extraction in time-series of measurements from a sequence of human movements. We use the elements of a Switching Linear Dynamic System model as building blocks to translate formal definitions and procedures from human movement analysis. Our approach provides a method for users with no expertise in signal processing to create models for movements using labeled dataset and latter use it for automatic assessment.

Preliminary tests were carried out involving six healthy adult subjects that executed common movements in functional tests and rehabilitation exercise sessions, such as sit-to-stand and lateral elevation of the arms. Also five elderly subjects, two of which with limited mobility, that executed the sit-to-stand movement. The proposed method worked on random motion sequences for the dual purpose of movement segmentation (accuracy of 72-100%) and motor performance assessment (mean error of 0-12%).

The results of this work have been accepted for publication in the journal IEEE Transactions in Neural Systems and Rehabilitation Engineering.



Movement sequence segmentation and movement type recognition



Motor performance parameters extraction

Figure 10. Dual purpose of the proposed approach: movement segmentation and movement assessment.

7.1.3. 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 other gait events such as stride length or dorsiflexion speed at heel on. A maximum of 30 subjects will be included in this experimental protocol. Equipped with motion capture targets on which an inertial sensor is set (Figure 11), subjects have to perform an experimental path on a gait carpet. EMG recordings are also performed to monitor and evaluate fatigue. In further works, algorithms from inertial data developed through these study will enable us to evolve toward close loop control, putting together inertial sensors and programmable stimulator in real time ([39]).

7.2. Modeling and identification of the sensory-motor system

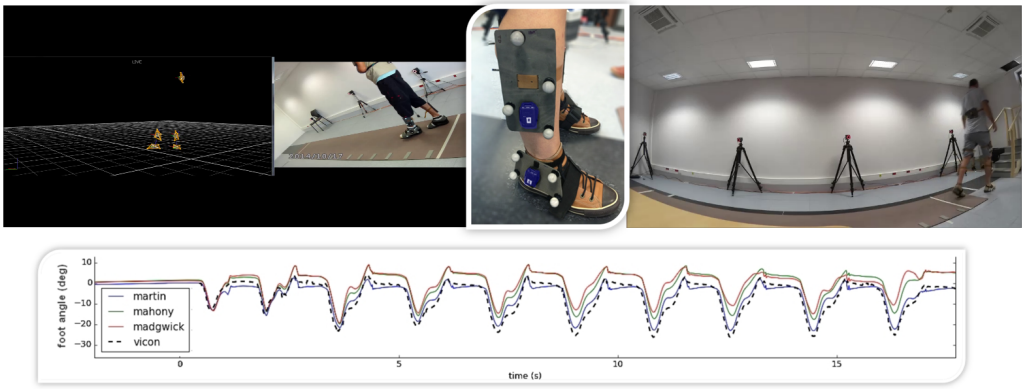


Figure 11. Gait analysis after stroke

7.2.1. Neuroplasticity and recovery in remote (sub)cortical structures following wide-awake surgery of infiltrative low-grade gliomas: investigation of fMRI and EEG signals by standard and nonlinear methods

Participants: Anthony Boyer, Jérémy Deverdun [CHU Montpellier], Hugues Duffau [CHU Montpellier], Emmanuelle Le Bars [CHU Montpellier], Sofiane Ramdani [LIRMM], David Guiraud, Nicolas Menjot de Champfleur [CHU Montpellier], François Bonnetblanc.

Wide-awake surgery of brain 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.

In a first attempt to better understand the direct consequences of wide-awake surgery we focused on the thalamus insofar as, topologically, it is the largest input source and output target of the cortex. It plays a major role in corticosubcortical and corticocortical interactions and is expected to be heavily impacted by the tumour removal while being essential to the recovery process. Studying the thalamus, based on its very particular anatomical properties, could provide essential indications regarding the behaviour of cortical and subcortical centers.

We carried out Amplitude of Low-Frequency fluctuations and Regional Homogeneity analyses on resting state fMRI data before and after the tumour removal, including an original 24h postoperative acquisition. We intended to assess possible changes in spontaneous neuronal activity over time, characterizing different facets of slow-wave hemodynamic fluctuations. We particularly sought evidences of disrupted and atypical neuronal activity emerging within deafferented thalamic subterritories.

This work revealed significant alterations of neuronal activity within distinct thalamic territories, in accordance with its neuro-anatomo-functional organization. We showed a transient decrease of neuronal activity intensity and homogeneity within the ipsilesional thalamus directly related with the anatomical dee- and deafferentation induced by the neurosurgery and a concomitant increase of neuronal activity and temporal synchrony in homologous regions of the contralesional thalamus, leading to a significant interhemispheric imbalance during the immediate postoperative period. Evidences of diaschisis-like phenomenon primarily affecting higher order thalamic nuclei of the ipsilesional thalamus and the extensive involvement of the contrale-

sional thalamus in the postoperative period promote the thesis of transient diaschisis-induced contralesional compensation for patients who underwent wide-awake surgery (Figure 12).

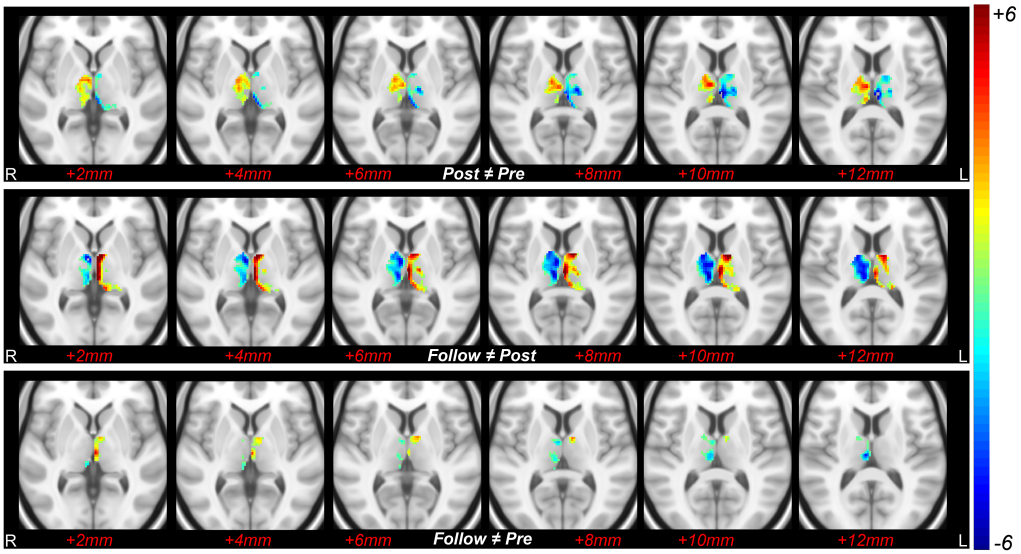


Figure 12. Voxelwise differences in ALFF score over time: ALFF maps were grouped depending on the acquisition date 1_Pre (-48h), 2_Post (+24h), 3_Follow (+3 months) and between-groups contrasts were generated as voxel-wise two-sample t-tests in order to highlight significant differences in scores over time (Post ≠ Pre, Follow ≠ Post and Follow ≠ Pre). Neuroradiological convention.

7.2.2. Understanding the effects of direct electrical stimulation of the brain during wide awake surgery

Participants: Marion Vincent, François Bonnetblanc, David Guiraud, Hugues Duffau, Mitsuhiro Hayashibe, Olivier Rossel.

Real-time functional mapping of the brain combined with direct electrical stimulation (DES) has been widely recommended for the awake neurosurgery of slow-growing and infiltrative brain tumors, to guide the resection [53]. Intra-operative DES is generally applied at 60 Hz in Europe (50 Hz in some other countries) (biphasic stimuli, single pulse duration 1 ms, intensity from 2 to 6 mA under local anesthesia, and during 1 to 5 s). By generating transient perturbations, it allows the real-time identification of both cortical areas and sub-cortical white matter pathways that are essential for the function. Its use lowers the probability of resecting essential functional areas near or within the tumor. However, the electrophysiological effects of DES remain poorly understood, locally and at a more remote distance [36], [9].

The investigation of this topic requires the recording of evoked potential. DES can be used to probe the spatio-temporal connectivity and dynamics of short- or long-range networks in vivo and in real time when combined with electrophysiological recordings (e.g. electroencephalography (EEG) or electroencephalography (ECoG)). This approach has been used for pre-surgical planning of drug-resistant epileptic patients by using an ECoG grid implanted at the surface of the grey matter. Matsumoto et al. [55] sought to measure in vivo connectivity with DES (rather than studying its propagation) but observed that a low-frequency cortical application of DES (1 Hz, constant current, and alternating rectangular wave pulses of 0.3 ms, with an intensity around 10-12 mA) induces 'cortico-cortical' evoked potentials (CCEPs) around 10-50 ms after stimulation. These properties are incompatible with the detection of EPs during awake brain surgery, when DES is classically

applied at 60 Hz due to stimulation artefacts. Conversely, 100 ms (i.e. a frequency of 10 Hz) seems to be a sufficient time-window that facilitates real-time averaging to detect these CCEPs for further on-line analysis of brain connectivity during the surgery.

In addition, in the studies mentioned above, ECoG signals were recorded in a classical common mode (CM) configuration, i.e. the signal was measured between each channel of interest and a reference electrode. Also, in all this literature, CCEPs were measured by averaging a large set of trials together. This off-line averaging actually prevents the use of ECoG recording to monitor the evoked potentials on-line. Recently, by lowering the DES frequency to 10 Hz and by using a differential recording mode (DM) for ECoG signals, in which the signal is measured between two adjacent electrodes, we were able to record for the first time on-line CCEPs easily with a standard current amplitude of stimulation (2 ms) and without averaging the data [41] (Figure 13 .

Recording ECoG in a DM enabled increasing the focality and the signal to noise ratio of the raw data. Ongoing experiments on new patients corroborate the reproducibility of this protocol. This unusual way of recording ECoG could improve the spatial resolution of the recordings in the three dimensions (in surface and in depth). Moreover, this method was used under general anesthesia but could also be performed on-line during the awake surgery. It would enable the investigation of the connectivity and to probe directly rapid plastic changes of cortical excitability.

7.2.3. A study on the effect of electrical stimulation as a user stimuli for motor imagery classification in Brain-Machine Interface

Participants: Saugat Bhattacharyya, Maureen Clerc, Mitsuhiro Hayashibe.

Functional Electrical Stimulation (FES) provides a neuroprosthetic interface to non-recovered muscle groups by stimulating the affected region of the human body. FES in combination with Brain-machine interfacing (BMI) has a wide scope in rehabilitation because this system directly links the cerebral motor intention of the users with its corresponding peripheral muscle activations. In this paper, we report the preliminary results of the effect of electrical stimulation during a motor imagery training task on healthy subjects and its comparison with visual stimuli.

The experiment designed for this work is divided into three sessions: only visual, only FES and both visual-FES stimuli. The sessions consist of instructing the subjects through a sequence of repetitive stimuli to execute the corresponding motor imagery task, which in our case, is left and right hand movement. The FES session is similar to the visual one except in place of the arrows, stimulation is directly induced in the fore-arm of the hand of interest, without providing any visual information. In the Visual-FES session, both the combined stimulations are time-synchronized to each other. After acquisition, the incoming raw EEG signal is band-pass filtered at 8-30 Hz. Then, common spatial filters (CSP) is applied to extract features relevant to left- and right-motor hand movement EEG signals. CSP is a spatial filter widely used in BMI because the spatial patterns contain highly discriminative features between two classes. In this study, we prepare the feature vectors using 6 spatial filters which is then transferred as inputs to a linear discriminant analysis (LDA) classifier. Finally, the classifier detects the corresponding motor intention of the subject, i.e., left and right motor movement. A block diagram of our experimental setup during Visual-FES session is illustrated in Fig. 14 .

Classification results shows a significant rise in accuracy for 2 (of 3) subjects which suggest a positive influence of FES during motor imagery training of the subjects. It was noted that both the subjects had no previous experience on BMI, then they were not familiar with generating motor imagery with visual stimuli. Visual stimuli are the widely accepted form of motor training but the subject requires constant training to reach an optimal result. Based on the results of this study, we can infer that electrical stimulation can also be used for motor training and it can potentially provide better performance as it can make natural proprioceptive feedback related to motor performance than visual stimuli which requires user's recognition regarding the visual cue.

7.2.4. A Study on the Effect of Electrical Stimulation During Motor Imagery Learning in Brain-Computer Interfacing

Participants: Saugat Bhattacharyya, Maureen Clerc, Mitsuhiro Hayashibe.

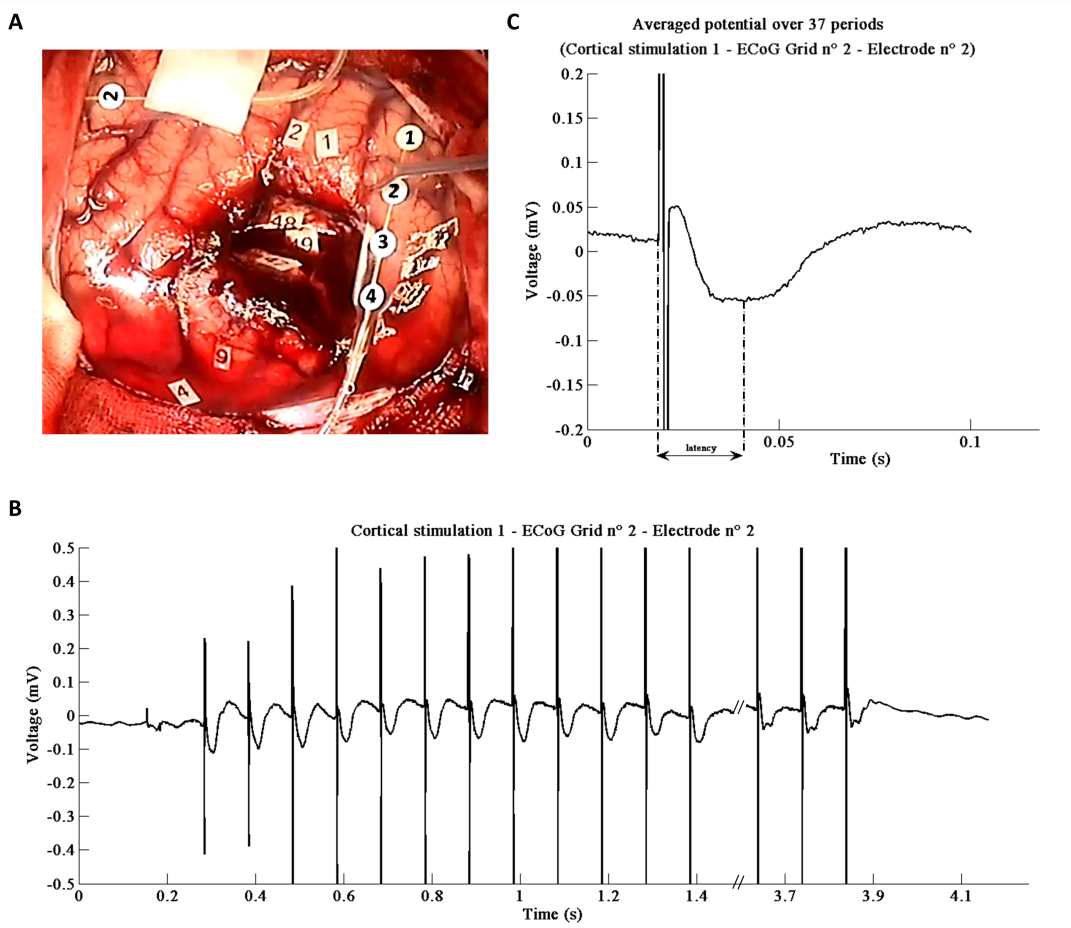


Figure 13. A: DES is applied cortically near the second electrode of the ECoG strip 2 during 3.7 s. B: Magnified view of the ECoG signal corresponding to the stimulation (with the 104 gain). CCEPs can be observed after each stimulation artefact. The last CCEPs are distorted due to the amplifier response. When the amplified signal exceeds the [-5 ; +5] volts range, an oscillation appears. C: Mean CCEP over 37 stimuli, with a latency of 23.7 $\pm$ 0.78 ms

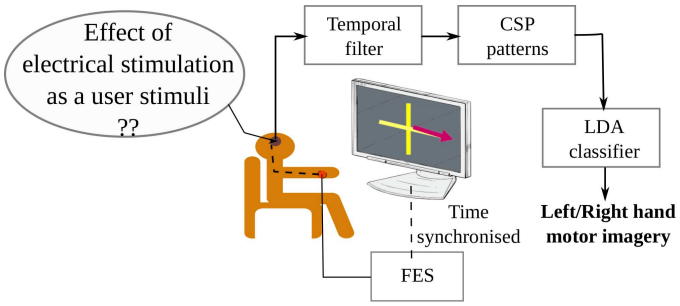


Figure 14. Block diagram of our experimental setup during motor imagery sessions where user stimuli is with conventional visual stimuli, electrical stimulation stimuli or the combined, respectively.

Functional Electrical Stimulation (FES) stimulates the affected region of the human body thus providing a neuroprosthetic interface to non-recovered muscle groups. FES in combination with Brain-computer interfacing (BCI) has a wide scope in rehabilitation because this system can directly link the cerebral motor intention of the users with its corresponding peripheral mucle activations. Such a rehabilitative system would contribute to improve the cortical and peripheral learning and thus, improve the recovery time of the patients.

To date, in BCI experiments feedback is commonly provided to the subject by means of a visual medium. On observing the feedback, the subject would attempt to perform his task. It is an interesting notion if one includes electrical stimulation to help in augmenting the performance of the motor task at hand. Thus, in this paper, we report the preliminary results of the effect of electrical stimulation on the learning of the subject during a motor imagery training task on healthy subjects. Through this study, we aim at employing FES as a proprioceptive feedback to the subject to improve the learning of the subject both in terms of accuracy and time.

In this experiment the participants performed four motor tasks: left hand movement, right hand movement, left foot movement and right foot movement across 6 separate sessions. A session provides instructions to the participant through a sequence of visual cues to execute one of the four motor tasks and each visual cue is termed as '*trial*'. Further, for data analysis, each trial are separated into time windows, termed as *epochs*. Each session consists of a feedback session provided visually to the participant at each trial, quantified by the hyperplane distance of the decoder. Before the start of the experiment, the participants undergo a training session for decoder training and to acclimatize to the tasks. Common Spatial Patterns is employed as features which is given as inputs to the Linear Discriminant decoder. The decoder designed in this work is a 2-level hierarchy. The first level classifies between left and right motor imagery and the second level discriminates between hand and foot motor imagery. In 3 of the 6 sessions, surface electrical stimulation (ES) is transmitted to the subject during the feedback period to aid the participant in performing the task. Thus, in this paper, we named the ES induced sessions as FES sessions and the sessions with only visual feedback as VIS sessions.

We report the learning during FES and VIS session feedback for each trial. For this purpose, we measure the distance of the feature vector from the hyperlane for each epoch updated at every 0.125 seconds. We took this parameter to study the feedback effect because the larger the distance from the hyperplane, the higher is the confidence of the classifier to detect the right output. The average feedback curve for all the correctly classified trials of both FES (in blue) and VIS (in red) are shown in Fig. 15 . From the curves we assume that greater the slope of the curve, faster is the learning demonstrated by the subject. Subject 1 demonstrates an increasing learning effect (greater slope) for FES feedback for all the limbs, except Right foot as compared to the VIS feedback. The figures for Subject 2 illustrates a more prominent learning effect during FES feedback and it is clearly differentiable for VIS feedback even though Subject 1 showcased a higher increase in accuracy across trials than Subject 2. It is also noted from the figures of both the subjects that VIS feedback has a frequent

increasing and decreasing trend of the curve. Subject 3 had a decrease in accuracy during FES feedback as compared to the VIS feedback which can be validated from the figure that the discriminability between the FES and VIS feedback are not as prominent in comparison to the other subjects. We can infer from these results that the electrical stimulation had a positive influence during motor task learning and with an increase in sessions one can assume ES to provide a faster learning. The steady increase of learning during FES sessions can be attributed to the fact that the subjects reported to be more motivated to perform the tasks when an ES was provided and they felt the inclusion of ES helped in their imagination. On the other hand, during VIS sessions the subjects reported to lose motivation in-between the tasks.

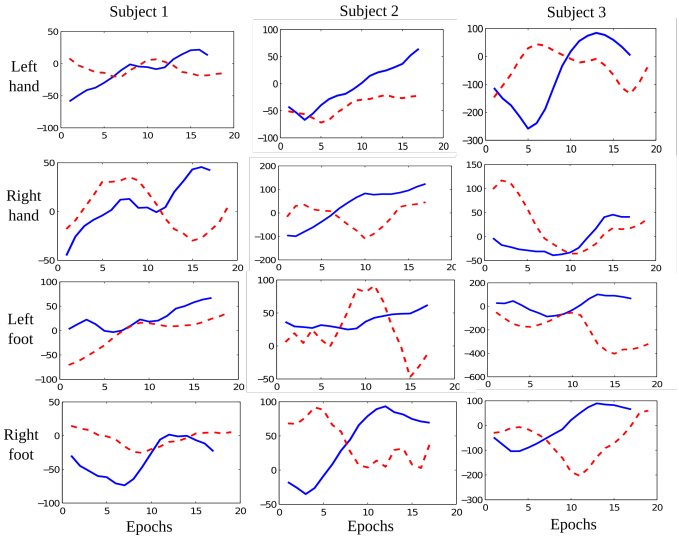


Figure 15. The learning curve of the 3 subjects for the motor imagery correctly classified tasks during FES sessions (in blue –) and VIS sessions (in red –) based on the average hyperplane distance.

7.2.5. NIRS-EEG joint imaging during transcranial direct current stimulation

Participants: Mehak Sood [IIT Hyderabad, India], Pierre Besson [Euromov, UM], Makii Muthalib [Euromov, UM], Utkarsh Jindal [IIT Hyderabad, India], Stéphane Perrey [Euromov, UM], Anirban Dutta, Mitsuhiro Hayashibe.

Transcranial direct current stimulation (tDCS) has been shown to perturb both cortical neural activity and hemodynamics during (online) and after the stimulation, however mechanisms of these tDCS-induced online and after-effects are not known. Here, online resting-state spontaneous brain activation may be relevant to monitor tDCS neuromodulatory effects that can be measured using electroencephalography (EEG) in conjunction with near-infrared spectroscopy (NIRS). We present a Kalman Filter based online parameter estimation of an autoregressive (ARX) model to track the transient coupling relation between the changes in EEG power spectrum and NIRS signals during anodal tDCS (2 mA, 10 min) using a 4x1 ring high-definition montage. Our online ARX parameter estimation technique using the cross-correlation between EEG band-power (0.5-11.25 Hz) and NIRS oxy-hemoglobin signal in the low frequency range was shown in 5 healthy subjects to be sensitive to detect transient EEG-NIRS coupling changes in resting-state spontaneous brain activation during anodal tDCS. Conventional sliding window cross-correlation calculations suffer a fundamental problem in computing the phase relationship as the signal in the window is considered time-invariant and the choice of the window length and step size are subjective. Here, Kalman Filter based method allowed online ARX parameter estimation using time-varying signals that could capture transients

in the coupling relationship between EEG and NIRS signals. Our new online ARX model based tracking method allows continuous assessment of the transient coupling between the electrophysiological (EEG) and the hemodynamic (NIRS) signals representing resting-state spontaneous brain activation during anodal tDCS. It is supported by Franco-Indian Inria-DST project funding and by the LabEx NUMEV (ANR-10-LABX-20).

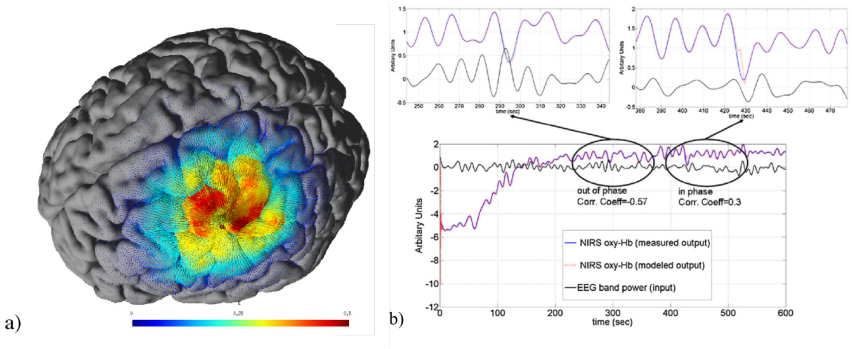


Figure 16. a) Electric field (V/m) estimated at the gray matter surface due to 2 mA anodal HD-tDCS. b) An illustrative example showing the NIRS oxy-Hb signal that was measured (in blue) as well as the predicted NIRS oxy-Hb signal using the ARX online tracking method (in dotted red).

7.2.6. Is EMG a good signal to assess fatigure under FES in different stimulation modes?.

Participants: Willy Fagart, Robin Candau [EUROMOV], Anthony Gelys [Propara Center], Mitsuhiro Hayashibe, David Guiraud.

The study that we have undertaken aims to analyse the neuromuscular fatigue in 3 paralysed subjects with spinal cord injury and to find if there is a link between the torque and the EMG signal. 6 series of 8 trains of stimulation (30 Hz, 400μs, 3 on / 2 off, in maximal intensity) were used to lead a muscular fatigue on the soleus muscle. At the beginning and the end of every series of stimulation, we measured the intermediate state of fatigue by evoking muscular twitch (1Hz, 400 μs and of maximal intensity) on the 2 legs. The temporal component and frequency of electromyographic activity were analyzed. These values were correlated with the torque. At the end of the protocol of stimulation, the torque decreased on 5 legs on 6 (ranging from -39 % to -20 %, p<0,05). A polynomial of degres 2 relation was found between the torque and the peak to peak value of the EMG signal. Nevertheless this relation does not remain reliable in a clinical context with regard to the variability of the data. This variability represents the processes of potentiation of the electric and mechanical answer as well as the expression of the mechanisms which contribute to the muscular fatigue.

7.2.7. EleVANT project: a diagnostic evaluation of acute stroke by near infrared spectroscopy and transcranial direct current stimulation coupling.

Participants: Victor Vagné, Vincent Costalat, David Guiraud, Emmanuelle Le Bars, Stéphane Perrey [EUROMOV], Olivier Rossel, Mitsuhiro Hayashibe.

Cerebral infarctions can now be treated with new techniques using intravenous thrombolysis and mechanical clearance. Their proven efficacy is directly correlated to the time lapse between the start of symptoms and initiation of treatment. Currently, a definitive diagnosis can only be made once the patient has realized a radiological imaging (CT scan or MRI) on a medical centre equipped with these expensive devices, thus enabling the medical team to initiate the appropriate treatment. Transit times during the pre-hospitalisation phase before diagnosis are therefore often longer and have the greatest negative impact on the patient's prognosis. The association of NIRS and tDCS enables recording modifications in cortical tissue oxygenation induced

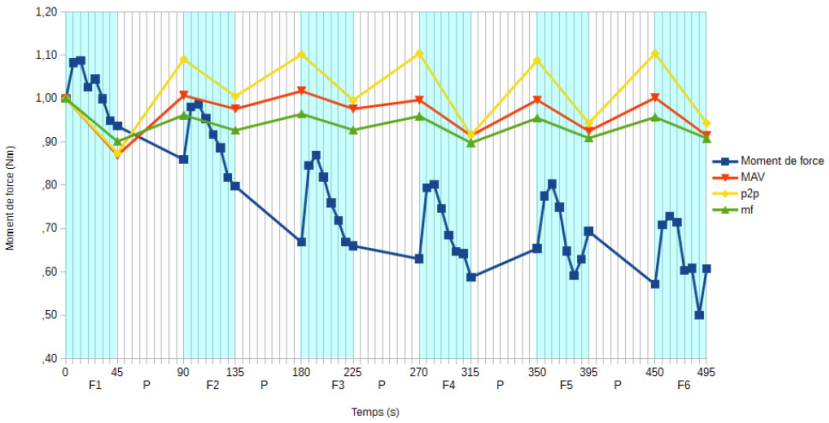


Figure 17. Correlation of different EMG features and torque during (Fn phases) and between (P phases) train of stimulations. Fatigue can be oberved and followed by EMG parameters.

by electrostimulation. A case-control study demonstrated the capacity of near infra-red spectroscopy (NIRS), combined with transcranial direct current stimulation (tDCS) to diagnose established cerebral ischaemia. According to this study, the affected hemisphere with impaired circulation showed significantly less change in cerebral hemoglobin oxygenation than the healthy side in response to anodal tDCS. This preliminary study showed the feasibility of identifying the lesioned hemisphere in subacute stroke patient. In collaboration with Gui de Chauliac Hospital, I2FH and Euromov, the EleVANT project is aiming to prospectively evaluate the NIRS-tDCS technique in the 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. A first NIRS-tDCS helmet prototype was developed to gather our Oxymon NIRS and Starstim tDCS devices, allowing good optodes-scalp and electrodes-scalp contact, while reducing both movement artifact and set-up time. This helmet was improved steps by steps as tests were done to attempt several parameters (among others electrodes location, amplitude and time of stimulation). A 4 NIRS optodes and 2 electrodes montage was retained to test and validate the proof of principle. Preliminary results are encouraging and need further investigation to be strongly validated.

Otherwise, as effects of anodal tDCS on hemodynamic response remain discussed, we'll proceed in parallel with the establishment of MRI protocols to attempt a validation of these effects.

7.3. Synthesis and Control of Human Functions

7.3.1. FES-assisted cycling in SCI individuals

Participants: Christine Azevedo Coste, Benoît Sijobert, Charles Fattal, Anne Daubigney [COS Divio, Dijon], Jérôme Parent, Antonio Padilha Bo [University Brasilia], Emerson Facin Martins [University Brasilia], Lucas Fonseca [University Brasilia], Juliana Guimaraes [University Brasilia].

During more than one year we have prepared one complete paraplegic patient to participate into FES-cycling discipline at Cybathlon 2016. A research protocol was associated to this physical preparation and several variables have been monitored during the training in order to evaluate performances, physical and psychological state. We have also developped a FES tricycle dedicated to the competition. We have modified a commercial trike and a commercial electrostimulator in order to have a mid cost system, adapted to complete paraplegia, easy to transport and compatible with safe transfers between wheelchair and trike seat. Our pilot reached the objectives: participating into the race, being qualified and cycle 750m in less than 8mn. He has been able during his training to cycle 1km200 in 13mn (fig.19).

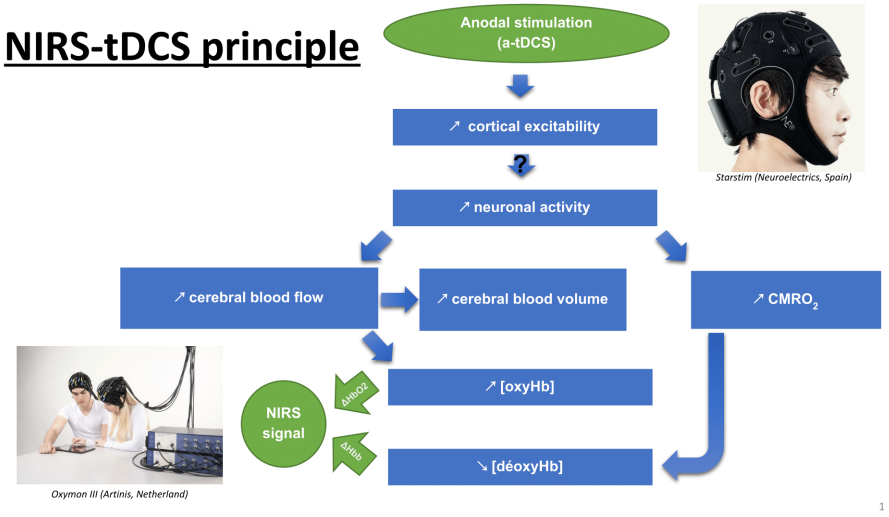


Figure 18.

In parallel, within CACAO associate team context, our Brazilian partner has trained several pilots using a similar training protocol [24], [22].

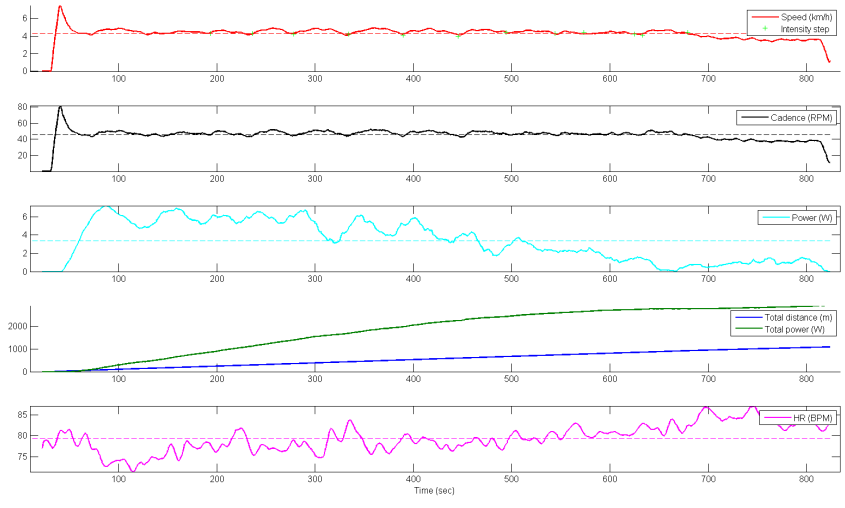


Figure 19. Best cycling performance. From top to bottom: speed (km/h), cadence (RPM), power (W), total distance (m) vs power (W), heart rate (BPM).

7.3.2. FES-assisted transfer in SCI individuals

Participants: Christine Azevedo Coste, Charles Fattal, Emerson Facin Martins [University Brasilia], Lucas Fonseca [University Brasilia], Ana Claudia Lopes [University Brasilia], Roberto Baptista [University Brasilia], Claudia Ochoa [University Brasilia].

One of the research axes investigated in CACAO associate team with Brasilia University is the assistance of seat to seat transfers in spinal cord injured (SCI) individuals. We have initated a research protocol to evaluate the feasibility to reduce arm efforts during pivot transfers by using feet support provided by lower limb muscles stimulation. 2 complete paraplegic patients were included for pilot experiments. Transfer is a key ability and allows greater interaction with the environment and social participation. Conversely, paraplegics have great risk of pain and injury in the upper limbs due to joint overloads during activities of daily living, like transfer. Preliminary results were promissing [30]. Further inclusions will be achieved to confirm these preliminary observations.

7.3.3. New cueing modality for Parkinson Disease

Participants: Christine Azevedo Coste, Benoît Sijobert, Christian Geny [CHU Montpellier].

Parkinson's Disease (PD) is the second most common neurodegenerative disorder in the world. It is often related to gait impairments and to a high risk of falls. Among different consequences of this disease, the Freezing of Gait (FOG) is defined as an episodic inability to generate an effective stepping. Subjects report the feeling of having their feet "glued to the ground". Numerous studies used auditory or visual stimulus to prevent FOG to happen.

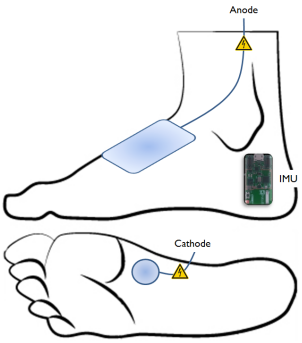


Figure 20. New cueing modality for Parkinson Disease. Electrodes and inertial sensor localization.

In our study, we aimed to investigate the effect of a sensitive cueing on gait disorders in subjects suffering from PD for improving gait and for reducing FOG occurrence. 13 participants with PD were equipped with an electrical stimulator and an inertial measurement unit (IMU) located under the lateral malleolus on the sagittal plane. Electrodes were positioned under the arch of the foot (Fig. 20) and electrical stimulation (ES) parameters adjusted to deliver a sensitive signal. Based on previous studies we achieved using IMU in Parkinson's Disease [52], [51], [56], in this protocol online IMU signal was processed in order to trigger ES at heel off detection (Fig. 21). Starting from a quiet standing posture, subjects were asked to walk at their preferred speed on a path including 5m straight line, u-turn and walk around tasks. 3 situations were considered: no stimulation baseline pre-condition, ES condition, no stimulation baseline post-condition. In ES condition the time to execute the different tasks was globally decreased in all the subjects. In "freezer" subjects, the time to complete the entire path was reduced by 19%. Freezing of Gait (FOG) episodes occurrence was decreased by 12% compared to baseline conditions. This preliminary work showed a positive global effect on gait and FOG in PD of a somatosensory cueing based on sensitive electrical stimulation [32].

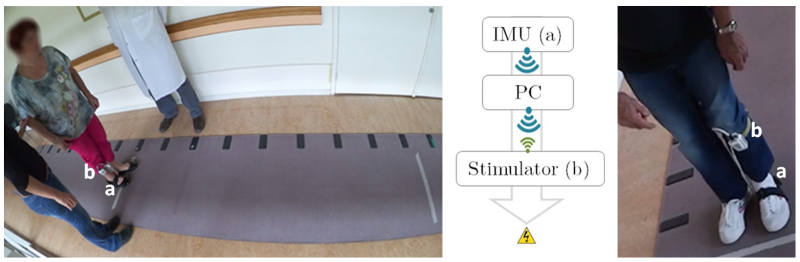


Figure 21. New cueing modality for Parkinson Disease. Experimental setup description.

7.3.4. Selective neural electrical stimulation of the upper limb nerves

Participants: Christine Azevedo Coste, David Guiraud, Wafa Tigra, Jacques Teissier [Clinique Beausoleil], Bertrand Coulet [CHU Montpellier], Charles Fattal, Anthony Gelis [Clinique PROPARA].

We have experimented a new approach of selective neural electrical stimulation of the upper limb nerves of two tetraplegic patients. Median and radial nerves are stimulated via a multipolar cuff electrode to elicit movements of wrist and hand in acute conditions during a surgical intervention. Various configurations corresponding to various combinations of a 12- poles cuff electrode contacts are tested. Video recording and electromyographic (EMG) signals recorded via sterile surface electrodes are used to evaluate the selectivity of each stimulation configuration in terms of activated muscles. We succeed to elicit graduated extension of wrist and fingers and graduated wrist flexion. We have also experimented a new human-machine interface to, at term trigger this electrical stimulation by individuals with tetraplegia. We investigated the feasibility of piloting an assistive device by processing supra-lesional muscle responses online. The ability to voluntarily contract a set of selected muscles was assessed in five spinal cord-injured subjects through electromyographic (EMG) analysis. Two subjects were also asked to use the EMG interface to control palmar and lateral grasping of a robot hand (Fig. 22). The use of different muscles and control modalities was also assessed. All patients are able to contract some of the evaluated muscles, preferential mode of pilot is patient dependent (Fig. 23).

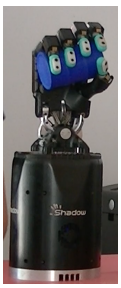


Figure 22. Closure posture of the Shadow robotic hand in the palmar grasp situation.

7.3.5. Spinal cord stimulation investigation

Participants: Christine Azevedo Coste, David Guiraud, Thomas Guiho, Charles Fattal, Luc Bauchet [CHU Montpellier].

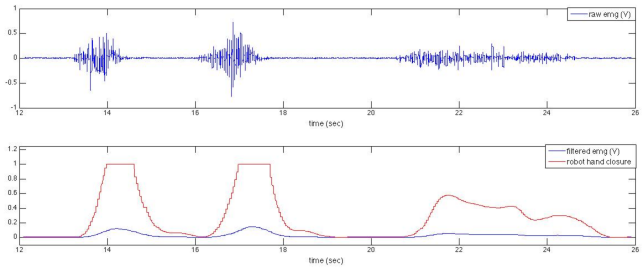


Figure 23. Robot hand trajectories generated from EMG recording for proportional mode. Top: raw EMG, bottom filtered EMG (blue) and hand trajectory (red). 0: hand is open, 1: hand is closed.

Spinal cord injury results in the loss of movement and sensory sensations but also in the disruption of some organ functions. Nearly all spinal cord injured subjects lose bladder control and are prone to kidney failure if they do not apply intermittent (self-) catheterization. Electrical stimulation of the sacral spinal roots with an implantable neuroprosthesis is one option besides self-catheterization to become continent and control micturition. However, many persons do not ask for this neuroprosthetic device (Brindley-Finetech implant) since deafferentation and loss of sensory functions and reflexes are serious side effects and since alternative treatments are available to patients (drugs, botulinus toxin. . . .). This PhD work aimed at investigating various techniques for spinal cord electrical stimulation in order to address dysfunctions in spinal cord injured individuals on lesion levels that have an impact on lower limb movements and bladder, bowel and sexual functions. Orderly recruitment of fibers at the spinal cord level should eventually lead to orderly recruitment of the detrusor muscle without activation of the bladder sphincter. Thereby, low pressure voiding, for example, should be obtained but is currently impossible with existing active implantable medical devices. A new large animal model – the domestic pig – was investigated to overcome size effects of rodent models and be able to translate results and technology more easily to human. [23].

7.4. Neuroprostheses and technology

7.4.1. Fast simulation and optimization tool to explore selective neural stimulation

Participants: Mélissa Dali, Olivier Rossel, David Guiraud.

In functional electrical stimulation, selective stimulation of axons is desirable to activate a specific target, in particular muscular function. This implies to simulate a fascicule without activating neighboring ones i.e. to be spatially selective. Spatial selectivity is achieved by the use of multicontact cuff electrodes over which the stimulation current is distributed. Because of the large number of parameters involved, numerical simulations provide a way to find and optimize electrode configuration. The present work offers a computation effective scheme and associated tool chain capable of simulating electrode-nerve interface and finds the best spread of current to achieve spatial selectivity. The software is protected to « Agence de Protection des Programmes « (APP), with the name MOS2SENS and identification IDDN.FR.001.490036.000.S.P.2014.000.31230 [21]

7.4.2. Numerical simulation of multipolar configuration

Participants: Mélissa Dali, Olivier Rossel, David Guiraud.

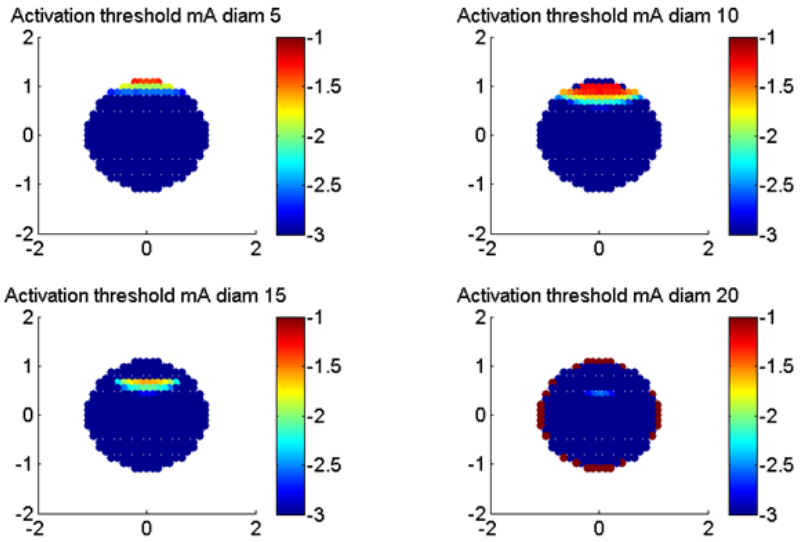


Figure 24. Spatially reverse recruitment order for fiber from 5μm to 20μm diameter obtained by multipolar configuration and prepulse technique

In the context of functional electrical stimulation of peripheral nerves, the control of a specific motor or sensory functions may need selective stimulation to target the desired effect without others. In implanted stimulation, spatial selectivity is obtained using multipolar CUFF electrodes with specific spread of the current over each contact. Furthermore, electrical stimulation recruits large fibers before small ones, whereas the targeted function could be elicited by a specific fiber type i.e. fiber diameter. In our work, numerical simulations were used to investigate the combination of multipolar configuration and prepulses, in order to obtain spatially reverse recruitment order. Multipolar stimulation provides efficient spatial selectivity, whereas subthreshold prepulses were used to reverse recruitment order with a reasonable increase of the injected charges. We compared several selective configurations combined with prepulses to show that some are able to guarantee both the spatial selectivity while one fiber's diameter can be preferentially activated [42].

7.4.3. Formal validation for critical 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. 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., when it is executed.

One solution has been studied in the PhD thesis of H. Leroux [54], which lays the foundations of translation rules from the designed model to the analyzed model integrating both implementation and execution constraints. 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.

A well-known drawback of such approach is that model checking is a technique that achieves properties verification through an exhaustive analysis of the state graph of the system model. The main limitation of this technique is the state space explosion problem because of its intrinsic exhaustivity. In a first part of the thesis (2015-2016), we proposed a compact state graph, called the Reduced Graph (see figure 25), which preserves all sequences of transitions firing as well as minimal and maximal duration of each sequence. To do so, we extend the partial order semantics to define temporal parallelism relations. According to covering steps approach, we compute our reduced graph reducing transitions interleaving, while keeping potential parallelism information.

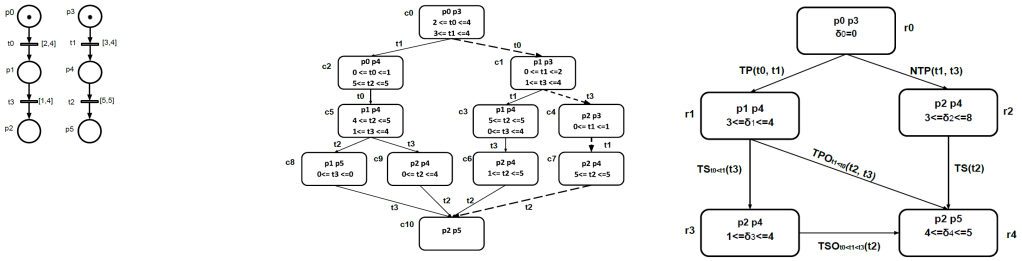


Figure 25. a Time Petri net, its classical State Class Graph and our Reduced Graph

But using classical analysis tools forces to analyze an over-set of the real behaviors, which limits the analytical capacities. In particular, the classical semantics of Petri nets considers an asynchronous execution, while in our context they are synchronously executed on FPGA with real parallelism and clock synchronization. Thus, we propose a new states graph which takes into account all the implementation and execution constraints related to the target hardware (non-functional properties): the Synchronous Behaviour Graph (SBG). We formally defined the graph and its semantics, illustrating this method on a simple example (see Figure 26). Then, we apply our method on a real industrial model, which is the execution engine embedded in our neurostimulator.

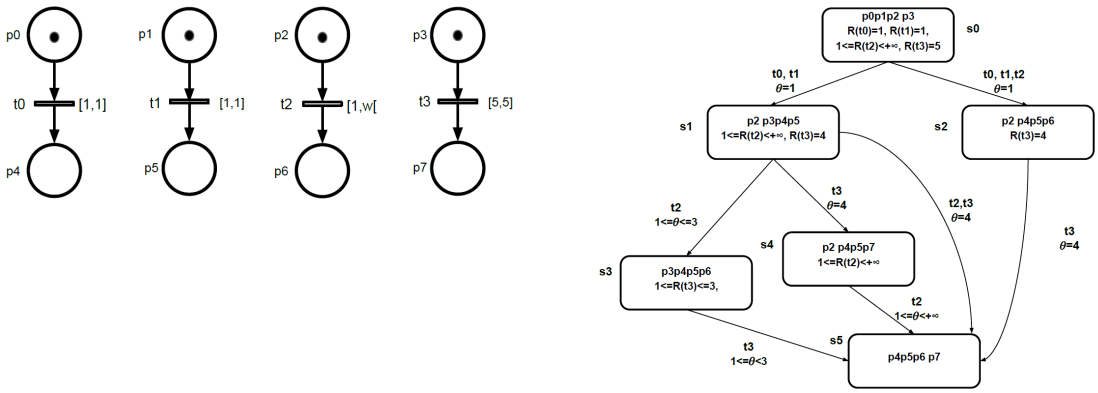


Figure 26. a Time Petri net and its Synchronous Behaviour Graph.

7.4.4. Control and scheduling co-design for stimulation systems

Participants: Daniel Simon, Zineddine Djellouli, David Andreu.

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 [14]. 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 [6].

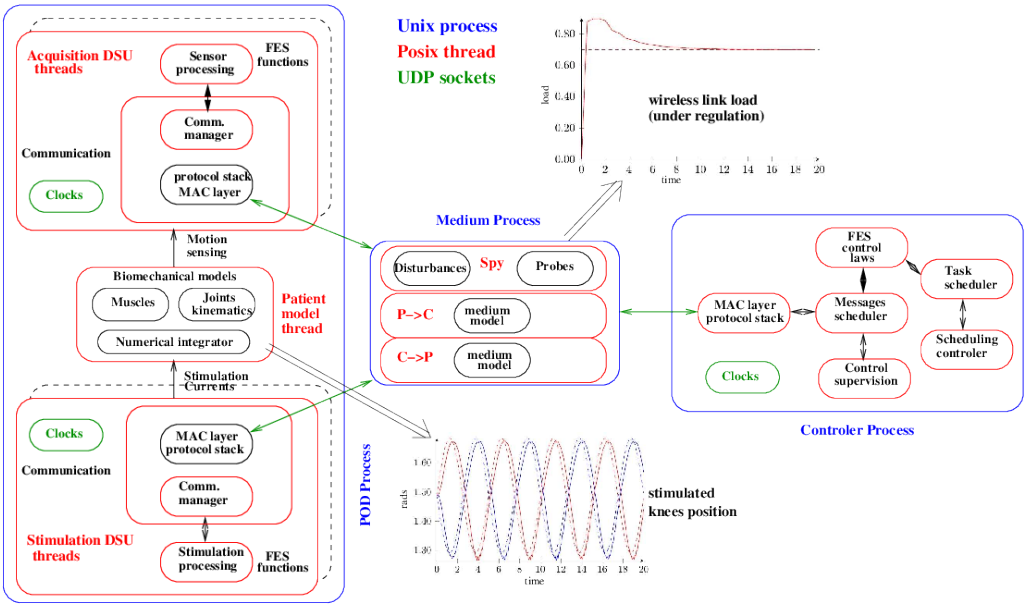


Figure 27. Real-time hybrid simulation architecture

Such simulator can be used for the design, testing and preliminary validation of new technologies and implementation. For example, it has been used to evaluate extensions of the STIMAP wireless communication protocol to optimize the network bandwidth when using multiple stimulation sites and control loops. Thanks to the hybrid nature of the simulation tool, the effect of the enhanced protocol can be directly observed on the controller output (e.g., concurrent controllers running to control several joints).

Another use is for the evaluation of closed-loop controllers acting on the execution resources of the distributed system. This approach provides adaptability and robustness, allowing for the design of fault-tolerant systems against varying and uncertain operating conditions [48]. It is especially usefull for embedded systems where these resources are scarce and fragile, as for the limited bandwith of wireless links between controllers and stimulation probes. Hence, a simple PI controller has been applied to the (m,k)-firm scheduling policy of control messages sent over the wireless link between the control device and the stimulation probes. It has been observed that such simple scheduling controller is able to jointly regulate both the communication load and the joints control quality.

GALEN Project-Team

7. New Results

7.1. Learning Grammars for Architecture-Specific Facade Parsing

Paticipants: Nikos Paragios (in collaboration with researchers from Université Paris-Est, LIGM, ENPC)

In [5], we present a novel framework to learn a compact grammar from a set of ground-truth images. To this end, parse trees of ground-truth annotated images are obtained running existing inference algorithms with a simple, very general grammar. From these parse trees, repeated subtrees are sought and merged together to share derivations and produce a grammar with fewer rules. Furthermore, unsupervised clustering is performed on these rules, so that, rules corresponding to the same complex pattern are grouped together leading to a rich compact grammar.

7.2. Non-Rigid Surface Registration

Paticipants: Dimitris Samaras, Nikos Paragios

This work [13] casts surface registration as the problem of finding a set of discrete correspondences through the minimization of an energy function, which is composed of geometric and appearance matching costs, as well as higher-order deformation priors. Two higher-order graph-based formulations are proposed under different deformation assumptions.

7.3. Monocular Surface Reconstruction using 3D Deformable Part Models

Paticipants: Maxim Berman, Stefan Kinauer, Iasonas Kokkinos

In this work [22] we train and detect part-based object models in 2D images, recovering 3D position and shape information (per part positions), allowing for a 3D reconstruction of the object. The resulting optimization problem is solved via a Branch&Bound approach, yielding detection results within a fraction of a second.

7.4. Learning with Non-modular loss functions

Paticipants: Jiaqian Yu, Matthew Blaschko

We have proposed an alternating direction method of multipliers (ADMM) based decomposition method loss augmented inference, that only depends on two individual solvers for the loss function term and for the inference term as two independent subproblems. In this way, we can gain computational efficiency and achieve more flexibility in choosing our non-modular loss functions of interest. We have proposed a novel supermodular loss function that empirically achieved better performance on the boundary of the objects, finding elongated structure [33]. We also introduced a novel convex surrogate operator for general non-modular loss functions, which provides for the first time a tractable solution for loss functions that are neither supermodular nor submodular, e.g. Dice loss. This surrogate is based on a canonical submodular-supermodular decomposition for which we have demonstrated its existence and uniqueness. It is further proven that this surrogate is convex, piecewise linear, an extension of the loss function, and for which subgradient computation is polynomial time [32][31].

7.5. Asymptotic Variance of MMD and Relative MMD

Paticipants: Eugene Belilovsky, Wacha Bounliphone, Matthew Blaschko (in collaboration with researchers at UCL and Deepmind)

Kernel mean embeddings allow for comparisons of complex distributions. They have been recently heavily used in hypothesis testing to compare distributions as well as in the nascent field of deep generative modeling. In this work we derived the asymptotic variance of the MMD and the cross covariance between joint MMD. We showed how this can be used effectively for model selection in complex Deep Generative Models where the likelihood metric is not accessible. Our results on the asymptotic variance of the MMD have already been used by other researchers to propose an efficient method for optimal testing and improved training of generative models.

7.6. Deconvolution and Deinterlacing of Video Sequences

Paticipants: 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 this work, 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.

7.7. A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise

Paticipants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with Y. Marnissi, PhD student at Univ. Paris-Est Marne la Vallée and Y. Zheng, IBM Research China)

In this work, 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. The Majorize-Minimize Subspace Algorithm and Block Parallelization

Paticipants: Emilie Chouzenoux and Jean-Christophe Pesquet (in collaboration with S. Cadoni, Master student at Univ. Paris-Est Marne la Vallée and Dr C. Chaux, Univ. Aix-Marseille)

State-of-the-art methods for solving smooth optimization problems are nonlinear conjugate gradient, low memory BFGS, and Majorize-Minimize (MM) subspace algorithms. The MM subspace algorithm which has been introduced more recently has shown good practical performance when compared with other methods on various optimization problems arising in signal and image processing. However, to the best of our knowledge, no general result exists concerning the theoretical convergence rate of the MM subspace algorithm. The paper [3] aims at deriving such convergence rates both for batch and online versions of the algorithm and, in particular, discusses the influence of the choice of the subspace. We also propose a Block Parallel Majorize-Minimize Memory Gradient (BP3MG) algorithm for solving large scale optimization problems in [16]. This

algorithm combines a block coordinate strategy with an efficient parallel update. The proposed method is applied to a 3D microscopy image restora- tion problem involving a depth-variant blur, where it is shown to lead to significant computational time savings with respect to a sequential approach.

7.9. Stochastic Forward-Backward and Primal-Dual Approximation Algorithms with Application to Online Image Restoration

Paticipants: Jean-Christophe Pesquet (In collaboration with Pr. P. L. Combettes, North Carolina State university)

Stochastic approximation techniques have been used in various contexts in data science. We propose a stochastic version of the forward-backward algorithm for minimizing the sum of two convex functions, one of which is not necessarily smooth. Our framework can handle stochastic approximations of the gradient of the smooth function and allows for stochastic errors in the evaluation of the proximity operator of the nonsmooth function. The almost sure convergence of the iterates generated by the algorithm to a minimizer is established under relatively mild assumptions. We also propose a stochastic version of a popular primal-dual proximal splitting algorithm, establish its convergence, and apply it to an online image restoration problem.

7.10. Random primal-dual proximal iterations for sparse multiclass SVM

Paticipants: Jean-Christophe Pesquet (in collaboration with Pr. G. Chierchia, Univ. Paris-Est Marne la Vallée, and Dr. N. Pustelnik, ENS Lyon)

Sparsity-inducing penalties are useful tools in variational methods for machine learning. In this paper, we propose two block-coordinate descent strategies for learning a sparse multiclass support vector machine. The first one works by selecting a subset of features to be updated at each iteration, while the second one performs the selection among the training samples. These algorithms can be efficiently implemented thanks to the flexibility offered by recent randomized primal-dual proximal methods. Experiments carried out for the supervised classification of handwritten digits demonstrate the interest of considering the primal-dual approach in the context of block-coordinate descent. The efficiency of the proposed algorithms is assessed through a comparison of execution times and classification errors.

7.11. PALMA, an improved algorithm for DOSY signal processing

Paticipants: Emilie Chouzenoux (in collaboration with Prof. 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. Used in conjunction with a Maximum Entropy and λ_1 hybrid regularisation [39], 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 in [35]. It has been implemented on the server [http://palma.labo.igbmc.fr] where users can have their datasets processed automatically.

7.12. Graph-based change detection and classification in satellite image pairs

Paticipants: Maria Vakalopoulou, Nikos Paragios

We proposed a scalable, modular, metric-free, single-shot change detection/registration method for remote sensing image pairs [11]. The framework exploits a decomposed interconnected graphical model formulation where in the presence of changes the iconic similarity constraints are relaxed. We employ a discretized, grid-based deformation space. State-of-the-art linear programming and duality principles have been used to optimize the joint solution space where local consistency is imposed on the deformation and the detection space. The proposed framework is working both in a unsupervised and supervised manner depending on the application. The developed method has been validated through large scale experiments on several multi-temporal very high resolution optical satellite datasets. Also a novel generic framework has been designed, developed and validated for addressing simultaneously the tasks of image registration, segmentation and change detection from multisensor, multiresolution, multitemporal satellite image pairs [30]. Our approach models the inter-dependencies of variables through a higher order graph. A patch-based deep learning strategy has been employed and used for segmentation likelihoods. The evaluation of the developed framework was performed on the '2016 IEEE GRSS Data Fusion Contest' dataset and indicate very promising results for all three different tasks.

7.13. Graphical models in artificial vision

Paticipants: Nikos Komodakis, M. Pawan Kumar, Stavros Alchatzidis, Enzo Ferrante, Evangelia Zacharaki, Nikos Paragios

Computer vision tasks are often reformulated as mathematical inference problems where the objective is to determine the set of parameters corresponding to the lowest potential of a task-specific objective function. Graphical models have been the most popular formulation in the field over the past two decades. In [7] we focus on the inference component of the problem and in particular we discuss in a systematic manner the most commonly used optimization principles in the context of graphical models. In [8] we briefly review hyper-graph representations as prominent tools in the casting of perception as a graph optimization problem. We discuss their strength and limitations, provide appropriate strategies for their inference and present their application to address a variety of problems in biomedical image analysis.

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 [2], 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.

7.14. Pattern analysis of EEG signals with epileptic activity

Paticipants: Evangelia Zacharaki (in collaboration with Prof. M. Megalooikonomou, University of Patras and M. Koutroumanidis, King's College, London)

We have addressed the needs of epileptic patients and healthcare professionals, aiming at the design and development of a non-intrusive personal health system for the monitoring and analysis of epilepsy-relevant multi-parametric data and the documentation of the epilepsy related symptoms. Specifically, we investigated the classification of epileptic and non-epileptic events from EEG based on temporal and spectral analysis and different fusion schemes [9]. We also studied the EEG brain activity during whole night sleep, since sleep is recognized as a major precipitator of epileptic activity [12].

MATHNEURO Team

5. New Results

5.1. Neural Networks as dynamical systems

5.1.1. A modular architecture for transparent computation in recurrent neural networks

Participants: Giovanni Carmantini [Plymouth University, UK], Peter Beim Graben [Humbolt University (Berlin), Germany], Mathieu Desroches [Inria MathNeuro], Serafim Rodrigues [Plymouth University, UK].

Computation is classically studied in terms of automata, formal languages and algorithms; yet, the relation between neural dynamics and symbolic representations and operations is still unclear in traditional eliminative connectionism. Therefore, we suggest a unique perspective on this central issue, to which we would like to refer as transparent connectionism, by proposing accounts of how symbolic computation can be implemented in neural substrates. In this study we first introduce a new model of dynamics on a symbolic space, the versatile shift, showing that it supports the real-time simulation of a range of automata. We then show that the Gödelization of versatile shifts defines nonlinear dynamical automata, dynamical systems evolving on a vectorial space. Finally, we present a mapping between nonlinear dynamical automata and recurrent artificial neural networks. The mapping defines an architecture characterized by its granular modularity, where data, symbolic operations and their control are not only distinguishable in activation space, but also spatially localizable in the network itself, while maintaining a distributed encoding of symbolic representations. The resulting networks simulate automata in real-time and are programmed directly, in the absence of network training. To discuss the unique characteristics of the architecture and their consequences, we present two examples: (i) the design of a Central Pattern Generator from a finite-state locomotive controller, and (ii) the creation of a network simulating a system of interactive automata that supports the parsing of garden-path sentences as investigated in psycholinguistics experiments.

This work has been published in Neural Networks and is available as [13].

5.1.2. Latching dynamics in neural networks with synaptic depression

Participants: Pascal Chossat [Inria MathNeuro], Martin Krupa [Inria MathNeuro], Frédéric Lavigne [Université de Nice - BCL].

Priming is the ability of the brain to more 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. 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 Hebbian learning, 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 submitted for publication and is available as [23].

5.1.3. On the Hamiltonian structure of large deviations in stochastic hybrid systems

Participants: Paul Bressloff [University of Utah, USA], Olivier Faugeras [Inria MathNeuro].

We present a new derivation of the classical action underlying a large deviation principle (LDP) for a stochastic hybrid system, which couples a piecewise deterministic dynamical system in $\mathbb{R}^d$ with a time-homogeneous Markov chain on some discrete space Γ. We assume that the Markov chain on Γ is ergodic, and that the discrete dynamics is much faster than the piecewise deterministic dynamics (separation of timescales). Using the Perron-Frobenius theorem and the calculus-of-variations, we show that the resulting Hamiltonian is given by the Perron eigenvalue of a $|\Gamma|$-dimensional linear equation. The corresponding linear operator depends on the transition rates of the Markov chain and the nonlinear functions of the piecewise deterministic system. We compare the Hamiltonian to one derived using WKB methods, and show that the latter is a reduction of the former. We also indicate how the analysis can be extended to a multi-scale stochastic process, in which the continuous dynamics is described by a piecewise stochastic differential equations (SDE). Finally, we illustrate the theory by considering applications to conductance-based models of membrane voltage fluctuations in the presence of stochastic ion channels.

This work has been submitted for publication and is available as [22].

5.1.4. Large Deviations of a Spatially-Stationary Network of Interacting Neurons

Participants: Olivier Faugeras [Inria MathNeuro], James Maclaurin [University of Sydney, USA].

In this work we determine a process-level Large Deviation Principle (LDP) for a model of interacting neurons indexed by a lattice $\mathbb{Z}^d$. The neurons are subject to noise, which is modelled as a correlated martingale. The probability law governing the noise is strictly stationary, and we are therefore able to find a LDP for the probability laws Π^n governing the stationary empirical measure $\widehat{\mu}^n$ generated by the neurons in a cube of length $(2n+1)$. We use this LDP to determine an LDP for the neural network model. The connection weights between the neurons evolve according to a learning rule / neuronal plasticity, and these results are adaptable to a large variety of neural network models. This LDP is of great use in the mathematical modelling of neural networks, because it allows a quantification of the likelihood of the system deviating from its limit, and also a determination of which direction the system is likely to deviate. The work is also of interest because there are nontrivial correlations between the neurons even in the asymptotic limit, thereby presenting itself as a generalisation of traditional mean-field models.

This work has been submitted for publication and is available as [25].

5.1.5. The Period adding and incrementing bifurcations: from rotation theory to applications

Participants: Albert Granados [Technical University of Denmark, Denmark], Lluís Alsedà [Autonomous University of Barcelona, Spain], Martin Krupa [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 accepted for publication in SIAM Review and is available as [26].

5.2. Neural Fields Theory

5.2.1. 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 [Inria MathNeuro], 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, center 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 submitted for publication and is available as [27].

5.3. Slow-Fast Dynamics in Neuroscience

5.3.1. Canards, folded nodes and mixed-mode oscillations in piecewise-linear slow-fast systems

Participants: Mathieu Desroches [Inria MathNeuro], Antoni Guillamon [Polytechnic University of Catalunya, Spain], Enrique Ponce [University of Seville, Spain], Rafel Prohens [University of the Balearic Islands, Spain], Antonio E. Teruel [University of the Balearic Islands, Spain], Serafim Rodrigues [Plymouth University, UK].

Canard-induced phenomena have been extensively studied in the last three decades, from both the mathematical and the application viewpoints. Canards in slow-fast systems with (at least) two slow variables, especially near folded-node singularities, give an essential generating mechanism for mixed-mode oscillations (MMOs) in the framework of smooth multiple timescale systems. There is a wealth of literature on such slow-fast dynamical systems and many models displaying canard-induced MMOs, particularly in neuroscience. In parallel, since the late 1990s several papers have shown that the canard phenomenon can be faithfully reproduced with piecewise-linear (PWL) systems in two dimensions, although very few results are available in the three-dimensional case. The present paper aims to bridge this gap by analyzing canonical PWL systems that display folded singularities, primary and secondary canards, with a similar control of the maximal winding number as in the smooth case. We also show that the singular phase portraits are compatible in both frameworks. Finally, we show using an example how to construct a (linear) global return and obtain robust PWL MMOs.

This work has been published in SIAM Review and is available as [16].

5.3.2. Spike-adding in parabolic bursters: the role of folded-saddle canards

Participants: Mathieu Desroches [Inria MathNeuro], Martin Krupa [Inria MathNeuro], Serafim Rodrigues [Plymouth University, UK].

The present work develops a new approach to studying parabolic bursting, and also proposes a novel four-dimensional canonical and polynomial-based parabolic burster. In addition to this new polynomial system, we also consider the conductance-based model of the Aplysia R15 neuron known as the Plant model, and a reduction of this prototypical biophysical parabolic burster to three variables, including one phase variable, namely the Baer-Rinzel-Carillo (BRC) phase model. Revisiting these models from the perspective of slow-fast dynamics reveals that the number of spikes per burst may vary upon parameter changes, however the spike-adding process occurs in an explosive fashion that involves special solutions called canards. This spike-adding canard explosion phenomenon is analysed by using tools from geometric singular perturbation theory in tandem with numerical bifurcation techniques. We find that the bifurcation structure persists across all considered systems, that is, spikes within the burst are incremented via the crossing of an excitability threshold given by a particular type of canard orbit, namely the true canard of a folded-saddle singularity. However there can be a difference in the spike-adding transitions in parameter space from one case to another, according to whether the process is continuous or discontinuous, which depends upon the geometry of the folded-saddle canard. Using these findings, we construct a new polynomial approximation of the Plant model, which retains all the key elements for parabolic bursting, including the spike-adding transitions mediated by folded-saddle canards. Finally, we briefly investigate the presence of spike-adding via canards in planar phase models of parabolic bursting, namely the theta model by Ermentrout and Kopell.

This work has been published in Physica D and is available as [17].

5.3.3. Slow-fast transitions to seizure states in the Wendling-Chauvel neural mass model

Participants: Mathieu Desroches [Inria MathNeuro], Olivier Faugeras [Inria MathNeuro], Martin Krupa [Inria MathNeuro].

We revisit the Wendling-Chauvel neural mass model by reducing it to eight ODEs and adding a differential equation that accounts for a dynamic evolution of the slow inhibitory synaptic gain. This allows to generate dynamic transitions in the resulting nine-dimensional model. The output of the extended model can be related to EEG patterns observed during epileptic seizure, in particular isolated pre-ictal spikes and low-voltage fast oscillations at seizure onset. We analyse the extended model using basic tools from slow-fast dynamical systems theory and relate the main transitions towards seizure states to torus canards, a type of solutions that has been shown to explain the spiking to bursting transition in many neural models. We find that the original ten-dimensional Wendling-Chauvel model can be reduced to eight dimensions, two variables being scaled versions of two other variables of the model. We then obtain a model with four PSP blocks, which is consistent with the block-diagrams typically presented to describe this model. Instead of varying the slow inhibitory synaptic gain parameter B quasi-statically, or just performing numerical bifurcation analysis in B as the structure of the fast subsystem of an hypothetical extended system, we construct a true slow dynamics for B, depending sensitively on the main PSP output of the model, Y0. Near fold bifurcation of limit cycles of the original model, the solution to the extended model performs fast low-amplitude oscillations close to both attracting and repelling branches of limit cycles, which is the signature of a torus canard phenomenon.

This work has been published in Opera Medica & Physiologica and is available as [14].

5.3.4. Canards in a minimal piecewise-linear square-wave burster

Participants: Mathieu Desroches [Inria MathNeuro], Soledad Fernández-García [University of Seville, Spain], Martin Krupa [Inria MathNeuro].

We construct a piecewise-linear (PWL) approximation of the Hindmarsh-Rose (HR) neuron model that is minimal, in the sense that the vector field has the least number of linearity zones, in order to reproduce all the dynamics present in the original HR model with classical parameter values. This includes square-wave bursting and also special trajectories called canards, which possess long repelling segments and organise the transitions between stable bursting patterns with n and n?+?1 spikes, also referred to as spike-adding canard explosions. We propose a first approximation of the smooth HR model, using a continuous PWL system, and show that its fast subsystem cannot possess a homoclinic bifurcation, which is necessary to obtain proper square-wave bursting. We then relax the assumption of continuity of the vector field across all zones, and we show that we can obtain a homoclinic bifurcation in the fast subsystem. We use the recently developed canard theory for PWL systems in order to reproduce the spike-adding canard explosion feature of the HR model as studied, e.g., in Desroches et al., Chaos 23(4), 046106 (2013).

This work has been published in Chaos and is available as [15].

5.3.5. From Canards of Folded Singularities to Torus Canards in a Forced van der Pol Equation

Participants: John Burke [Boston University, USA], Mathieu Desroches [Inria MathNeuro], Albert Granados [Technical University of Denmark, Denmark], Tasso J. Kaper [Boston University, USA], Martin Krupa [Inria MathNeuro], Theodore Vo [Boston University, USA].

In this article, we study canard solutions of the forced van der Pol equation in the relaxation limit for low-, intermediate-, and high-frequency periodic forcing. A central numerical observation made herein is that there are two branches of canards in parameter space which extend across all positive forcing frequencies. In the low-frequency forcing regime, we demonstrate the existence of primary maximal canards induced by folded saddle nodes of type I and establish explicit formulas for the parameter values at which the primary maximal canards and their folds exist. Then, we turn to the intermediate- and high-frequency forcing regimes and show that the forced van der Pol possesses torus canards instead. These torus canards consist of long

segments near families of attracting and repelling limit cycles of the fast system, in alternation. We also derive explicit formulas for the parameter values at which the maximal torus canards and their folds exist. Primary maximal canards and maximal torus canards correspond geometrically to the situation in which the persistent manifolds near the family of attracting limit cycles coincide to all orders with the persistent manifolds that lie near the family of repelling limit cycles. The formulas derived for the folds of maximal canards in all three frequency regimes turn out to be representations of a single formula in the appropriate parameter regimes, and this unification confirms the central numerical observation that the folds of the maximal canards created in the low-frequency regime continue directly into the folds of the maximal torus canards that exist in the intermediate- and high-frequency regimes. In addition, we study the secondary canards induced by the folded singularities in the low-frequency regime and find that the fold curves of the secondary canards turn around in the intermediate-frequency regime, instead of continuing into the high-frequency regime. Also, we identify the mechanism responsible for this turning. Finally, we show that the forced van der Pol equation is a normal form-type equation for a class of single-frequency periodically driven slow/fast systems with two fast variables and one slow variable which possess a non-degenerate fold of limit cycles. The analytic techniques used herein rely on geometric desingularisation, invariant manifold theory, Melnikov theory, and normal form methods. The numerical methods used herein were developed in Desroches et al. (SIAM J Appl Dyn Syst 7:1131–1162, 2008, Nonlinearity 23:739–765 2010).

This work has been published in Journal of Nonlinear Science and is available as [12].

5.3.6. Mixed-mode oscillations in a piecewise-linear system with multiple time scale coupling

Participants: Soledad Fernández-García [University of Seville, Spain], Martin Krupa [Inria MathNeuro], Frédérique Clément [Inria Mycenae].

In this work, we analyze a four dimensional slow?fast piecewise linear system with three time scales presenting Mixed-Mode Oscillations. The system possesses an attractive limit cycle along which oscillations of three different amplitudes and frequencies can appear, namely, small oscillations, pulses (medium amplitude) and one surge (largest amplitude). In addition to proving the existence and attractiveness of the limit cycle, we focus our attention on the canard phenomena underlying the changes in the number of small oscillations and pulses. We analyze locally the existence of secondary canards leading to the addition or subtraction of one small oscillation and describe how this change is globally compensated for or not with the addition or subtraction of one pulse.

This work has been published in Physica D and is available as [18].

5.4. Plasticity

5.4.1. Time-code neurotransmitter release at excitatory and inhibitory synapses

Participants: Serafim Rodrigues [Plymouth University, UK], Mathieu Desroches [Inria MathNeuro], Martin Krupa [Inria MathNeuro], Jesus M. Cortes [Biocruces Institute, Spain], Terrence J. Sejnowski [Salk Institute, USA], Afia B. Ali [University College London, UK].

Communication between neurons at chemical synapses is regulated by hundreds of different proteins that control the release of neurotransmitter that is packaged in vesicles, transported to an active zone, and released when an input spike occurs. Neurotransmitter can also be released asynchronously, that is, after a delay following the spike, or spontaneously in the absence of a stimulus. The mechanisms underlying asynchronous and spontaneous neurotransmitter release remain elusive. Here, we describe a model of the exocytotic cycle of vesicles at excitatory and inhibitory synapses that accounts for all modes of vesicle release as well as short-term synaptic plasticity (STSP). For asynchronous release, the model predicts a delayed inertial protein unbinding associated with the SNARE complex assembly immediately after vesicle priming. Experiments are proposed to test the model?s molecular predictions for differential exocytosis. The simplicity of the model will also facilitate large-scale simulations of neural circuits.

This work has been published in Proceedings of the National Academy of Sciences of the USA (PNAS) and is available as [20].

5.5. Vision in Neuroscience

5.5.1. The relative contribution of noise and adaptation to competition during tri-stable motion perception

Participants: Andrew Meso [Bournemouth University, UK], James Rankin [Center for Neural Science, NYU, USA], Olivier Faugeras [Inria MathNeuro], Pierre Kornprobst [Inria BioVision], Guillaume Masson [Institut de Neuroscience de la Timone, France].

Animals exploit antagonistic interactions for sensory processing and these can cause oscillations between competing states. Ambiguous sensory inputs yield such perceptual multi-stability. Despite numerous empirical studies using binocular rivalry or plaid pattern motion, the driving mechanisms behind the spontaneous transitions between alternatives remain unclear. In the current work, we used a tri-stable barberpole motion stimulus combining empirical and modelling approaches to elucidate the contributions of noise and adaptation to underlying competition. We first robustly characterised the coupling between perceptual reports of transitions and continuously recorded eye direction, identifying a critical window of 480ms before button presses within which both measures were most strongly correlated. Second, we identified a novel non monotonic relationship between stimulus contrast and average perceptual switching rate with an initially rising rate before a gentle reduction at higher contrasts. A neural fields model of the underlying dynamics introduced in previous theoretical work and incorporating noise and adaptation mechanisms was adapted, extended and empirically validated. Noise and adaptation contributions were confirmed to dominate at the lower, and higher, contrasts respectively. Model simulations with two free parameters, controlling adaptation dynamics and direction thresholds, captured the measured mean transition rates for participants. We verified the shift from noise dominated towards adaptation-driven in both the eye direction distributions and inter-transition duration statistics. This work combines modelling and empirical evidence to demonstrate the signal strength dependent interplay between noise and adaptation during tri- stability. We propose that the findings generalise beyond the barberpole stimulus case to ambiguous perception in continuous feature spaces.

This work has been published in Journal of Vision and is available as [19].

MIMESIS Team

5. New Results

5.1. Augmented Reality for Hepatic Surgery

Participants: Rosalie Plantefève, Bruno Marques, Frederick Roy, Nazim Haouchine, Igor Peterlik, Stéphane Cotin.

Liver cancer is the 2nd most common cause of cancer death worldwide, with more than 745,000 deaths from liver cancer in 2012. When including deaths from liver cirrhosis, the toll reaches nearly 2 million people worldwide. Today, surgical tumor ablation remains the best treatment for liver cancer.To localize the hepatic tumors and to define the resection planes, clinicians rely on pre-operative medical images (obtained with computed tomography scanner or magnetic resonance imaging). However, the liver lesions and vascular system are difficult to localize during surgery. This may lead to incomplete tumor resection or haemorrhage.

We provide surgeons with an augmented view of the liver and its internal structures during surgery to help them to optimally resect the tumors while limiting the risk of vascular lesion. Therefore, an elastic registration method to align the pre-operative and intra-operative data has been developed [26]. This method, which uses a biomechanical model and anatomical landmarks, was designed to limit its impact on the clinical workflow and reaches a registration accuracy below the resection margin even when the liver is strongly deformed between its pre-operative and intra-operative state. This registration algorithm has been integrated into a software, SOFA-OR, to conduct the first clinical tests.

5.2. Augmented Reality for Mini-Invasive Surgery

Participants: Nazim Haouchine, Lionel Untereiner, Frederick Roy, Igor Peterlik, Stéphane Cotin.

We have addressed the ill-posed problem of initial alignment of pre-operative to intra-operative data for augmented reality during minimally invasive hepatic surgery. This problem consists in finding the rigid transformation that relates the scanning reference and the endoscopic camera pose, and the non-rigid transformation undergone by the liver with respect to its scanned state. Most of the state-of-the-art methods assume a known initial registration.

We have proposed in [16] a method that permits to recover the deformation undergone by the liver while simultaneously finding the rotational and translational parts of the transformation. Our formulation considers the boundaries of the liver with its surrounding tissues as hard constraints directly encoded in an energy minimization process. We performed experiments on real in-vivo data of human hepatic surgery and synthetic data, and compared our method with related works (Figure 7).

5.3. Detecting topological changes during non-rigid registration

Participants: Christoph Paulus, David Cazier, Stéphane Cotin.

Augmented reality has shown significant promise in overcoming certain visualization and interaction challenges in various domains such as medicine, construction, advertising, manufacturing, and gaming. Despite the promise of augmented reality and its successful application to many domains, significant research challenges remain. Among these challenges is the augmentation of non-rigid structures that can undergo topological changes, such as fracture, tearing or cutting. This is for instance the case in minimally invasive surgery, which has gained popularity and became a well-established procedure thanks to its benefits for the patient, in particular with shortened recovery times.

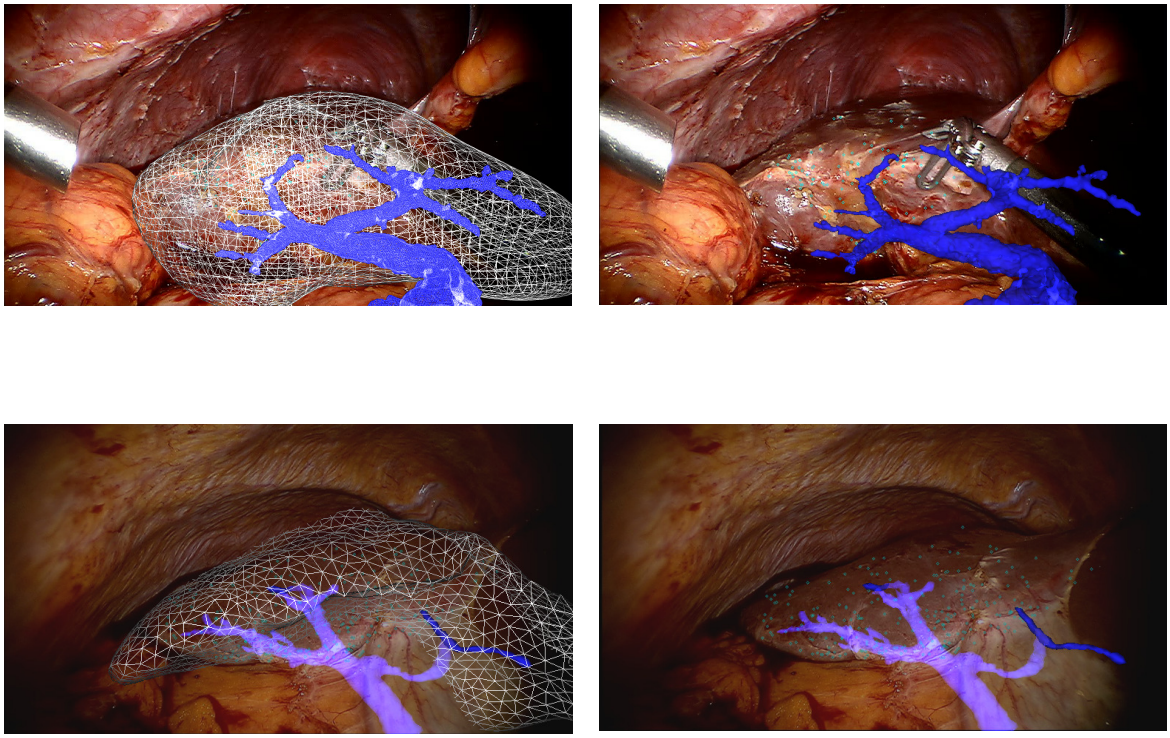


Figure 6. Non-rigid registration between intra-operative and pre-operative data. The overlay of the liver and its vascular network help the surgeon during the operation.

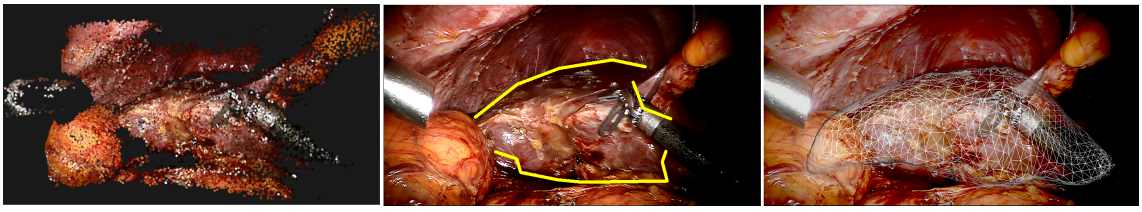


Figure 7. Left: a 3D map is reconstructed from the intra-operative view. Middle: The contours of the liver are extracted. Right: They are used as constraint to pilot a biomechanical model.

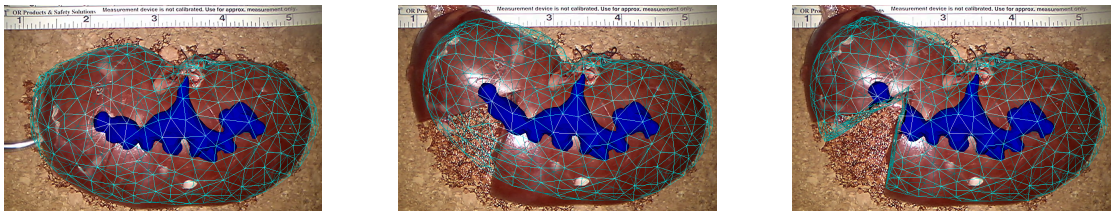


Figure 8. Left: A kidney whose internal structures have been scanned and segmented is cut and deformed. Middle: Standart methods do not detect the cut. Right: Our method detects the cut and applies it to the virtual model.

Current methods dealing with non-rigid augmented reality only provide an augmented view when the topology of the tracked object is not modified, which is an important limitation. We solve this shortcoming by introducing a method for physics-based non-rigid augmented reality [11]. Singularities caused by topological changes are detected by analyzing the displacement field of the underlying deformable model. These topological changes are then applied to the physics-based model to approximate the real cut. All these steps, from deformation to cutting simulation, are performed in real-time. This significantly improves the coherence between the actual view and the model, and provides added value.

5.4. Augmented Reality for Vascular Surgery

Participants: Raffaella Trivisonne, Igor Peterlik, Hadrien Courtecuisse, Stéphane Cotin.

Significant changes have taken place over the past 20 years in medicine with the development of minimally invasive procedures. While surgery evolved towards laparoscopy for instance, interventional radiology has become another alternative for many pathologies. Regarding catheter-based interventions, the lack of depth perception in projective grey-scale images, and the extensive use of X-ray imaging to visualize the instrument and the anatomy through which it must be inserted, are among the main issues. We propose to address these different problems by developing an advanced navigation system which relies on a combination of real-time simulation and information extracted from intra-operative images to assess the current position of the catheter. Such a method would have direct applications in endovascular procedures allowing for an enhanced view of the operating field, both in term of 3D perception and quality of the images. Our approach combines advanced modeling of the device, 2D-3D registration and constraint-based simulation.

We have developed a method [18] based on constraint-based simulation allowing for the enhancement of fluoroscopic images with a 3D real-time catheter insertion and 3D vessel visualization. Our method relies mainly on image features, without the need of any information about the surrounding 3D vasculature, nor does it require any tracking device (Figure 9).

5.5. Image analysis for the characterization of cell mobility

Participant: Igor Peterlik.

The complex behaviour of motile cells plays a crucial role in biological processes such as tissue growth and tumorigenesis. Biomedical research that focuses on understanding the mechanisms of cell motility generates large amounts of multidimensional image data acquired by fully automated optical microscopes. Manual analysis of such data is extremely laborious and therefore, it is necessary to develop reliable automatic methods of image analysis. However, evaluation and assessment of such methods remains a challenging task, since in the case of real data, no ground truth is available to establish simple and robust metrics. Therefore, an important task in development of automatic methods of image analysis is the synthetic generation of realistic images allowing for quantitative assessment based on the ground truth.

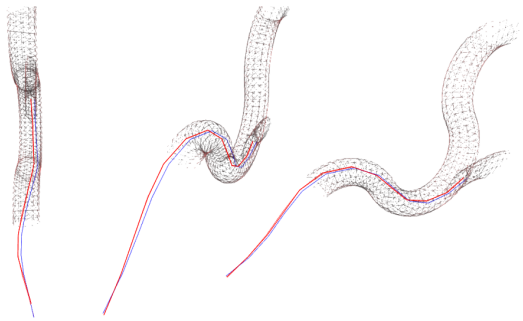


Figure 9. A 3D catheter reconstruction (red) and the real catheter (blue).

We collaborate with the Centre of Biomedical Image Analysis (CBIA) at Masaryk University, Czech Republic on the development of reliable image analysis methods for quantitative characterization of cell motility driven by cellular protrusions at the leading edge of crawling cells. In particular, we develop physics-based models of living cells which are used to generate synthetic time-lapse 3D image series that realistically mimic the motile cells with protrusions (Figure 10). Although modeling of living cells has many specificities, we successfully exploit the modeling algorithms originally designed and developed for the simulation soft tissues. We have already demonstrated that realistic simulation of living cells can be achieved using SOFA and are working toward more complex models and scenarios, involving interactions among the cells and mitosis.

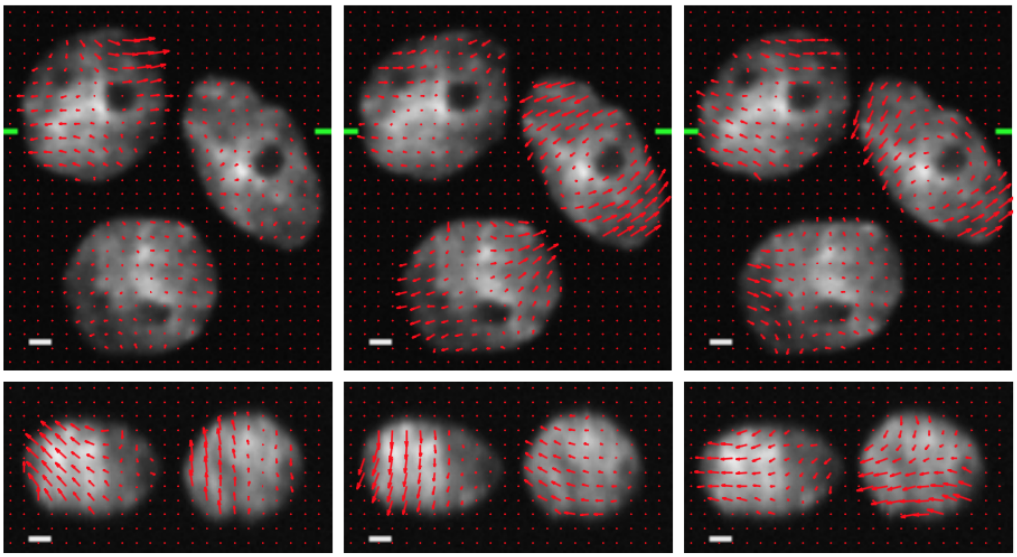


Figure 10. Sample synthetic nuclei generated with our method.

5.6. Training for retina surgery

Participants: Rémi Bessard Duparc, Stéphane Cotin.

Retina surgery is an increasingly performed procedure for the treatment of a wide spectrum of retinal pathologies. Yet, as most micro-surgical techniques, it requires long training periods before being mastered. To properly answer requests from clinicians for highly realistic training on one hand, and new requirements from accreditation or recertification from surgical societies on the other hand, we are developing a high-fidelity training system for retinal surgery.

This simulator is built upon our strong scientific expertise in the field of real-time simulation and a success story for technology transfer in the field of cataract surgery simulation. The simulation system is based on the Open Source simulation platform SOFA, and relies on expertise from our partners to ensure clinical and industrial relevance (this work is funded through the ANR project RESET). A first version of the training system has been developed and demonstrated in different ophthalmology conferences.

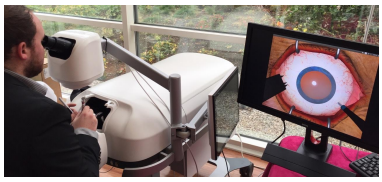
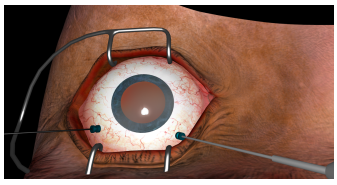
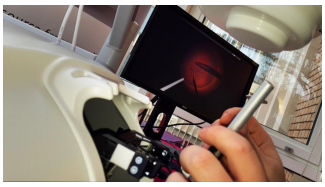




Figure 11. Left: the simulation is performed on a dedicated hardware including a microscope and instruments. Middle: The instruments are inserted in the eye. Right: The epiretinal membrane is removed.

5.7. Robotic control of flexible needle insertion

Participants: Yinoussa Adagolodjo, Hadrien Courtecuisse.

We introduce a new method for automatic robotic needle steering in deformable tissues [13]. It uses an inverse Finite Element (FE) simulation to control an articulated robot interacting with deformable structures. We consider a flexible needle, embedded in the end effector of a 6 arm Mitsubishi RV1A robot, and its insertion into a silicone phantom. Given a trajectory on the rest configuration of the silicone phantom, our method provides in real-time the displacements of the articulated robot which guarantee the permanence of the needle within the predefined path, taking into account any undergoing deformation on both the needle and the trajectory itself. A forward simulation combines i) a kinematic model of the robot, ii) FE models of the needle and phantom gel iii) an interaction model allowing the simulation of friction and puncture force. A Newton-type method is then used to provide the displacement of the robot to minimize the distance between the needle's tip and the desired trajectory. We validate our approach with a simulation in which a virtual robot can successfully perform the insertion while both the needle and the trajectory undergo significant deformations.

5.8. Compensation of brain shift in brain tumor surgery

Participant: Hadrien Courtecuisse.

During brain tumor surgery, planning and guidance are based on pre-operative images which do not account for brain-shift. However, this shift is a major source of error in neuro-navigation systems and affects the accuracy of the procedure. The vascular tree is extracted from pre-operative Magnetic Resonance Angiography and from intra-operative Doppler ultrasound images, which provides sparse information on brain deformations. The pre-operative images are then updated based on an elastic registration of the blood vessels, driven by a patient-specific biomechanical model. We develop a biomechanical model [17] to extrapolate the deformation to the surrounding soft tissues. Our method has proved to efficiently compensate for brain deformation while being compatible with a surgical process.

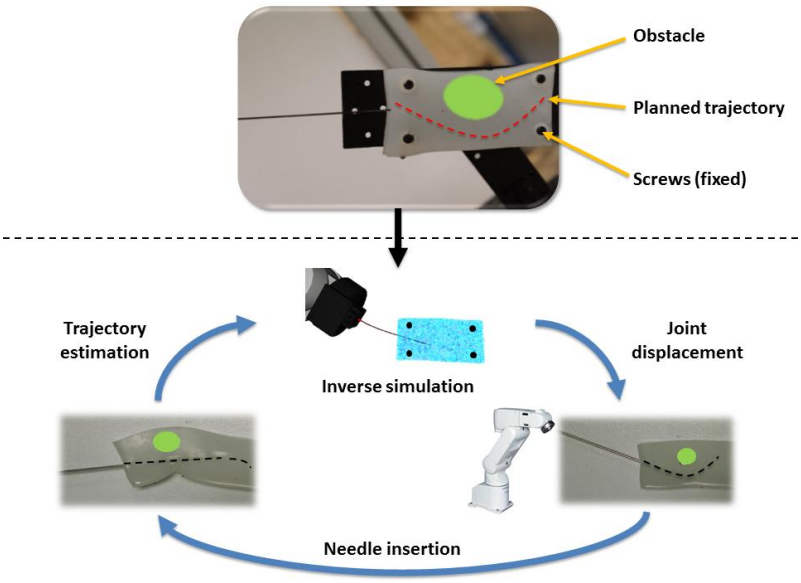


Figure 12. Automatic robotic needle steering in deformable tissues.

5.9. Regional anaesthesia

Participants: Rémi Bessard Duparc, Stéphane Cotin.

The **RASimAs** project (Regional Anaesthesia Simulator and Assistant) is a European research project funded by the European Union's 7th Framework Program. It aims at providing a virtual reality simulator and assistant to doctors performing regional anaesthesia by developing the patient-specific Virtual Physiological Human models. Our work lead to the following journal article (submitted in Sept 2016) : *Real-time error controlled adaptive mesh refinement: Application to needle insertion simulation*.

This paper presents the first real-time discretization-error-driven adaptive finite element approach for corotational elasticity problems involving strain localization. We propose a hexahedron-based finite element method, combined with a posteriori error estimation driven local h-refinement, for simulating soft tissue deformation. This enables to control the local error and global error level in the mechanical fields (e.g. displacement or gradient) during the simulation. The local error level is used to refine the mesh only where it is needed, while maintaining a coarser mesh elsewhere. We investigate the convergence of the algorithm on academic examples, and demonstrate its practical usability on a percutaneous procedure involving needle insertion in soft tissues.

2016 was the third year of the project during which we developed new models of the biomechanics of the leg and arm, as well as the simulation of the insertion of the anaesthesiology needle.

See the RASimAs web site for more details.

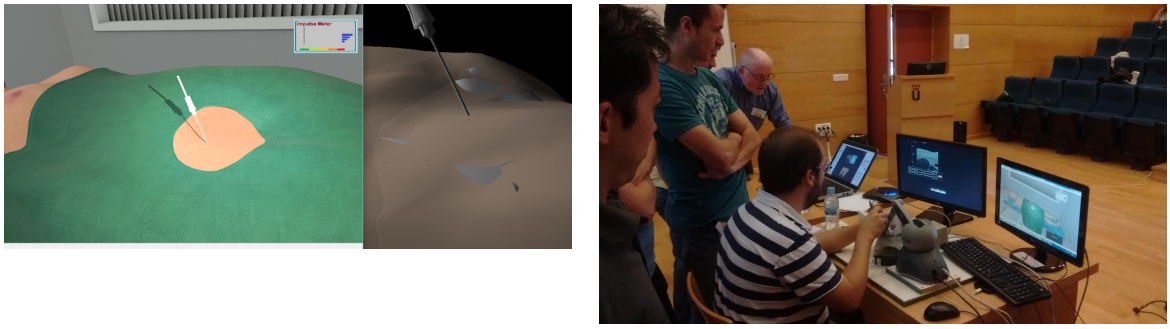


Figure 13. Left: Interface of the RASimAs simulator during femoral nerve block. Right: The RASimas developer team at the General Assembly.

MNEMOSYNE Project-Team

7. New Results

7.1. Overview

This year we first explored two main 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 also began this year to study some characteristics of 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 explored the limbic loop by describing a series of neural mechanisms that propose how responding conditioning results from interactions between the amygdala, the nucleus accumbens and the limbic pole of the frontal cortex. In our models [1], this learning is also fed by exchanges with the hippocampus (episodic memory) and the sensory cortex (semantic memory) and we studied the major role of acetylcholine in these exchanges. This also allowed us to address the difficult question of the articulation between the respondant and operant conditioning in particular in the nucleus accumbens. We proposed an original mechanism whereby noradrenaline could modulate the balance between exploration and exploitation [12] based on an assessment of the level of uncertainty and its impact on performance [13].

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 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.

Lastly, we carried out a thorough study about the behavior of our model of associative memory in the hippocampus [23], and particularly about its resistance to interference.

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. We have explored this year the dynamics of information flows within this circuit through a computational model described at the neuron and synapse level. Taking into account previous experimental observations from primates and earlier computational models, we incrementally developed a network capable of learning to perform behavioral tasks under several protocols and conditions [5]. The development of this computational model was conducted in parallel with the development of an experimental model of decision making in the salamander (Pleurodeles waltlii) [2]. The result here is a computational model of learning and decision making in the basal ganglia that allows for the testing of experimental hypotheses and also to conduct in silico pathophysiological or pharmacological investigations at the cellular level.

7.4. Systemic integration

We have worked this year on the integration of goal-oriented and habitual behaviors, two modes of learning associated to the motor loop. There is an apparent contradiction between experimental data showing that the basal ganglia are involved in goal-oriented and routine behaviors and clinical observations. Lesion or disruption by deep brain stimulation of the globus pallidus interna has been used for various therapeutic purposes ranging from the improvement of dystonia to the treatment of Tourette's syndrome. None of these approaches has reported any severe impairment in goal-oriented or automatic movement. To solve this conundrum, we trained two monkeys to perform a variant of a two-armed bandit-task (with different reward contingencies). Bilateral inactivation of the globus pallidus interna, by injection of muscimol, prevents animals from learning new contingencies while performance remains intact, although slower for the familiar stimuli. We replicate in silico these data by adding lateral competition and Hebbian learning in the cortical layer of the theoretical model of the cortex–basal ganglia loop that provided the framework of our experimental approach [7]. These results suggest that a behavioral decision results from both the cooperation (acquisition) and competition (expression) of two distinct but entangled memory systems, the goal-directed system and the habitual system that may represent the two ends of the same graded phenomenon.

We began our first works of systemic integration associating our models developed in the limbic and motor loops, for the study of the taking into account of the uncertainty in the selection of the action [1]. This preliminary work using the VirtualEnaction platform (*cf.* § 6.4) will be continued this year with a PhD that begins.

We have more generally proposed a study [11], analyzing the role of neuromodulation in adaptation to uncertainty, whose potential systemic impact is evident, particularly because it provides precious characteristics for autonomous learning [10].

7.5. Machine Learning

In Machine Learning, we were interested this year in two phenomena for which we consider classical paradigms of modeling and for which we wonder how they could be adapted by bio-inspiration.

The first paradigm concerns the manipulation of temporal sequences. In a perspective of better understanding how brain learn structured sequences we extended a model on syntax acquisition using the Reservoir Computing framework (using random recurrent networks) [16], [9], [19], [20]. The extended model is also used in a Human-Robot Interaction architecture to enable users to use more natural language with robots [14], [15], [18]. This work will be extended with our collaborators at the University of Hamburg (*cf.* § 9.3).

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.

The second paradigm concerns the extraction of characteristics and the use of hierarchical networks, as in the case of deep networks. An industrial application whose study has just begun (*cf.* § 9.2) will lead us to revisit these models to make them more easily usable in constrained frameworks, for example with limited size corpuses.

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Francesco Giovannini
In collaboration with Beate Knauer and Motoharu Yoshida (Ruhr University) and LieJune Shiau (University of Houston)

7.1.1. Memory and Anaesthesia

7.1.1.1. The CAN-In model of hippocampal theta oscillations

During working memory tasks, the hippocampus exhibits synchronous theta-band activity, which is thought to be correlated with the short-term memory maintenance of salient stimuli. Recent studies indicate that the hippocampus contains the necessary circuitry allowing it to generate and sustain theta oscillations without the need of extrinsic drive. However, the cellular and network mechanisms supporting synchronous rhythmic activity are far from being fully understood. Based on electrophysiological recordings from hippocampal pyramidal CA1 cells, we have presented a possible mechanism for the maintenance of such rhythmic theta-band activity in the isolated hippocampus [3]. Our model network, based on the Hodgkin-Huxley formalism, comprising pyramidal neurons equipped with calcium-activated non-specific cationic (CAN) ion channels, is able to generate and maintain synchronized theta oscillations ($4 - 12\,Hz$), following a transient stimulation. The synchronous network activity is maintained by an intrinsic CAN current (I_{CAN}), in the absence of constant external input. The analysis of the dynamics of model networks of pyramidal-CAN and interneurons (CAN-In) reveals that feedback inhibition improves the robustness of fast theta oscillations, by tightening the synchronisation of the pyramidal CAN neurons. The frequency and power of the theta oscillations are both modulated by the intensity of the I_{CAN}, which allows for a wide range of oscillation rates within the theta band.

This biologically plausible mechanism for the maintenance of synchronous theta oscillations in the hippocampus aims at extending the traditional models of septum-driven hippocampal rhythmic activity.

7.1.1.2. Generation of gamma oscillations in a network of adaptive exponential integrate and fire neurons

Fast neuronal oscillations in the Gamma rhythm (20-80 Hz) are observed in the neocortex and hippocampus during behavioral arousal. Through a conductance-based, four-dimensional Hodgkin-Huxley type neuronal model, Wang and Buzsáki have numerically demonstrated that such rhythmic activity can emerge from a random network of GABAergic interneurons when their intrinsic neuronal characters and network structure act as the main drive of the rhythm. We investigate Gamma oscillations through a randomly connected network model comprising low complexity, two-dimensional adaptive exponential integrate-and-fire (AdEx) neurons that have subthreshold and spike-triggered adaptation mechanisms. Despite the simplicity of our network model, it shares two important results with the previous biophysical model: the minimal number of necessary synaptic inputs to generate coherent Gamma-band rhythms remains the same, and this number is weakly-dependant on the network size. Using AdEx model, we also investigate the necessary neuronal, synaptic and connectivity properties that lead to random network synchrony with Gamma rhythms. These findings suggest a computationally more tractable framework for studying sparse and random networks inducing cortical rhythms in the Gamma band (Laure Buhry submitted an article to Journal of Computational Neuroscience, currently under major revision).

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Tamara Tošić, Mariia Fedotenkova, Meysam Hashemi, Cecilia Lindig-Leon, Jimmy, Nex, Sébastien Rimbert.

In collaboration with Stéphanie Fleck (Univ. Lorraine), Nathalie Gayraud (Inria Sophia Antipolis) and Maureen Clerc (Inria Sophia Antipolis)

7.2.1. Level of Consciousness

Participants: Axel Hutt, Meysam Hashemi

Meysam Hashemi defended his thesis about analytical and numerical studies of thalamo-cortical neural population models during general anesthesia. The findings of this thesis provide new insights into the mechanisms responsible for the specific changes in EEG patterns that are observed during propofol-induced sedation. Our results indicate that depending on the mean potential values of the system resting states, an increase or decrease in the thalamo-cortical gain functions results in an increase or decrease in the alpha power, respectively. In contrast, the evolution of the delta power is rather independent of the system resting states; the enhancement of spectral power in delta band results from the increased synaptic or extra-synaptic GABAergic inhibition. Furthermore, we aim to identify the parameters of a thalamo-cortical model by fitting the model power spectrum to the EEG recordings. To this end, we address the task of parameter estimation in the models that are described by a set of stochastic ordinary or delay differential equations [2].

7.2.2. Motor system

Participants: Laurent Bougrain, Cecilia Lindig-Leon, Jimmy, Nex, Sébastien Rimbert.

In collaboration with Stéphanie Fleck (Univ. Lorraine), Nathalie Gayraud (Inria Sophia Antipolis) and Maureen Clerc (Inria Sophia Antipolis)

7.2.2.1. Incremental motor imagery learning for rehabilitation after stroke

After a stroke, Brain-Computer Interfaces (BCI) allows improving rehabilitation of the motor cortex to recover the autonomy of the patient. The design of BCIs has to be done with an in-depth analysis concerning user's conditions during the learning of BCI. Since strokes affect mainly senior citizens, it is very important to guide the design of BCIs to make it usable. We propose to improve the experimental conditions through a new BCI protocol including an incremental motor imagery learning [21].

7.2.2.2. Motor neuroprostheses

We wrote a review that aims to position current neuroprosthetics research between reality and fiction, expectations of persons under a disability, fantasies of the augmented Man and scientific difficulties. Beyond the buzz effect to get the attention of the public and funders, and enthusiasm by journalists for novelty what are the expectations of potential users, the disappointments and the satisfactions of patients, how many persons are equipped, what are the price and the opportunities to use such devices outside of laboratories [5].

7.2.2.3. Classification of Motor patterns

In order to build systems that are able to detect several motor patterns, multiclass schemes need to be applied. We compared a series of multiclass approaches to assert the benefits of hierarchical classification. The compared methods are based on two effective techniques for MI-discrimination, namely, Common Spatial Patterns (CSP) and Riemannian geometry, for which the hierarchical and non-hierarchical approaches have been considered. We include the CSP by Joint Diagonalization method, which corresponds with a non-hierarchical approach; and its hierarchical counterpart, namely, Binary CSP. In addition, the non-hierarchical Minimum Distance to Riemannian Mean method (MDRM) is also evaluated, together with its analogous hierarchical approach; a contribution of the present work called Hierarchical MDRM algorithm (HMDRM). All these methods have been applied on dataset 2a of the BCI competition IV to facilitate their comparison. The highest accuracies were reached by the BCSP and HMDRM methods, confirming the effectiveness of hierarchical algorithms [7].

7.2.2.4. Discrete Motor Imageries for a Faster Detection

We are investigating differences between continuous MIs and discrete MIs corresponding to a 2s MI. Results show that both discrete and continuous MIs modulate ERD and ERS components. Both ERSs are different but ERDs are close in term of power of (de)synchronization. These results show that discrete motor imageries may be preferable for BCI systems design in order to faster detect MIs and reduce user fatigue. [8]

7.2.3. Pain under General Anaesthesia

Participants : Mariia Fedotenkova, Axel Hutt, Tamara Tošić
In collaboration with Peter beim Graben and James W. Sleigh.

7.2.3.1. Detection of EEG-signal Features for Pain under General Anaesthesia

Mariia Fedotenkova defended her thesis about extraction of multivariate components in brain signals obtained during general anesthesia. We studied analgesia effect of general anesthesia, more specifically, on patients reaction to nociceptive stimuli. We also study differences in the reaction between different anesthetic drugs. The study was conducted on a dataset consisting of 230 EEG signals: pre- and post-incision recordings obtained from 115 patients, who received desflurane and propofol. Combining features obtained with power spectral analysis and recurrence symbolic analysis [22], [6], [23], classification was carried out on a two-class problem, distinguishing between pre-/post-incision EEG signals, as well as between two different anesthetic drugs, desflurane and propofol [1].

PARIETAL Project-Team

6. New Results

6.1. Dictionary Learning for Massive Matrix Factorization

Sparse matrix factorization is a popular tool to obtain interpretable data decompositions, which are also effective to perform data completion or denoising. Its applicability to large datasets has been addressed with online and randomized methods, that reduce the complexity in one of the matrix dimension, but not in both of them. In this paper, we tackle very large matrices in both dimensions. We propose a new factoriza-tion method that scales gracefully to terabyte-scale datasets, that could not be processed by previous algorithms in a reasonable amount of time. We demonstrate the efficiency of our approach on massive functional Magnetic Resonance Imaging (fMRI) data, and on matrix completion problems for recommender systems, where we obtain significant speed-ups compared to state-of-the art coordinate descent methods.

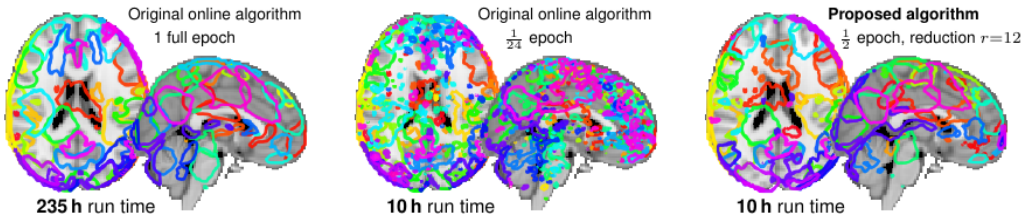


Figure 3. Brain atlases: outlines of each map obtained with dictionary learning. Left: the reference algorithm on the full dataset. Middle: the reference algorithm on a twentieth of the dataset. Right: the proposed algorithm with a similar run time: half the dataset and a compression factor of 9. Compared to a full run of the baseline algorithm, the figure explore two possible strategies to decrease computation time: processing less data (middle), or our approach (right). Our approach achieves a result closer to the gold standard in a given time budget. See [22] for more information.

See Fig. 3 for an illustration and [22] for more information.

6.2. Learning brain regions via large-scale online structured sparse dictionary-learning

We propose a multivariate online dictionary-learning method for obtaining de-compositions of brain images with structured and sparse components (aka atoms). Sparsity is to be understood in the usual sense: the dictionary atoms are constrained to contain mostly zeros. This is imposed via an 1-norm constraint. By "structured", we mean that the atoms are piece-wise smooth and compact, thus making up blobs, as opposed to scattered patterns of activation. We propose to use a Sobolev (Laplacian) penalty to impose this type of structure. Combining the two penalties, we obtain decompositions that properly delineate brain structures from functional images. This non-trivially extends the online dictionary-learning work of Mairal et al. (2010), at the price of only a factor of 2 or 3 on the overall running time. Just like the Mairal et al. (2010) reference method, the online nature of our proposed algorithm allows it to scale to arbitrarily sized datasets. Experiments on brain data show that our proposed method extracts structured and denoised dictionaries that are more intepretable and better capture inter-subject variability in small medium, and large-scale regimes alike, compared to state-of-the-art models.

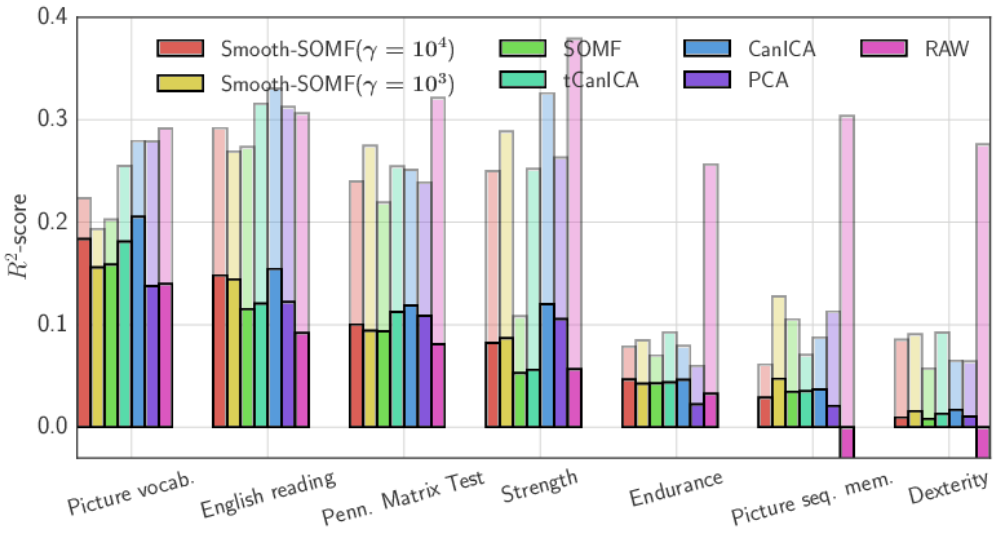


Figure 4. Predicting behavioral variables of the Human Connectome Project dataset using subject-level brain activity maps and various intermediate representations obtained with variants of dictionary learning. Bold bars represent performance on test set while faint bars in the background represent performance on train set. See [19] for more information.

See Fig. 4 for an illustration and [19] for more information.

6.3. Social-sparsity brain decoders: faster spatial sparsity

Spatially-sparse predictors are good models for brain decoding: they give accurate predictions and their weight maps are interpretable as they focus on a small number of regions. However, the state of the art, based on total variation or graph-net, is computationally costly. Here we introduce sparsity in the local neighborhood of each voxel with social-sparsity, a structured shrinkage operator. We find that, on brain imaging classification problems, social-sparsity performs almost as well as total-variation models and better than graph-net, for a fraction of the computational cost. It also very clearly outlines predictive regions. We give details of the model and the algorithm.

See Fig. 5 for an illustration and [32] for more information.

6.4. Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example

Resting-state functional Magnetic Resonance Imaging (R-fMRI) holds the promise to reveal functional biomarkers of neuropsychiatric disorders. However, extracting such biomarkers is challenging for complex multi-faceted neuropatholo-gies, such as autism spectrum disorders. Large multi-site datasets increase sample sizes to compensate for this complexity, at the cost of uncontrolled heterogeneity. This heterogeneity raises new challenges, akin to those face in realistic diagnostic applications. Here, we demonstrate the feasibility of inter-site classification of neuropsychiatric status, with an application to the Autism Brain Imaging Data Exchange (ABIDE) database, a large (N=871) multi-site autism dataset. For this purpose, we investigate pipelines that extract the most predictive biomarkers from the data. These R-fMRI pipelines build participant-specific connectomes from functionally-defined brain areas. Connectomes are then compared across participants to

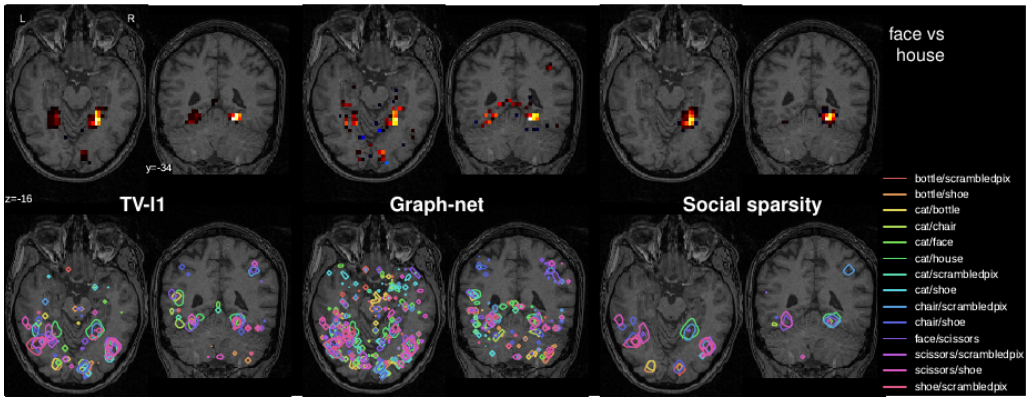


Figure 5. Decoder maps for the object-classification task – Top: weight maps for the face-versus-house task. Overall, the maps segment the right and left parahippocampal place area (PPA), a well-known place-specific regions, although the left PPA is weak in TV-l1, spotty in graph-net, and absent in social sparsity. Bottom: outlines at 0.01 of the other tasks. Beyond the PPA, several known functional regions stand out such as primary or secondary visual areas around the prestriate cortex as well as regions in the lateral occipital cortex, responding to structured objects. Note that the graphnet outlines display scattered small regions even thought the value of the contours is chosen at 0.01, well above numerical noise. See [32] for more information.

learn patterns of connectivity that differentiate typical controls from individuals with autism. We predict this neuropsychiatric status for participants from the same acquisition sites or different, unseen, ones. Good choices of methods for the various steps of the pipeline lead to 67% prediction accuracy on the full ABIDE data, which is significantly better than previously reported results. We perform extensive validation on multiple subsets of the data defined by different inclusion criteria. These enables detailed analysis of the factors contributing to successful connectome-based prediction. First, prediction accuracy improves as we include more subjects, up to the maximum amount of subjects available. Second, the definition of functional brain areas is of paramount importance for biomarker discovery: brain areas extracted from large R-fMRI datasets outperform reference atlases in the classification tasks.

See Fig. 6 for an illustration and [1] for more information.

6.5. Seeing it all: Convolutional network layers map the function of the human visual system

Convolutional networks used for computer vision represent candidate models for the computations performed in mammalian visual systems. We use them as a detailed model of human brain activity during the viewing of natural images by constructing predictive models based on their different layers and BOLD fMRI activations. Analyzing the predictive performance across layers yields characteristic fingerprints for each visual brain region: early visual areas are better described by lower level convolutional net layers and later visual areas by higher level net layers, exhibiting a progression across ventral and dorsal streams. Our predictive model generalizes beyond brain responses to natural images. We illustrate this on two experiments, namely retinotopy and face-place oppositions, by synthesizing brain activity and performing classical brain mapping upon it. The synthesis recovers the activations observed in the corresponding fMRI studies, showing that this deep encoding model captures representations of brain function that are universal across experimental paradigms.

See Fig. 7 for an illustration and [10] for more information.

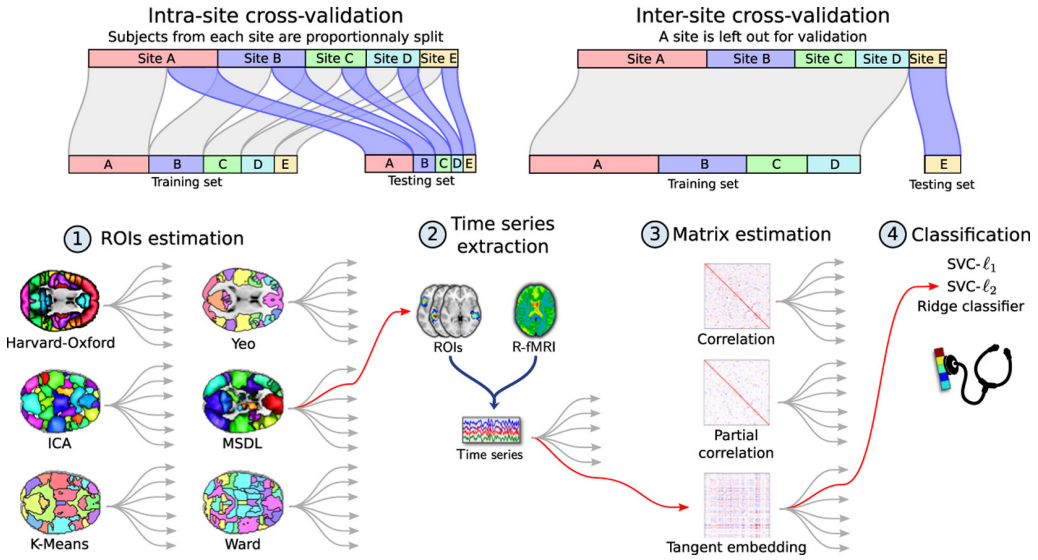


Figure 6. Validation of an fMRI-based pipeline for autism prediction. Several variants are considered for each pipeline step. See [1] for more information.

6.6. Formal Models of the Network Co-occurrence Underlying Mental Operations

Systems neuroscience has identified a set of canonical large-scale networks in humans. These have predominantly been characterized by resting-state analyses of the task-uncon-strained, mind-wandering brain. Their explicit relationship to defined task performance is largely unknown and remains challenging. The present work contributes a multivariate statistical learning approach that can extract the major brain networks and quantify their configuration during various psychological tasks. The method is validated in two extensive datasets (n = 500 and n = 81) by model-based generation of synthetic activity maps from recombination of shared network topographies. To study a use case, we formally revisited the poorly understood difference between neural activity underlying idling versus goal-directed behavior. We demonstrate that task-specific neural activity patterns can be explained by plausible combinations of resting-state networks. The possibility of decomposing a mental task into the relative contributions of major brain networks, the "network co-occurrence architecture" of a given task, opens an alternative access to the neural substrates of human cognition.

See Fig. 8 for an illustration and [6] for more information.

6.7. Transmodal Learning of Functional Networks for Alzheimer's Disease Prediction

Functional connectivity describes neural activity from resting-state functional magnetic resonance imaging (rs-fMRI). This noninvasive modality is a promising imaging biomarker of neurodegenerative diseases, such as Alzheimer's disease (AD), where the connectome can be an indicator to assess and to understand the pathology. However, it only provides noisy measurements of brain activity. As a consequence, it has shown fairly limited discrimination power on clinical groups. So far, the reference functional marker of AD is the fluorodeoxyglucose positron emission tomography (FDG-PET). It gives a reliable quantification of metabolic activity, but it is costly and invasive. Here, our goal is to analyze AD populations solely based on rs-fMRI, as functional connectivity is correlated to metabolism. We introduce transmodal learning: leveraging a prior

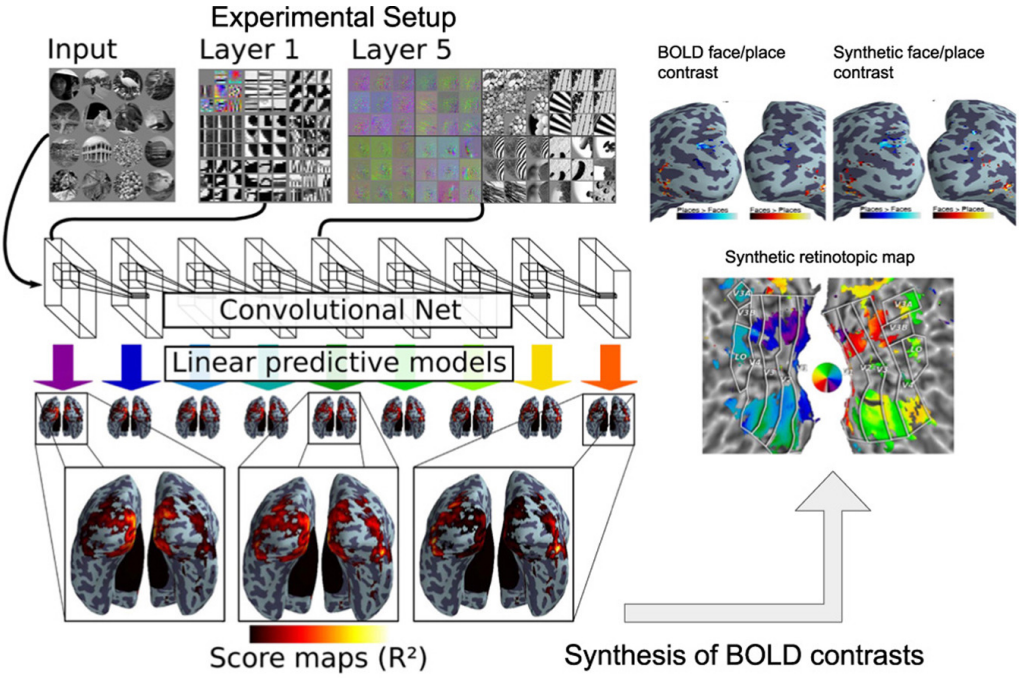


Figure 7. Overview of the vision mapping experiment: Convolutional network image representations of different layer depth explain brain activity throughout the full ventral visual stream. This mapping follows the known hierarchical organisation. Results from both static images and video stimuli. A model of brain activity for the full brain, based on the convolutional network, can synthesize brain maps for other visual experiments. Only deep models can reproduce observed BOLD activity. See [10] for more information.

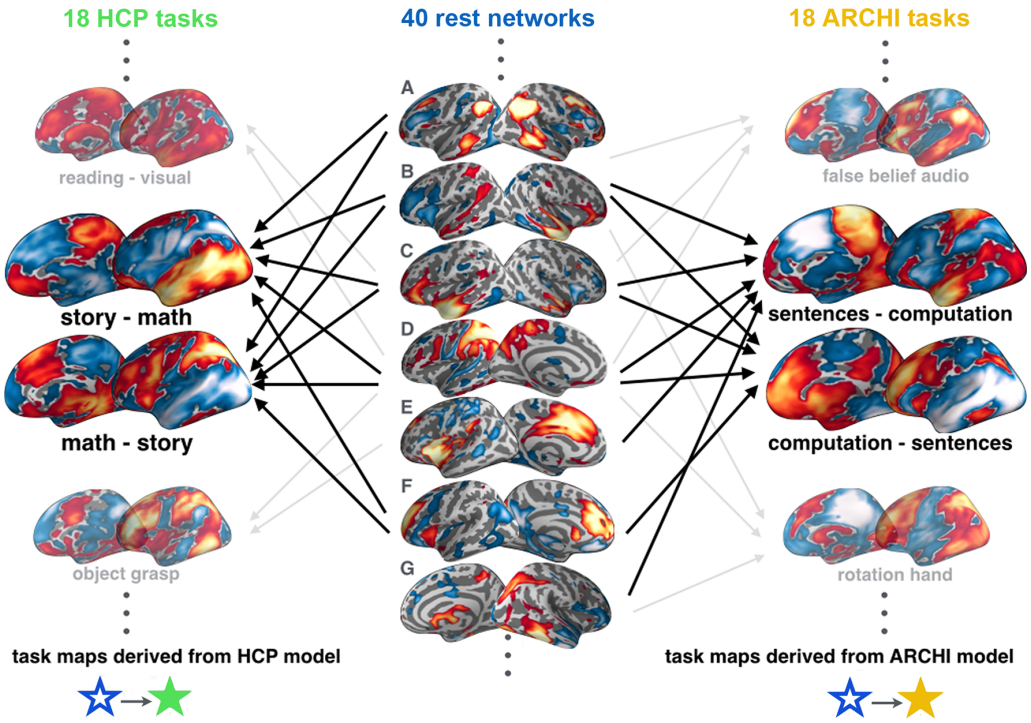


Figure 8. Task-rest correspondence: Reconstructing two similar tasks from two different datasets based on the same resting networks. 40 sparse PCA networks were discovered from the same rest data and used for feature engineering as a basis for classificationof 18 psychological tasks from HCP (left) and from ARCHI (right). Middle column: Examples of resting-state networks derived from decomposing rest data using sparse PCA. Networks B and C might be related to semantics processing in the anterior temporal lobe, network D covers extended parts of the parietal cortex, while networks E and F appear to be variants of the so-called "salience" network. Left/Right column: Examples of task-specific neural activity generated from network co-occurrence models of the HCP/ARCHI task batteries. Arrows: A diagnostic subanalysis indicated what rest networks were automatically ranked top-five in distinguishing a given task from the respective 17 other tasks. Although the experimental tasks in the HCP and ARCHI repositories, "story versus math" and "sentences versus computation" were the most similar cognitive contrasts in both datasets. For these four experimental conditions the model-derived task maps are highly similar. Consequently, two independent classification problems in two independent datasets with a six-fold difference in sample size resulted in two independent explicit models that, nevertheless, generated comparable task-specific maps. This indicated that network co-occurrence modeling indeed captures genuine aspects of neurobiology rather than arbitrary discriminatory aspects of the data. See [6] for more information.

from one modality to improve results of another modality on different subjects. A metabolic prior is learned from an independent FDG-PET dataset to improve functional connectivity-based prediction of AD. The prior acts as a regularization of connectivity learning and improves the estimation of discriminative patterns from distinct rs-fMRI datasets. Our approach is a two-stage classification strategy that combines several seed-based connectivity maps to cover a large number of functional networks that identify AD physiopathology. Experimental results show that our transmodal approach increases classification accuracy compared to pure rs-fMRI approaches, without resorting to additional invasive acquisitions. The method successfully recovers brain regions known to be impacted by the disease.

6.8. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines

Decoding, ie prediction from brain images or signals, calls for empirical evaluation of its predictive power. Such evaluation is achieved via cross-validation, a method also used to tune decoders' hyper-parameters. This paper is a review on cross-validation procedures for decoding in neuroimaging. It includes a didactic overview of the relevant theoretical considerations. Practical aspects are highlighted with an extensive empirical study of the common decoders in within-and across-subject predictions, on multiple datasets –anatomical and functional MRI and MEG– and simulations. Theory and experiments outline that the popular " leave-one-out " strategy leads to unstable and biased estimates, and a repeated random splits method should be preferred. Experiments outline the large error bars of cross-validation in neuroimaging settings: typical confidence intervals of 10%. Nested cross-validation can tune decoders' parameters while avoiding circularity bias. However we find that it can be more favorable to use sane defaults, in particular for non-sparse decoders.

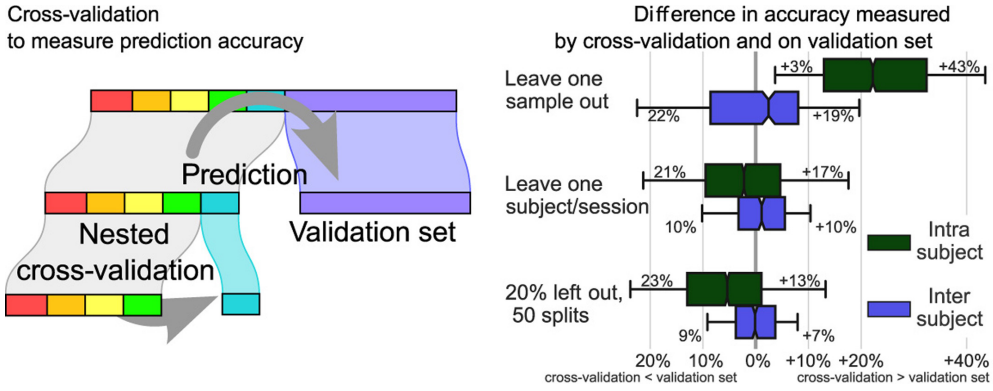


Figure 9. (Left) Illustration of the nested cross-validation principle. (Right) Typical cross-validated accuracy result: leave-one-out cross validation, when applied to imaging data, yields to optimistic bias (top) when used on dependent data, and in other cases leads to estimated with inflated variance. See [16] for more information.

See Fig. 9 for an illustration and [16] for more information.

6.9. A projection algorithm for gradient waveforms design in Magnetic Resonance Imaging

Collecting the maximal amount of information in a given scanning time is a major concern in Magnetic Resonance Imaging (MRI) to speed up image acquisition. The hardware constraints (gradient magnitude, slew rate, ...), physical distortions (e.g., off-resonance effects) and sampling theorems (Shannon, compressed sensing) must be taken into account simultaneously, which makes this problem extremely challenging. To date,

the main approach to design gradient waveform has consisted of selecting an initial shape (e.g. spiral, radial lines, ...) and then traversing it as fast as possible using optimal control. In this paper, we propose an alternative solution which first consists of defining a desired parameterization of the trajectory and then of optimizing for minimal deviation of the sampling points within gradient constraints. This method has various advantages. First, it better preserves the density of the input curve which is critical in sampling theory. Second, it allows to smooth high curvature areas making the acquisition time shorter in some cases. Third, it can be used both in the Shannon and CS sampling theories. Last, the optimized trajectory is computed as the solution of an efficient iterative algorithm based on convex programming. For piecewise linear trajectories, as compared to optimal control reparameterization, our approach generates a gain in scanning time of 10% in echo planar imaging while improving image quality in terms of signal-to-noise ratio (SNR) by more than 6 dB. We also investigate original trajectories relying on traveling salesman problem solutions. In this context, the sampling patterns obtained using the proposed projection algorithm are shown to provide significantly better reconstructions (more than 6 dB) while lasting the same scanning time.

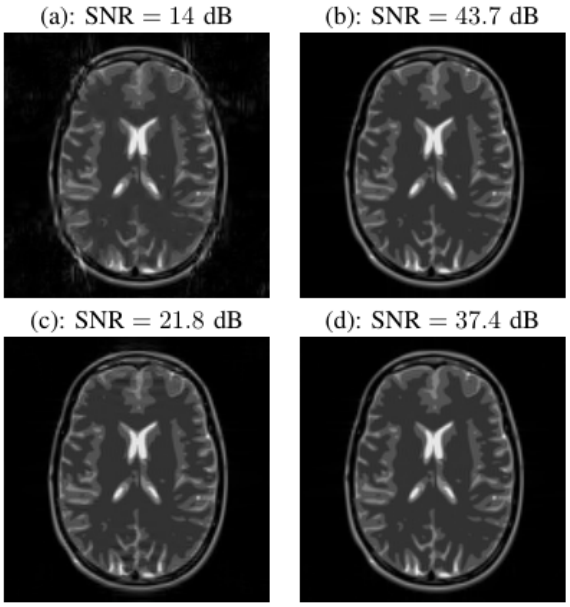


Figure 10. Reconstructed images from data collected along EPI-like trajectories. (a)-(b): Reconstruction results from the optimally reparameterized EPI readout. (c)-(d): Reconstructed results from data collected using the projected EPI trajectories. See [9] for more information.

See Fig. 10 for an illustration and [9] for more information.

6.10. Impact of perceptual learning on resting-state fMRI connectivity: A supervised classification study

Perceptual learning sculpts ongoing brain activity. This finding has been observed by statistically comparing the functional connectivity (FC) patterns computed from resting-state functional MRI (rs-fMRI) data recorded before and after intensive training to a visual attention task. Hence, functional connectivity serves a dynamic role in brain function, supporting the consolidation of previous experience. Following this line of research, we trained three groups of individuals to a visual discrimination task during a magneto-encephalography (MEG) experiment. The same individuals were then scanned in rs-fMRI. Here, in a supervised classification

framework, we demonstrate that FC metrics computed on rs-fMRI data are able to predict the type of training the participants received. On top of that, we show that the prediction accuracies based on tangent embedding FC measure outperform those based on our recently developed multivariate wavelet-based Hurst exponent estimator, which captures low frequency fluctuations in ongoing brain activity too.

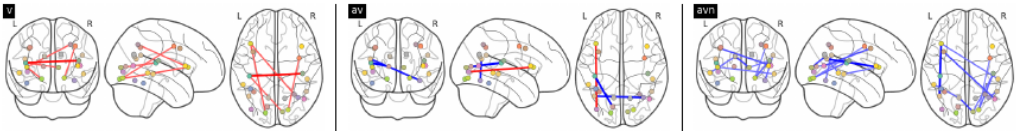


Figure 11. Statistical significant functional interactions (positive and negative values are color coded in red and blue, respectively) within each group of individuals (V: purely visual traing, AV: audio-visual training and AVn: unmatched audio-visual), Bonferroni-corrected for multiple comparisons at $\alpha = 0.05$. *See [24] for more information.*

See Fig. 11 for an illustration and [24] for more information.

SISTM Project-Team

7. New Results

7.1. High dimensional data

Approaches Applied in Genomics Context [13]
Motivation: The association between two blocks of ?omics? data brings challenging issues in computational biology due to their size and complexity. Here, we focus on a class of multivariate statistical methods called partial least square (PLS). Sparse version of PLS (sPLS) operates integration of two datasets while simultaneously selecting the contributing variables. However, these methods do not take into account the important structural or group effects due to the relationship between markers among biological pathways. Hence, considering the predefined groups of markers (e.g. genesets), this could improve the relevance and the efficacy of the PLS approach. Results: We propose two PLS extensions called group PLS (gPLS) and sparse gPLS (sgPLS). Our algorithm enables to study the relationship between two different types of omics data (e.g. SNP and gene expression) or between an omics dataset and multivariate phenotypes (e.g. cytokine secretion). We demonstrate the good performance of gPLS and sgPLS compared with the sPLS in the context of grouped data. Then, these methods are compared through an HIV therapeutic vaccine trial. Our approaches provide parsimonious models to reveal the relationship between gene abundance and the immunological response to the vaccine.

Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure [26]
High-dimensional data classification is a challenging problem. A standard approach to tackle this problem is to perform variables selection, e.g. using step-wise or LASSO procedures. Another standard way is to perform dimension reduction, e.g. by Principal Component Analysis or Partial Least Square procedures. The approach proposed in this paper combines both dimension reduction and variables selection. First, a procedure of clustering of variables is used to built groups of correlated variables in order to reduce the redundancy of information. This dimension reduction step relies on the R package ClustOfVar which can deal with both numerical and categorical variables. Secondly, the most relevant synthetic variables (which are numerical variables summarizing the groups obtained in the first step) are selected with a procedure of variable selection using random forests, implemented in the R package VSURF. Numerical performances of the proposed methodology called CoV/VSURF are compared with direct applications of VSURF or random forests on the original p variables. Improvements obtained with the CoV/VSURF procedure are illustrated on two simulated mixed datasets (cases n>p and n<p)

Arbres CART et Forêts aléatoires, Importance et sélection de variables [27]
Two algorithms proposed by Leo Breiman : CART trees (Classification And Regression Trees for) introduced in the first half of the 80s and random forests emerged, meanwhile, in the early 2000s, are the subject of this article. The goal is to provide each of the topics, a presentation, a theoretical guarantee, an example and some variants and extensions. After a preamble, introduction recalls objectives of classification and regression problems before retracing some predecessors of the Random Forests. Then, a section is devoted to CART trees then random forests are presented. Then, a variable selection procedure based on permutation variable importance is proposed. Finally the adaptation of random forests to the Big Data context is sketched.

Comments on: " A Random Forest Guided Tour " [8]
This paper is a comment on the survey paper by Biau and Scornet (2016) about random forests. We focus on the problem of quantifying the impact of each ingredient of random forests on their performance. We show that such a quantification is possible for a simple pure forest , leading to conclusions that could apply more generally. Then, we consider " hold-out " random forests, which are a good middle point between " toy " pure forests and Breiman's original random forests.

Targeting HIV-1 Env gp140 to LOX-1 Elicits Immune Responses in Rhesus Macaques. [18]
Improved antigenicity against HIV-1 envelope (Env) protein is needed to elicit vaccine-induced protective immunity in humans. Here we describe the first tests in non-human primates (NHPs) of Env gp140 protein fused to a humanized anti-LOX-1 recombinant antibody for delivering Env directly to LOX-1-bearing antigen presenting cells, especially dendritic cells (DC). These data, as well as the safety of this protein vaccine, justify further exploration of this DC-targeting vaccine approach for protective immunity against HIV-1.

Significant changes in HIV-1 Capsid stability induced by common CTL-driven viral sequence mutations. [46]
HIV-1-infected individuals with protective HLA class I alleles exhibit better control of viremia and slower disease progression. Virus control in these individuals has been associated with strong and potent HIV-1-specific cytotoxic-T-lymphocyte (CTL) responses restricted by protective HLA alleles, but control of viremia also occurs in the presence of selected CTL escape mutations. Taken together, these data demonstrate that CTL-driven escape mutations within p24 Gag restricted by protective HLA class I alleles have a significant impact on capsid stability that might contribute to the persistent control of viral replication observed despite viral escape from CTL responses.

Optimization and evaluation of luminex performance with supernatants of Peripheral Blood Mononuclear Cell culture. [48]
The Luminex bead-based multiplex assay is useful for quantifying immune mediators such as cytokines and chemokines. Cross-comparisons of reagents for this technique from different suppliers have already been performed using serum or plasma but rarely with supernatants collected from antigen-stimulated peripheral blood mononuclear cells (PBMC). Here, we first describe an optimization protocol for cell culture including quantity of cells and culture duration to obtain reproducible cytokine and chemokine quantifications. Then, we compared three different Luminex kit suppliers.

7.2. Modeling biomarkers and Mecanistic modeling

- **Dynamic models for estimating the effect of HAART on CD4 in observational studies: Application to the Aquitaine Cohort and the Swiss HIV Cohort Study. [15]**
  Highly active antiretroviral therapy (HAART) has proved efficient in increasing CD4 counts in many randomized clinical trials. Because randomized trials have some limitations (e.g., short duration, highly selected subjects), it is interesting to assess the effect of treatments using observational studies. This is challenging because treatment is started preferentially in subjects with severe conditions. This general problem had been treated using Marginal Structural Models (MSM) relying on the counterfactual formulation. Another approach to causality is based on dynamical models. We present three discrete-time dynamic models based on linear increments models (LIM): the first one based on one difference equation for CD4 counts, the second with an equilibrium point, and the third based on a system of two difference equations, which allows jointly modeling CD4 counts and viral load. We also consider continuous-time models based on ordinary differential equations with non-linear mixed effects (ODE-NLME). These mechanistic models allow incorporating biological knowledge when available, which leads to increased statistical evidence for detecting treatment effect. Because inference in ODE-NLME is numerically challenging and requires specific methods and softwares, LIM are a valuable intermediary option in terms of consistency, precision, and complexity. We compare the different approaches in simulation and in illustration on the ANRS CO3 Aquitaine Cohort and the Swiss HIV Cohort Study.

- **Use of dynamical models for treatment optimization in HIV infected patients : a sequential Bayesian analysis approach. [15]**
  The use of dynamic mechanistic models based on ordinary differential equations (ODE) has greatly improved the knowledge of the dynamics of HIV and of the immune system. Their flexibility for fitting data and prediction abilities make them a good tool for optimization of the design delivery and efficacy of new intervention in the HIV field. We present the problem of inference in ODE models with mixed effects on parameters. We introduce a Bayesian estimation procedure based on the maximization of the penalized likelihood and a normal approximation of posteriors, which is implemented in the NIMROD software. We investigate the impact of pooling different data by using a sequential Bayesian analysis (SBA), which uses posteriors of a previous study as new priors. We show that the normal approximation of the posteriors, which constrains the shape of new priors, leads to gains in accuracy of estimation while reducing computation times. The illustration is from two clinical trials of combination of antiretroviral therapies (cART): ALBI ANRS 070 and PUZZLE ANRS 104. This paper reproduces some unpublished work from my PhD thesis. It is an extension of my oral presentation on the same topic at the 47th Journées de Statistique organized by the French Statistical Society (SFdS) in Lille, France, May 2015, when being awarded the Marie-Jeanne Laurent-Duhamel prize.
- **Surveillance of $\gamma\delta$T Cells Predicts Cytomegalovirus Infection Resolution in Kidney Transplants. [11]**
  Cytomegalovirus (CMV) infection in solid-organ transplantation is associated with increased morbidity and mortality, particularly if a CMV mutant strain with antiviral resistance emerges. Monitoring CMV specific T cell response could provide relevant information for patient care. We assessed if V delta 2 neg gamma delta T cell kinetics in peripheral blood predict CMV infection resolution and emergence of a mutant strain in high risk recipients of kidney transplants, including 168 seronegative recipients receiving organs from seropositive donors and 104 seropositive recipients receiving antithymocyte globulins (R+/ATG). In conclusion, longitudinal surveillance of V delta 2 neg gamma delta T cells in recipients of kidney transplants may predict CMV infection resolution and antiviral drug resistance.
- **Early CD4+ T Cell Responses Are Associated with Subsequent CD8+ T Cell Responses to an rAd5-Based Prophylactic Prime-Boost HIV Vaccine Strategy. [12]**
  Initial evaluation of a candidate vaccine against HIV includes an assessment of the vaccine's ability to generate immune responses. However, the dynamics of vaccine-induced immune responses are unclear. We hypothesized that the IFN-gamma producing cytotoxic CD8+ T cell responses could be predicted by early IL-2 producing CD4+ helper T cell responses, and we evaluated this hypothesis using data from a phase I/II prophylactic HIV vaccine trial. The objective was to assess the dynamics after vaccination with a recombinant adenoviral serotype 5 (rAd5) HIV vaccine. Regression models confirmed this relationship with a significant association between the two markers. These results suggest an early and leading role of CD4+ T cells in the cellular response to the rAd5-rAd5 vaccine and in particular the stimulation of cytotoxic CD8+ T cell responses. These results could inform better timing of CD4+ T cell measurements in future clinical trials.
- **Reference curves for CD4 T-cell count response to combination antiretroviral therapy in HIV-1-infected treatment-naïve patients. [29]**
  The aim of this work was to provide a reference for the CD4 T-cell count response in the early months after the initiation of combination antiretroviral therapy (cART) in HIV-1-infected patients. All patients in the Collaboration of Observational HIV Epidemiological Research Europe (COHERE) cohort who were aged > 18 years and started cART for the first time between 1 January 2005 and 1 January 2010 and who had at least one available measurement of CD4 count and a viral load < 50 HIV-1 RNA copies/mL at 6 months (+- 3 months) after cART initiation were included in the study. Unadjusted and adjusted references curves and predictions were obtained using quantile regressions. Reference curves aid the evaluation of the immune response early after antiretroviral therapy initiation that leads to viral control.

- **Repeated Cycles of Recombinant Human Interleukin 7 in HIV-Infected Patients With Low CD4 T-Cell Reconstitution on Antiretroviral Therapy: Results of 2 Phase II Multicenter Studies. [17].**
  Phase I/II studies in human immunodeficiency virus (HIV) infected patients receiving antiretroviral therapy have shown that a single cycle of 3 weekly subcutaneous injections of recombinant human interleukin 7 (r-hIL-7) is safe and improves immune CD4 T-cell restoration. Herein, we report data from 2 phase II trials evaluating the effect of repeated cycles of r-hIL-7 (20 microg/kg) with the objective of restoring a sustained CD4 T-cell count >500 cells/microL. INSPIRE 2 was a single-arm trial conducted in the United States and Canada. INSPIRE 3 was a 2 arm trial with 3:1 randomization to r-hIL-7 versus control conducted in Europe and South Africa. Participants with plasma HIV RNA levels <50 copies/mL during antiretroviral therapy and with CD4 T-cell counts between 101 and 400 cells/microL were eligible. A repeat cycle was administered when CD4 T-cell counts fell to <550 cells/microL. A total of 107 patients were treated and received 1 (n = 107), 2 (n = 74), 3 (n = 14), or 4 (n = 1) r-hIL-7 cycles during a median follow-up of 23 months. r-hIL-7 was well tolerated. Four grade 4 events were observed, including 1 case of asymptomatic alanine aminotransferase elevation. After the second cycle, anti-r-hIL-7 binding antibodies developed in 82% and 77% of patients in INSPIRE 2 and 3, respectively (neutralizing antibodies in 38% and 37%), without impact on the CD4 T-cell response. Half of the patients spent >63% of their follow-up time with a CD4 T-cell count >500 cells/microL. CONCLUSIONS: Repeated cycles of r-hIL-7 were well tolerated and achieved sustained CD4 T-cell restoration to >500 cells/microL in the majority of study participants.

7.3. Implication in analysis of results from Clinical trials and cohorts

- **Superior efficacy of an HIV vaccine combined with ARV prevention in SHIV challenged non-human primates. [38]**
  Although vaccines and antiretroviral (ARV) prevention have demonstrated partial success against human immunodeficiency virus (HIV) infection in clinical trials, their combined introduction could provide more potent protection. Furthermore, combination approaches could ameliorate the potential increased risk of infection following vaccination in the absence of protective immunity. We used a nonhuman primate model to determine potential interactions of combining a partially effective ARV microbicide with an envelope-based vaccine. These important findings suggest that combined implementation of new biomedical prevention strategies may provide significant gains in HIV prevention.
- **A Method to Estimate the Size and Characteristics of HIV-positive Populations Using an Individual-based Stochastic Simulation Model. [14]**
  It is important not only to collect epidemiologic data on HIV but to also fully utilize such information to understand the epidemic over time and to help inform and monitor the impact of policies and interventions. We describe and apply a novel method to estimate the size and characteristics of HIV-positive populations. In the pseudo-epidemic example, HIV estimates have narrower plausibility ranges and are closer to the true number, the greater the data availability to calibrate the model. We demonstrate that our method can be applied to settings with less data, however plausibility ranges for estimates will be wider to reflect greater uncertainty of the data used to fit the model.
- **Immunologic response in treatment-naïve HIV-2-infected patients: the IeDEA West Africa cohort. [9]**
  Response to antiretroviral therapy (ART) among individuals infected with HIV-2 is poorly described. We compared the immunological response among patients treated with three nucleoside reverse-transcriptase inhibitors (NRTIs) to boosted protease inhibitor (PI) and unboosted PI-based regimens in West Africa. In this observational study using African data, boosted PI-containing regimens had better immunological response compared to triple NRTI combinations and unboosted PI-based regimens at 12 months. A randomized clinical trial is still required to determine the best initial regimen for treating HIV-2 infected patients.

- **Intrinsic defect in keratinocyte function leads to inflammation in Hidradenitis suppurativa. [10]**
  Hidradenitis suppurativa (HS) is a chronic, inflammatory, debilitating, follicular disease of the skin. Despite a high prevalence in the general population, the physiopathology of HS remains poorly understood. The use of antibiotics and immunosuppressive agents for therapy suggests a deregulated immune response to microflora. These findings point out a functional defect of keratinocytes in HS leading to a balance prone to inflammatory responses. This is likely to favor a permissive environment for bacterial infections and chronic inflammation characterizing clinical outcomes in patients with HS.
- **Uptake of Home-Based HIV Testing, Linkage to Care, and Community Attitudes about ART in Rural KwaZulu-Natal, South Africa: Descriptive Results from the First Phase of the ANRS 12249 TasP Cluster-Randomised Trial. [34]**
  The 2015 WHO recommendation of antiretroviral therapy (ART) for all immediately following HIV diagnosis is partially based on the anticipated impact on HIV incidence in the surrounding population. We investigated this approach in a cluster-randomised trial in a high HIV prevalence setting in rural KwaZulu-Natal. We present findings from the first phase of the trial and report on uptake of home-based HIV testing, linkage to care, uptake of ART, and community attitudes about ART. Home-based HIV testing was well received in this rural population, although men were less easily contactable at home; immediate ART was acceptable, with good viral suppression and retention. However, only about half of HIV-positive people accessed care within 6 mo of being identified, with nearly two-thirds accessing care by 12 mo. The observed delay in linkage to care would limit the individual and public health ART benefits of universal testing and treatment in this population.

7.4. Conferences

Members of the team were involved in more than 20 talks during conferences and colloquium. In particular, [20], [21], [23], [19], [22], [24] and [25] have proceedings.

VISAGES Project-Team

7. New Results

7.1. Image Computing: Detection, Segmentation, Registration and Analysis

7.1.1. Quantitative analysis of T2/T2 relaxation time alteration*

Participants: Benoit Combès, Anne Kerbrat, Olivier Commowick, Christian Barillot.

T2 and T2* relaxometric data [0] becomes a standard tool for the quantitative assessment of brain tissues and of their changes along time or after the infusion of a contrast agent. Being able to detect significant changes of T2/T2* relaxation time is an important issue. Generally, such a task is performed by comparing the variability level in the regions of interest to the variability in the normal appearance white matter. However, in the case of T2 and T2* relaxometry, this solution is highly problematic. Indeed the level of noise in the normal appearance white matter is significanlty smaller than the level of noise in more intense region (e.g. MS lesions). Our aim is to provide a Bayesian analysis of T2/T2* relaxometry estimation and alteration. More specifically, we build posterior distributions for the relaxation time and the relaxation offset by elucidating the dedicated Jeffreys priors. Then the resulting posterior distributions can be evaluated using a Monte Carlo Markov Chain algorithm. Such an analysis has three main advantages over the classical estimation procedure. First it allows in a simple way to compute many estimators of the posterior including the mode, the mean, the variance and confidence intervals. Then, it allows to include prior information. Finally, becauce one can extract confidence interval from the posterior, testing properly whether the true relaxometry time is included within a certain range of value given a confidence level is simple. This work was published as a conference paper in MICCAI 2016 [22].

7.1.2. 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 accepted for publication in IEEE Transactions in Medical Imaging 2017.

7.1.3. An a contrario approach for the detection of patient-specific brain perfusion abnormalities with arterial spin labelling

Participants: Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

[0] https://en.wikipedia.org/wiki/Relaxation_(NMR)

In this work, we introduce a new locally multivariate procedure to quantitatively extract voxel-wise patterns of abnormal perfusion in individual patients. This a contrario approach uses a multivariate metric from the computer vision community that is suitable to detect abnormalities even in the presence of closeby hypo- and hyper-perfusions. This method takes into account local information without applying Gaussian smoothing to the data. Furthermore, to improve on the standard a contrario approach, which assumes white noise, we introduce an updated a contrario approach that takes into account the spatial coherency of the noise in the probability estimation. Validation is undertaken on a dataset of 25 patients diagnosed with brain tumors and 61 healthy volunteers. We show how the a contrario approach outperforms the massively univariate General Linear Model usually employed for this type of analysis. This work as been published in Neuroimage [14].

7.1.4. Dictionary Learning for Pattern Classification in Medical Imaging: Why Does Size Matter?

Participants: Hrishikesh Deshpande, Pierre Maurel, Christian Barillot.

Sparse representation based dictionary learning (DL) technique has proved to be an effective tool for image classification. While standard DL methods are effective in data representation, several discriminative DL methods have been proposed for learning dictionaries better suited for classification. Majority of these methods, in pattern recognition applications, learn the dictionaries for each class and com- pare the error terms of sparse reconstruction for each dictionary. However this raises a question that is still an open problem in the sparsity community: What role does the size of each dictionary play in the classification process? In this work, we prove that this parameter is pivotal, especially in cases where there are variability differences between classes. We illustrate our assertion on standard and discriminative DL techniques in two applications: Lips detection in face images and the classification of multiple sclerosis lesions in multi-channel brain MR images.

7.2. Image processing on Diffusion Weighted Magnetic Resonance Imaging

7.2.1. Maximum Likelihood Estimators of Brain White Matter Microstructure

Participant: Olivier Commowick.

Diffusion MRI is a key in-vivo non invasive imaging capability that can probe the microstructure of the brain. However, its limited resolution requires complex voxelwise generative models of the diffusion. Diffusion Compartment (DC) models divide the voxel into smaller compartments in which diffusion is homogeneous. We developed a comprehensive framework for maximum likelihood estimation (MLE) of such models that jointly features ML estimators of (i) the baseline MR signal, (ii) the noise variance, (iii) compartment proportions, and (iv) diffusion-related parameters. ML estimators are key to providing reliable mapping of brain microstructure as they are asymptotically unbiased and of minimal variance. We compare our algorithm (which efficiently exploits analytical properties of MLE) to alternative implementations and a state-of-the-art strategy. Simulation results show that our approach offers the best reduction in computational burden while guaranteeing convergence of numerical estimators to the MLE. In-vivo results also reveal remarkably reliable microstructure mapping in areas as complex as the centrum semiovale. Our ML framework accommodates any DC model and is available freely for multi-tensor models as part of the ANIMA software. This work was published as a conference paper in MICCAI 2016 [24].

7.3. EEG and MR Imaging

7.3.1. Multi-Modal EEG and fMRI Source Localization using Sparse Constraints

Participants: Saman Noorzadeh, Pierre Maurel, Christian Barillot.

In this work a multi-modal approach is introduced to estimate the brain neuronal sources based on EEG and fMRI. These two imaging techniques can provide complementary information about the neuronal activities of the brain. Each of these data modalities are first modeled linearly based on the sources. The sources are then estimated with a high spatio-temporal resolution based on a symmetrical integrated approach of these models. For a better estimation, a sparse constraint is also applied to the method based on the physiological knowledge that we have about the brain function. The results which are validated on the real data, shows the reconstruction of neuronal sources with the high spatio-temporal resolution. This is a joint work with Remi Gribonval.

7.3.2. Unimodal versus bimodal EEG-fMRI neurofeedback of a motor imagery task

Participants: Lorraine Perronnet, Marsel Mano, Élise Bannier, Christian Barillot.

In the context of the HEMISFER project, we proposed a simultaneous EEG-fMRI experimental protocol in which 10 healthy participants performed a motor-imagery task in unimodal and bimodal neurofeedback conditions. With this protocol we were able to compare for the first time the effects of unimodal EEG-neurofeedback and fMRI-neurofeedback versus bimodal EEG-fMRI-neurofeedback 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). Such a feedback is intended to relieve the cognitive load of the subject by presenting the bimodal neurofeedback task as a single regulation task instead of two. Additionally, this integrated feedback metaphor gives flexibility on defining a bimodal neurofeedback target. Participants were able to regulate activity in their motor regions in all neurofeedback conditions. Moreover, motor activations as revealed by offline fMRI analysis were stronger during EEG-fMRI-neurofeedback than during EEG-neurofeedback. This result suggests that EEG-fMRI-neurofeedback could be more specific or more engaging than EEG-neurofeedback. Our results also suggest that during EEG-fMRI-neurofeedback, participants tended to regulate more the modality that was harder to control. Taken together our results shed light on the specific mechanisms of bimodal EEG-fMRI-neurofeedback and on its added-value as compared to unimodal EEG-neurofeedback and fMRI-neurofeedback.

This work was done in collaboration with the Inria Hybrid and Athena teams. Experiments were conducted at the Neurinfo MRI research facility from University of Rennes 1. This was presented during the poster session of the 2016 Organization for Human Brain Mapping (OHBM) conference.

7.3.3. Brain training with Neurofeedback

Participants: Lorraine Perronnet, Christian Barillot.

We published a book chapter called Brain training with Neurofeedback in the book “Brain Computer Interfaces 1: Methods and Perspectives” (published in French and English) [26]. The first section of the chapter defines the concept of neurofeedback and gives an overall view of the current status in this domain. The second section describes the design of a NF training program and the typical course of a NF session, as well as the learning mechanisms underlying NF. The third section retraces the history of NF, explaining the origin of its questionable reputation and providing a foothold for understanding the diversity of existing approaches. The fourth section discusses how the fields of NF and BCIs might potentially overlap in future with the development of "restorative" BCIs. Finally, the fifth and last section presents a few applications of NF and summarizes the state of research of some of its major clinical applications.

7.3.4. Design of an Experimental Platform for Hybrid EEG-fMRI Neurofeedback Studies

Participants: Marsel Mano, Élise Bannier, Lorraine Perronnet, Christian Barillot.

During a neurofeedback (NF) experiment one or more brain activity measuring technologies are used to estimate the changes of the acquired neural signals that reflect the changes of the subject’s brain activity in real-time. There exist a variety of NF research applications that use only one type of neural signals (i.e. uni-modal) like EEG or fMRI, but there are very few NF researches that use two or more neural signals (i.e. multi-modal). This is primarily because of the associated technical burdens.

We have developed, installed and successfully conducted used a hybrid EEG-fMRI platform for bi-modal NF experiments, as part of the project Hemisfer. Our system is based on the integration and the synchronization of an MR-compatible EEG and fMRI acquisition subsystems. The EEG signals are acquired with a 64 channel MR-compatible solution from Brain Products and the MR imaging is performed on a 3T Verio Siemens scanner (VB17) with a 12-ch head coil. We have developed two real-time pipelines for EEG and fMRI that handle all the necessary signal processing, the Joint NF module that calculates and fuses the NF and a visualize module that displays the NF to the subject. The control and the synchronization of both subsystems with eachother 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. The entire fMRI process from acquisition to NF takes always less than 200ms, well below the TR of regular EPI sequences (2s). The same process for EEG, with NF update cycles varying 2-5Hz, is done in virtually real time (50Hz). Various NF tasks scenarios for regulating the measured brain activity were tested with subjects. In particular, the platform was used for a NF study on 10 subjects with over 50 sessions using three NF protocols based on motor imagery related brain activity: a) fMRI-NF, b) EEG-NF and c) EEG and fMRI-NF; and two online brain activity regulating protocols without NF. Our hybrid EEG-fMRI NF platform has been a very reliable environment for the NF experiments in our project. Its modular architecture is easily adaptable to different experimental environments, and offers high efficiency for optimal real-time NF applications.

7.4. Applications in Neuroradiology and Neurological Disorders

7.4.1. Imaging biomarkers in Multiple Sclerosis: from image analysis to population imaging

Participants: Christian Barillot, Gilles Edan, Olivier Commowick.

The production of imaging data in medicine increases more rapidly than the capacity of computing models to extract information from it. The grand challenges of better understanding the brain, offering better care for neurological disorders, and stimulating new drug design will not be achieved without significant advances in computational neuroscience. The road to success is to develop a new, generic, computational methodology and to confront and validate this methodology on relevant diseases with adapted computational infrastructures. This new concept sustains the need to build new research paradigms to better understand the natural history of the pathology at the early phase; to better aggregate data that will provide the most complete representation of the pathology in order to better correlate imaging with other relevant features such as clinical, biological or genetic data. In this context, one of the major challenges of neuroimaging in clinical neurosciences is to detect quantitative signs of pathological evolution as early as possible to prevent disease progression, evaluate therapeutic protocols or even better understand and model the natural history of a given neurological pathology. Many diseases encompass brain alterations often not visible on conventional MRI sequences, especially in normal appearing brain tissues (NABT). MRI has often a low specificity for differentiating between possible pathological changes which could help in discriminating between the different pathological stages or grades. The objective of medical image analysis procedures is to define new quantitative neuroimaging biomarkers to track the evolution of the pathology at different levels. We have published a position paper in Medical Image Analysis [2] that illustrates this issue in one acute neuro-inflammatory pathology: Multiple Sclerosis (MS). It exhibits the current medical image analysis approaches and explains how this field of research will evolve in the next decade to integrate larger scale of information at the temporal, cellular, structural and morphological levels.

7.4.2. Multiple Sclerosis lesion segmentation using an automated multimodal Graph Cut

Participants: Jérémy Beaumont, Olivier Commowick, Christian Barillot.

In this work, we present an algorithm for Multiple Sclerosis (MS) lesion segmentation. Our method is fully automated and includes three main steps: 1. the computation of a rough total lesion load in order to optimize the parameter set of the following step; 2. the detection of lesions by graph cut initialized with a robust Expectation-Maximization (EM) algorithm; 3. the application of rules to remove false positives and to adjust the contour of the detected lesions. This work was part of the FLI 2016 MSSEG challenge data organized at MICCAI 2016 [25].

7.4.3. Automatic Multiple Sclerosis lesion segmentation from Intensity-Normalized multi-channel MRI

Participants: Jérémy Beaumont, Olivier Commowick, Christian Barillot.

In the context of the FLI MICCAI 2016 MSSEG challenge for lesion segmentation, we present a fully automated algorithm for Multiple Sclerosis (MS) lesion segmentation. Our method is composed of three main steps. First, the MS patient images are registered and intensity normalized. Then, the lesion segmentation is done using a voxel-wise comparison of multi-channel Magnetic Resonance Images (MRI) against a set of controls. Finally, the segmentation is refined by applying several lesion appearance rules. This work was part of the FLI 2016 MSSEG challenge data organized at MICCAI 2016 [21].

7.5. Management of Information in Neuroimaging

Participants: Michael Kain, Olivier Commowick, Élise Bannier, Inès Fakhfakh, Justine Guillaumont, Florent Leray, Yao Chi, Christian Barillot.

The major topic that is addressed in this period concern the sharing of data and processing tools in neuroimaging (through the “Programme d’Investissement d’Avenir” project such as OFSEP and FLI-IAM) which led to build a suitable architecture to share images and processing tools, started from the NeuroBase project (supported by the French Ministry of Research). Our overall goal within these projects is to set up a computer infrastructure to facilitate the sharing of neuroimaging data, as well as image processing tools, in a distributed and heterogeneous environment. These consortium gathered expertise coming from several complementary domains of expertise: image processing in neuroimaging, workflows and GRID computing, ontology development and ontology-based mediation. 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 As an example, 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 Mangment (IAM) node of the France Life Imaging national infrastructure (FLI-IAM). Our team 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 software platform.

XPOP Team

7. New Results

7.1. Identifiability in mixed effects models

We considered the question of model identifiability within the context of nonlinear mixed effects models. Although there has been extensive research in the area of fixed effects models, much less attention has been paid to random effects models.

In this context we distinguish between theoretical identifiability, in which different parameter values lead to non-identical probability distributions, structural identifiability which concerns the algebraic properties of the structural model, and practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise.

We have explored a number of pharmacokinetic models which are known to be non-identifiable at an individual level but can become identifiable at the population level if a number of specific assumptions on the probabilistic model hold. Essentially if the probabilistic models are different, even though the structural models are non-identifiable, then they will lead to different likelihoods. The findings are supported through simulations.

7.2. Enhanced Method for Diagnosing Pharmacometric Models

For nonlinear mixed-effects pharmacometric models, diagnostic approaches often rely on individual parameters, also called empirical Bayes estimates (EBEs), estimated through maximizing conditional distributions. When individual data are sparse, the distribution of EBEs can "shrink" towards the same population value, and as a direct consequence, resulting diagnostics can be misleading.

Instead of maximizing each individual conditional distribution of individual parameters, we propose to randomly sample them in order to obtain values better spread out over the marginal distribution of individual parameters.

We have evaluated, through diagnostic plots and statistical tests, hypothesis related to the distribution of the individual parameters and shown that the proposed method leads to more reliable results than using the EBEs. In particular, diagnostic plots are more meaningful, the rate of type I error is correctly controlled and its power increases when the degree of misspecification increases. An application to the warfarin pharmacokinetic data confirms the interest of the approach for practical applications.

7.3. A Shrinkage-Thresholding Metropolis Adjusted Langevin Algorithm for Bayesian Variable Selection

We have introduced a new Markov Chain Monte Carlo method for Bayesian variable selection in high dimensional settings. The algorithm is a Hastings-Metropolis sampler with a proposal mechanism which combines a Metropolis Adjusted Langevin (MALA) step to propose local moves associated with a shrinkage-thresholding step allowing to propose new models.

The geometric ergodicity of this new trans-dimensional Markov Chain Monte Carlo sampler was established. An extensive numerical experiment, on simulated and real data, illustrates the performance of the proposed algorithm in comparison with some more classical trans-dimensional algorithms

7.4. Maximum likelihood estimation of a low-order building model

Our objective was to investigate the accuracy of the estimates learned with an open loop model of a building whereas the data is actually collected in closed loop, which corresponds to the true exploitation of buildings. We have proposed a simple model based on an equivalent RC network whose parameters are physically interpretable. We also described the maximum likelihood estimation of these parameters by the EM algorithm, and derived their statistical properties.

The numerical experiments clearly show the potential of the method, in terms of accuracy and robustness. We have emphasized the fact that the estimations are linked to the generating process for the observations, which includes the command system. For instance, the features of the building are correctly estimated if there is a significant gap between the heating and cooling setpoint.

7.5. LP-convergence of a Girsanov theorem based particle filter

We have analyzed the LP-convergence of a previously proposed Girsanov theorem based particle filter for discretely observed stochastic differential equation (SDE) models. We proved the convergence of the algorithm with the number of particles tending to infinity by requiring a moment condition and a step-wise initial condition boundedness for the stochastic exponential process giving the likelihood ratio of the SDEs. The practical implications of the condition are illustrated with an Ornstein–Uhlenbeck model and with a non-linear Bene's model.

7.6. Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding

In the random coefficients binary choice model, a binary variable equals 1 iff an index $X^T\beta$ is positive. The vectors X and β are independent and belong to the sphere S^{d-1} in R^d. We have proven lower bounds on the minimax risk for estimation of the density f_β over Besov bodies where the loss is a power of the $L^p(S^{d-1})$ norm for $1 \leq p \leq \infty$. We have shown that a hard thresholding estimator based on a needlet expansion with data-driven thresholds achieves these lower bounds up to logarithmic factors.

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.

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). In [20], we present a Baroclinic Jet test case set in a zonally reentrant channel that provides a controllable test of a model capacity at resolving submesoscale dynamics. We compare simulations from two models, ROMS and NEMO, at different mesh sizes (from 20 to 2 km). Through a spectral decomposition of kinetic energy and its budget terms, we identify the characteristics of numerical dissipation and effective resolution. It shows that numerical dissipation appears in different parts of a model, especially in spatial advection-diffusion schemes for momentum equations (KE dissipation) and tracer equations (APE dissipation) and in the time stepping algorithms. Effective resolution, defined by scale-selective dissipation, is inadequate to qualify traditional ocean models with low-order spatial and temporal filters, even at high grid resolution. High- order methods are better suited to the concept and probably unavoidable. Fourth-order filters are suited only for grid resolutions less than a few kilometers and momentum advection schemes of even higher-order may be justified. The upgrade of time stepping algorithms (from filtered Leapfrog), a cumbersome task in a model, appears critical from our results, not just as a matter of model solution quality but also of computational efficiency (extended stability range of predictor-corrector schemes). Effective resolution is also shaken by the need for non scale-selective barotropic mode filters and requires carefully addressing the issue of mode splitting errors. Possibly the most surprising result is that submesoscale energy production is largely affected by spurious diapycnal mixing (APE dissipation). This result justifies renewed efforts in reducing tracer mixing errors and poses again the question of how much vertical diffusion is at work in the real ocean.

7.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Mehdi-Pierre Daou, Laurent Debreu, Florian Lemarié, Charles Pelletier, Antoine Rousseau.

7.1.2.1. Coupling heterogeneous models in hydrodynamics

The coupling of models of different kinds is gaining more and more attention, due in particular to a need for more global modeling systems encompassing different disciplines (e.g. multi-physics) and different approaches (e.g. multi-scale, nesting). In order to develop such complex systems, it is generally more pragmatic to assemble different modeling units inside a user friendly modelling software platform rather than to develop new complex global models.

In the context of hydrodynamics, global modeling systems have to couple models of different dimensions (1D, 2D or 3D) and representing different physics (Navier-Stokes, hydrostatic Navier-Stokes, shallow water...). We have been developing coupling approaches for several years, based on so-called Schwarz algorithms. Our recent contributions address the development of absorbing boundary conditions for Navier-Stokes equations [4], and of interface conditions for coupling hydrostatic and nonhydrostatic Navier-Stokes flows [5]. In the context of our partnership with ARTELIA Group (PhD thesis of Medhi Pierre Daou), implementations of Schwarz coupling algorithms have been performed for hydrodynamics industrial codes (Mascaret, Telemac and OpenFoam), using the PALM coupling software. A first series of experiments was realized in an academic test case, and a second one in the much more realistic context of the Rusumo hydropower plant, coupling Telemac-3D (Navier-Stokes equations) with OpenFoam (diphasic solver) - see Figure 1 . M.-P. Daou defended his PhD on September 27, 2016 [1].

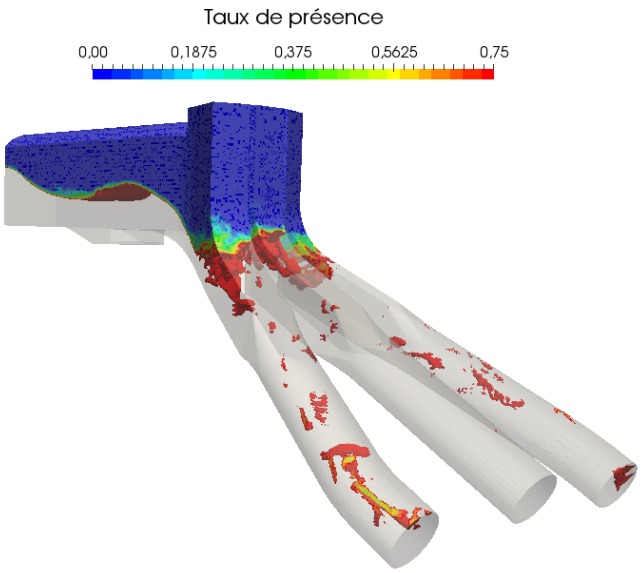


Figure 1. Biphasic simulation of the air/water flow in the Rusumo hydropower plant (PhD of M. P. Daou)

7.1.2.2. Ocean-atmosphere coupling

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approximation of one [63]. For the last few years we have been actively working on the analysis of Schwarz waveform relaxation to apply this type of iterative coupling method to air-sea coupling [65], [66], [64]. In the context of the simulation of tropical cyclone, sensitivity tests to the coupling method have been carried out using ensemble simulations (through perturbations of the coupling frequency and initial conditions). We showed that the use of the Schwarz iterative coupling methods leads to a significantly reduced spread in the ensemble results (in terms of cyclone trajectory and intensity), thus suggesting that a source of error is removed w.r.t coupling methods en vogue in existing coupled models [68].

Motivated by this encouraging result, our activities over the last few years can be divided into four general topics

1. *Stability and consistency analysis of existing coupling methods*: in [63] 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 last year in [3] 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. 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. A publication is currently in preparation for the Quarterly Journal of the Royal Meteorological Society. In parallel we have contributed to a general review gathering the main international specialists on the topic [38].

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 in the framework of the SIMBAD project and should be coupled in early 2017 (collaboration with Mercator-océan).
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. [67]) in large-scale realistic ocean-atmosphere coupled simulations [16], [15].

These four topics are addressed through strong collaborations between the applied mathematics 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. 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 will officially start in January 2017.

7.1.2.3. 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. So far, three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a simple coupled problem. 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 [28].

The aforementioned system has been recoded within the OOPS framework (Object Oriented Prediction System) in order to ease the transfer to more complex/realistic models.

7.1.3. 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 [40].

7.2. Model reduction / multiscale algorithms

7.2.1. Multigrid Methods for Variational Data Assimilation.

Participants: Laurent Debreu, François-Xavier Le Dimet, Arthur Vidard.

In order to lower the computational cost of the variational data assimilation process, we investigate the use of multigrid methods to solve the associated optimal control system. On a linear advection equation, we study the impact of the regularization term on the optimal control and the impact of discretization errors on the efficiency of the coarse grid correction step. We show that even if the optimal control problem leads to the solution of an elliptic system, numerical errors introduced by the discretization can alter the success of the multigrid methods. The view of the multigrid iteration as a preconditioner for a Krylov optimization method leads to a more robust algorithm. A scale dependent weighting of the multigrid preconditioner and the usual background error covariance matrix based preconditioner is proposed and brings significant improvements. This work is summarized in ([7]).

7.2.2. 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 [61], 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 [62]. In [60], 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 [11].

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 [10], 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 [11]. In [32], we propose a generalization of the probabilistic goal-oriented error estimation in [10] 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 for Forecasting Ocean Models

Participants: Eric Blayo, Laurent Gilquin, Céline Helbert, François-Xavier Le Dimet, Elise Arnaud, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Laurence Viry, Federico Zertuche.

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. Data assimilation and second order sensitivity analysis

Sensitivity Analysis is defined by some scalar response function giving an evaluation of the state of a system with respect to parameters. By definition, sensitivity is the gradient of this response function. In the case of Variational Data Assimilation, sensitivity analysis have to be carried out on the optimality system because this is the only system in which all the information is located. An important application is the sensitivity, for instance, of the prediction with respect to observations. It's necessary to derive the optimality system and to introduce a second order adjoint.We have applied it to a simulated pollution transport problem and in the case of an oceanic model [18], [19]. More applications to water pollution using a complex hydrological model are under development.

7.3.2. Estimating variance-based sensitivity indices

Participants: Elise Arnaud, Laurent Gilquin, Clémentine Prieur, Simon Nanty, Céline Helbert, Laurence Viry.

In variance-based sensitivity analysis, a classical tool is the method of Sobol' [74] which allows to compute Sobol' indices using Monte Carlo integration. One of the main drawbacks of this approach is that the estimation of Sobol' indices requires the use of several samples. For example, in a d-dimensional space, the estimation of all the first-order Sobol' indices requires $d+1$ samples. Some interesting combinatorial results have been introduced to weaken this defect, in particular by Saltelli [72] and more recently by Owen [70] but the quantities they estimate still require $O(d)$ samples.

In a recent work [80] we introduce a new approach to estimate all first-order Sobol' indices by using only two samples based on replicated latin hypercubes and all second-order Sobol' indices by using only two samples based on replicated randomized orthogonal arrays. This method is referred as the replicated method. We establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, we propose to apply this new approach to a marine ecosystem model of the Ligurian sea (northwestern Mediterranean) in order to study the relative importance of its several parameters. The calibration process of this kind of chemical simulators is well-known to be quite intricate, and a rigorous and robust — i.e. valid without strong regularity assumptions — sensitivity analysis, as the method of Sobol' provides, could be of great help. The computations are performed by using CIGRI, the middleware used on the grid of the Grenoble University High Performance Computing (HPC) center. We are also applying these estimates to calibrate integrated land use transport models. As for these models, some groups of inputs are correlated, Laurent Gilquin extended the approach based on replicated designs for the estimation of grouped Sobol' indices [58].

We can now wonder what are the asymptotic properties of these new estimators, or also of more classical ones. In [60], the authors deal with asymptotic properties of the estimators. In [57], the authors establish also a multivariate central limit theorem and non asymptotic properties.

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.

To address this challenge, we propose 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 [36]. A second approach combines the use of quasi-Monte Carlo sampling and the construction of a new stopping criterion [9], [39].

Extension of the replication method has also been proposed to face constraints arising in an application on the land use and transport model Tranus, such as the presence of dependency among the model inputs, as well as multivariate outputs [37].

7.3.2.1. 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 [59] in the case where input parameters are correlated. Clémentine Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [49]. We obtained first results [50], 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 [48], [51]. We also considered (see the paragraph 7.3.2) the estimation of groups Sobol' indices, with a procedure based on replicated designs. These indices provide information at the level of groups, and not at a finer level, but their interpretation is still rigorous.

Céline Helbert and Clémentine Prieur 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 [14]. The whole methodology proposed during the PhD is presented in [13].

More recently, the Shapley value, from econometrics, was proposed as an alternative to quantify the importance of random input variables to a function. Owen [71] derived Shapley value importance for independent inputs and showed that it is bracketed between two different Sobol' indices. Song et al. [75] recently advocated the use of Shapley value for the case of dependent inputs. In a very recent work [42], 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.

7.3.3. Optimal Control of Boundary Conditions

Participants: Eric Blayo, Eugene Kazantsev, Florian Lemarié.

A variational data assimilation technique is applied to the identification of the optimal boundary conditions for a simplified configuration of the NEMO model. A rectangular box model placed in mid-latitudes, and subject to the classical single or double gyre wind forcing, is studied. The model grid can be rotated on a desired angle around the center of the rectangle in order to simulate the boundary approximated by a staircase-like coastlines. The solution of the model on the grid aligned with the box borders was used as a reference solution and as artificial observational data. It is shown in [24], [25] that optimal boundary has a rather complicated geometry which is neither a staircase, nor a straight line. The boundary conditions found in the data assimilation procedure bring the solution toward the reference solution allowing to correct the influence of the rotated grid.

Adjoint models, necessary to variational data assimilation, have been produced by the TAPENADE software, developed by the SCIPORT team. This software is shown to be able to produce the adjoint code that can be used in data assimilation after a memory usage optimization.

7.3.4. Non-Parametric Estimation for Kinetic Diffusions

Participants: Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Venezuela and is partly funded by an ECOS Nord project.

We are focusing our attention on models derived from the linear Fokker-Planck equation. From a probabilistic viewpoint, these models have received particular attention in recent years, since they are a basic example for hypercoercivity. In fact, even though completely degenerated, these models are hypoelliptic and still verify some properties of coercivity, in a broad sense of the word. Such models often appear in the fields of mechanics, finance and even biology. For such models we believe it appropriate to build statistical non-parametric estimation tools. Initial results have been obtained for the estimation of invariant density, in conditions guaranteeing its existence and unicity [45] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [46] (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 [46], in collaboration with J.R. Leon (Caracas, Venezuela) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [47], 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. A paper has been submitted.

In [6], we focused on damping Hamiltonian systems under the so-called fluctuation-dissipation condition.

Note that Professor Jose R. Leon (Caracas, Venezuela) is now 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.

7.3.5. 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 [52] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [73] 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 [53] 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 [54], [56], we proposed non parametric estimators of a multivariate extension of the CTE. As might be expected, the properties of these estimators deteriorate when considering extreme risk levels. In collaboration with Elena Di Bernardino (CNAM, Paris), Clémentine Prieur is working on the extrapolation of the above results to extreme risk levels.

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [55]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervises the PhD thesis of Patricia Tencaliec. We are working on risk assessment, concerning flood data for the Durance drainage basin (France). The PhD thesis started in October 2013 and will be defended in next February. A first paper on data reconstruction has been accepted [79]. It was a necessary step as the initial series contained many missing data. A second paper is in preparation, considering the modeling of precipitation amount with semi-parametric sparse mixtures.

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.

MOISE was the leader of the ANR ADDISA project dedicated to the assimilation of images, and is a member of its follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France) that ended in 2013.

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [77], which was recently extended [76]. Thanks to the nonlinear brightness assumption, we proposed an algorithm to estimate the motion between two images, based on the minimization of a nonlinear cost function. We proved its efficiency and robustness on simulated and experimental geophysical flows. As part of the ANR project GeoFluids, we are investigating new ways to define distance between a couple of images. One idea is to compare the gradient of the images rather than the actual value of the pixels. This leads to promising results. Another idea, currently under investigation, consists in comparing main structures within each image. This can be done using, for example, a wavelet representation of images. Both approaches have been compared, in particular their relative merits in dealing with observation errors. This work has been extended to the progressive assimilation of different scales contained in the observations [22]

In recent developments we have also used "Level Sets" methods to describe the evolution of the images. The advantage of this approach is that it permits, thanks to the level sets function, to consider the images as a state variable of the problem. We have derived an Optimality System including the level sets of the images. This approach is being applied to the tracking of oceanic oil spills [41]

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. Figure 2 illustrates the advantage of using Wasserstein over L^2 : imagine that the density ρ_0 represents the observation, ρ_1 the background, and that we wish to find the "best" interpolation of the two. The "middle point" between them in the sense of the L^2 distance does not have the correct characteristics: its amplitude is smaller, and its shape is not correct as well. On the contrary, the W^2 middle point presents a similar structure and is indeed physically a better candidate for the interpolation of ρ_0 and ρ_1. Data assimilation experiments with the Shallow Water model have been realised and confirm the interest of the Wasserstein distance. Results have been presented at ISDA conference [35] and a paper has been submitted [34]. N. Feyeux will defend his PhD thesis on Dec. 8th.

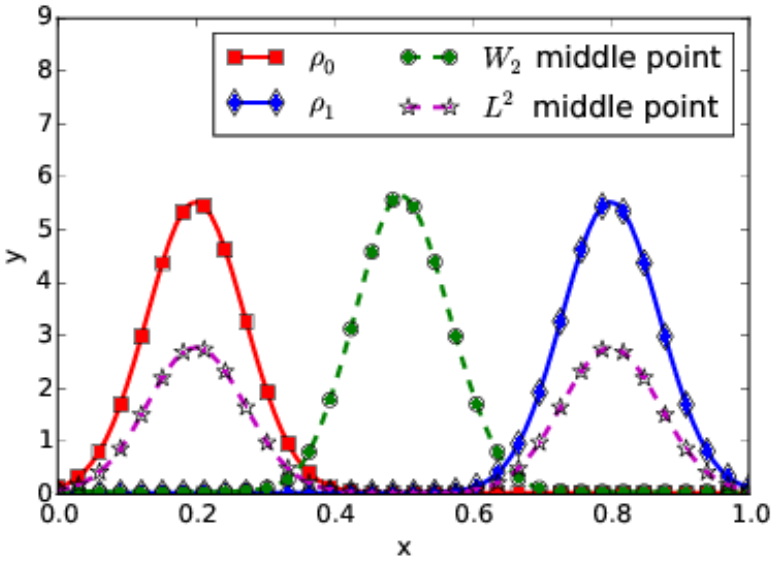


Figure 2. Illustration of the impact of using the Wasserstein distance in Data Assimilation instead of the classical L^2 distance.

7.5. Tracking of Mesoscale Convective Systems

Participant: Clémentine Prieur.

We are interested in the tracking of mesoscale convective systems.A particular region of interest is West Africa. Data and hydrological expertise is provided by T. Vischel and T. Lebel (LTHE, Grenoble).

A first approach involves adapting the multiple hypothesis tracking (MHT) model originally designed by the NCAR (National Centre for Atmospheric Research) for tracking storms [78] to the data for West Africa. With A. Makris (working on a post-doctoral position), we proposed a Bayesian approach [69], which consists in considering that the state at time t is composed on one hand by the events (birth, death, splitting, merging) and on the other hand by the targets' attributes (positions, velocities, sizes, ...). The model decomposes the state into two sub-states: the events and the targets positions/attributes. The events are updated first and are conditioned to the previous targets sub-state. Then given the new events the target substate is updated. A simulation study allowed to verify that this approach improves the frequentist approach by Storlie et al. (2009). It has been tested on simulations [69] and investigated in the specific context of real data on West Africa [12]. Using PHD (probability hypothesis density) filters adapted to our problem, generalizing recent developments in particle filtering for spatio-temporal branching processes (e.g. [44]) could be an interesting alternative to explore. The idea of a dynamic, stochastic tracking model should then provide the base for generating rainfall scenarios over a relatively vast area of West Africa in order to identify the main sources of variability in the monsoon phenomenon.

7.6. 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 [2] and Thomas Capelle will defend his in February 2017.

On top of the development on calibration procedure and sensitivity analysis for LUTI models, a study was conducted to understand in what extend modelling is or could be more integrated into urban planning [17].

ANGE Project-Team

7. New Results

7.1. Modelling of complex flows

7.1.1. The Shallow Water model with Roof: derivation and simulation

Participants: Edwige Godlewski, Cindy Guichard, Martin Parisot, Jacques Sainte-Marie, Fabien Wahl.

In view of taking into account interactions with floating structures, a shallow water type model is derived. In a first step a constraint corresponding to a static roof is considered and a relaxation approach is proposed in order to solve the model numerically. A particular attention is paid to the energy law as an application to marine energy devices is planned. The CPR scheme proposed in [17] is adapted to our case and implemented in one space dimension. Finally the numerical results are tested on analytical solutions, as well stationary as non-stationary ones [26].

7.1.2. Modelling of Sediment Transport

Participants: Emmanuel Audusse, Léa Boittin, Martin Parisot, Jacques Sainte-Marie.

A new model for sediment transport in river context is proposed. The model is derived from the Navier-Stokes equations by performing simultaneously the thin layer approximation and the diffusive limit. The well-posedness of the model is studied in a simplified case.

7.1.3. Layer-averaged Euler and Navier-Stokes equation

Participants: Marie-Odile Bristeau, Bernard Di Martino, Cindy Guichard, Jacques Sainte-Marie.

In [3], we propose a strategy to approximate incompressible hydrostatic free surface Euler and Navier-Stokes models. The proposed strategy extends previous works approximating the Euler and Navier-Stokes systems using a multilayer description. Here, the required closure relations are obtained using an energy-based optimality criterion instead of an asymptotic expansion. Moreover, the layer-averaged description is successfully applied to the Navier-Stokes system with a general form of the Cauchy stress tensor.

7.1.4. Layerwise Discretization for Non-Hydrostatic flows

Participants: Martin Parisot, Yohan Penel, Jacques Sainte-Marie.

In collaboration with Enrique Fernández-Nieto (Sevilla).

The work presented in [25] aims at deriving a new semi-discretisation with respect to the vertical variable of the Euler equations. It results in a hierarchy of multilayer model involving both hydrostatic and non-hydrostatic parts of the pressure field. All models are proven to satisfy an energy inequality. Moreover, the linear dispersion relation is given for each one with an explicit formula which converges to the exact Airy formula when the number of layers goes to infinity.

7.1.5. Two-phase (grains/fluid) model for geophysical debris flows

Participant: Anne Mangeney.

We developped a thin-layer depth-averaged model describing the two-phase flow made of granular material saturated by a fluid and include compression/dilatation effects. We solved numerically these equations and were able to accurately reproduce laboratory experiments.

7.1.6. Multi-layer model for viscoplastic granular flows

Participant: Anne Mangeney.

In collaboration with Enrique Fernández-Nieto and Gladys Narbona-Reina (Sevilla).

A multi-layer model was developped to simulate granular flow dynamics and deposit based on viscoplastic behaviour ($\mu(I)$-rheology). The numerical model made it possible to reproduce for the first time the increase of runout distance of granular material when flowing on erodible beds.

7.2. Assessments of models by means of experimental data

7.2.1. Hydrodynamics and biology coupling in the context of algae growth

Participants: Marie-Odile Bristeau, Jacques Sainte-Marie.

In collaboration with BIOCORE (especially O. Bernard) in the framework of the IPL Algae in Silico.

Hydrodynamics in a high rate production reactor for microalgae cultivation affects light history perceived by the cells. The interplay between cell movement and medium turbidity leads to a light pattern forcing photosynthesis dynamics. The purpose of this multidisciplinary downscaling study is to reconstruct single cell trajectories in an open raceway and experimentally reproduce such high frequency light pattern to observe its effect on growth. We show that the frequency of such a realistic signal plays a determinant role in the dynamics of photosynthesis. This study highlights the need for experiments with more realistic light stimuli in order to better understand microalgal growth at high cell density.

7.2.2. 2D Drucker-Prager and $\mu(I)$ granular flow model

Participant: Anne Mangeney.

In collaboration with François Bouchut, Ioan Ionescu, Alexandre Ern, Christelle Lusso and Nathan Martin.

We developped 2D (horizontal/vertical) models of granular flows solving the yield behaviour of Drucker-Prager type laws using either a duality method or a regularization method. We included the effect of the lateral wall friction and get very good agreement with laboratory experiments of granular collapses over horizontal and inclined planes.

7.2.3. Analytical and numerical description of the static/flowing interface deduced from 2D Drucker-Prager model

Participant: Anne Mangeney.

In collaboration with François Bouchut, Alexandre Ern and Christelle Lusso.

We proposed analytical and numerical solution of the static/flowing interface and compared it with laboratory experiments of granular flows. Our study show how the static/flowing interface dynamics depends on the slope, friction angle, viscosity and normal velocity profiles.

7.2.4. Seismic inversion and numerical modelling of the force generated by landslides on the topography or by iceberg calving

Participant: Anne Mangeney.

By inverting the long period seismic signal to recover the force generating seismic waves and simulating this force with mechanical models of granular flows, we can provide a unique constraint on the dynamics of the phenomenon at stake and on its characteristics.

7.2.5. Data assimilation

Participants: Sebastian Reyes-Riffo, Julien Salomon.

In collaboration with Felix Kwok.

Taking advantage of a PROCORE-FRANCE/HONG KONG grant obtained in the latter spring, we work on a time-parallelization strategy for an assimilation algorithm. The target application also deals with wave energy: we aim at forecasting in real-time the characteristics of the wave coming on an extracting device, in order to adapt it in a continuous way.

7.3. Analysis of models in Fluid Mechanics

7.3.1. Weak solutions of multilayer Hydrostatic Flows

Participants: Bernard Di Martino, Boris Haspot, Yohan Penel.

We investigate the existence of global weak solutions for the multilayer model introduced by Audusse et al. [2] which is related to incompressible free surface flows. More precisely, in [22] we prove the global stability of weak solutions over the torus. We observe that this model admits the so called BD-entropy and a gain of integrability on the velocity in the spirit of the work of Mellet and Vasseur. The main difficulty comes from the terms describing the transfer of flux between the layers which are not taken into account in the immiscible case.

7.3.2. Hyperbolicity of the Layerwise Discretized Hydrostatic Euler equation: the bilayer case

Participants: Emmanuel Audusse, Edwige Godlewski, Martin Parisot.

In collaboration with Nina Aguillon (UPMC).

Several model of free surface flows described in the literature are based on a layerwise discretization of the Euler equations. The question addressed in the current work is about the hyperbolicity of the layerwise discretized model. More precisely, we focus on the 2-layer case and we prove the well-posedness of the Riemann problem in two dimensional framework. Due to the mass exchange, the 2D Riemann problem is not a simple extension of the 1D Riemann problem.

7.3.3. Normal mode perturbation for the shallow water equations

Participants: Emmanuel Audusse, Albin Grataloup, Yohan Penel.

This work focuses on the shallow water equations for a fluid flow in subcritical regime with an arbitrary topography. A normal mode perturbation was performed around a 1D steady state in the 2D model. The resulting system of ODE was studied in terms of eigenvalues of the corresponding matrix and the derivation of a dispersion relation.

7.3.4. Global well-posdness of the Euler-Korteweg system for small irrotational data

Participant: Boris Haspot.

In collaboration with C. Audiard (UPMC).

The Euler-Korteweg equations are a modification of the Euler equations that takes into account capillary effects. In the general case they form a quasi-linear system that can be recast as a degenerate Schrödinger type equation. We prove here that under a natural stability condition on the pressure, global well-posedness holds in dimension $d \geq 3$ for small irrotational initial data. The proof is based on a modified energy estimate, standard dispersive properties if $d \geq 5$, and a careful study of the nonlinear structure of the quadratic terms in dimensions 3 and 4 involving the theory of space time resonance.

7.4. Numerical methods for free-surface flows

7.4.1. A two-dimensional method for a dispersive shallow water model

Participants: Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, Jacques Sainte-Marie.

We propose a numerical method for a two-dimensional dispersive shallow water system with topography. This model is a depth averaged Euler system and takes into account a non-hydrostatic pressure which implies to solve an incompressible system. From the variational formulation of the mixed problem proposed in [6], we apply a finite element method with compatible spaces to the two-dimensional problem on unstructured grids.

7.4.2. Numerical Discretization for Coriolis Effects

Participants: Emmanuel Audusse, Do Minh Hieu, Yohan Penel.

Efficient computations near the geostrophic equilibrium need to carefully design numerical schemes. This question is investigated in the context of colocated finite volume approach and extends previous works by Bouchut et al. [32], Dellacherie [35], Buet and Despres [36].

7.4.3. Optimization of topography

Participants: Sebastian Reyes-Riffo, Julien Salomon.

We work on a method to compute optimal topographies for wave-energy production. The first part of the work was devoted to the numerical analysis of the scheme used to simulate waves. In this way, we have obtained stability conditions that enable to couple it with an optimization loop.

7.4.4. An adaptive numerical scheme for solving incompressible two-phase and free-surface flows

Participant: Dena Kazerani.

We present a numerical scheme for solving two-phase or free surface flows. The interface/free surface is modelled using the level-set formulation. Besides, the mesh is anisotriopic and adapted at each iteration. The incompressible Navier–Stokes equations are temporally discretized using the method of characteristics and are solved at each time iteration by a first order Lagrange–Galerkin method. The level-set function representing the interface/free surface satisfies an advection equation which is also solved using the method of characteristics.

7.4.5. Propeler design

Participants: Jérémy Ledoux, Julien Salomon.

We work on a usual algorithm in propeler design: based on the so-called “Blade Element Momentum Theory”, this approach reduces the simulation to a 2D system by coupling the latter with a outer loop of low computational cost. So far, this method has not been analyzed mathematically, hence our interest.

7.5. Software developments and assessments

7.5.1. Improvements in the FRESHKISS3D code

Participants: Marie-Odile Bristeau, David Froger, Jacques Sainte-Marie, Fabien Souillé.

Several tasks have been achieved in the FRESHKISS3D software:

- Cython branch finalization (integration of second order in time and space numerical schemes)
- Project exportation on Gitlab.inria collaborative development platform
- New development tools set-up (Gitlab-ci, Git-lfs)
- Definition of new development rules and practices with gitlab
  - Use of the issue board
  - Review system and merge request rework
- Chlorides propagation in Vilaine river (Saur project)
  - Case definition in freshkiss3d
  - TracerSource class definition for floodgate modeling
  - VerticalDebit class definition for special boundary condition (siphon)
  - Simplified scenarios set-up (1day, 2days simulated)
  - First simulations and post processing
- New examples structure with introduction of two new cases to illustrate VerticalDebit and TracerSource class
- Various documentation updates

CASTOR Project-Team

5. New Results

5.1. Mathematical theory of reduced MHD models

Participant: Hervé Guillard.

In the modelling of strongly magnetized plasma, one of the fundamental model used 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 plasma 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 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.

5.2. Behavior of upwind finite volume scheme for Low Mach number flows

Participants: Hervé Guillard, Boniface Nkonga.

We have performed a review of different modifications proposed to enable compressible flow solvers to compute accurately flows near the incompressible limit. The reasons of the failure of upwind solvers to obtain accurate solutions in the low Mach number regime have been explained and different corrections proposed in the literature have been reviewed and discussed. Numerical experiments to illustrate the behavior of the different methods have been done and presented. This work will be published in 2017 as a contribution for the "Handbook of numerical analysis" collection.

5.3. Finite volume approximations for fusion plasma

Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

The MHD model used for plasma studies in tokamak is very often based on the magnetic vector potential form of the equations where the vector potential satisfies $\nabla \times \mathbf{A} = \mathbf{B}$ with $\mathbf{B}$ the magnetic field and only a small number of numerical models uses the conservative formulation based on $\mathbf{B}$. One of the shortcomings of this latter formulation is the necessity to enforce numerically the divergence free constraint on the magnetic field that can be difficult to achieve and/or computationally costly. Another difficulty is that the equilibrium solution of the MHD equation given by the Grad-Shafranov equation is not an exact solution of the discrete equation.
We have begun to investigate the use of the $\mathbf{B}$ formulation for tokamak studies. The divergence free constraint is taken into account by a projection at each time step on a rotated gradient field. This step ensures a strict respect of the divergence free constraint while being extremely cheap since the scalar field is simply advected by the flow. The numerical experiments performed show that this method is efficient for the study of discontinuous MHD flows. For plasma fusion flows, the method experiences presently some difficulties to compute steady equilibrium flows.

5.4. Bi-temperature Euler equations

Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

A particular class of extended MHD models uses a description of the plasma where the ionic and electronic temperatures are different while velocities and densities are common to the two species. This preliminary work has examined the construction of finite volume numerical schemes for two-temperature models in the context of the Euler equations. The finite volume scheme uses the assumption that the electronic entropy is constant across the shocks to define the weak solutions of the system and the numerical fluxes are obtained with a relaxation scheme. Numerical simulations of several test-cases involving strong shocks show that this numerical strategy is efficient even in the presence of strong temperature differences between ions and electrons.

5.5. Domain segmentation using the Reeb Graph

Participants: Hervé Guillard, Adrien Loseille (gamma3 Inria-Saclay), Alexis Loyer.

The generation of block structured meshes is a difficult task that is not easily automated and very often ask for manual intervention and specific expertise. We show in this work that if the required mesh is constrained to be aligned on the contour lines of a Morse function, then the mesh generation process can be done in a fully automatic way and reduces to only two basic meshing operations. This technique can be useful for a large number of potential applications. It is here studied for the construction of flux surface aligned meshes in the framework of the EoCoE project.

5.6. Equilibrium reconstruction

Participants: Blaise Faugeras, Jacques Blum, Cédric Boulbe, Holger Heumann.

Within the framework of the European Integrated Tokamak Modelling WPCD project we have been involved in a benchmark study between the equilibrium reconstruction codes EQUINOX, EQUAL and CLISTE on AUG (Asdex Upgrade) equilibriums. This work has been presented at the 2016 EPS conference.

The benchmark study leads us to include new functionalities to EQUINOX such as the possibility to use a radially variable regularization and the computation of error bars on the reconstructed profiles.
In order to be used on the WEST tokamak, EQUINOX has been adapted to the ITER standard "IMAS" using IDS as data type.
A numerical method for equilibrium reconstruction using magnetic measurements as well as polarimetry measurements with their Stokes vector representation has been developped in order to take into account the so-called Cotton-Mouton effect. The algorithm is based on optimal control of a coupled partial and ordinary differential equations system. The method is being tested on an ITER test case.

5.7. FEB-BEM numerical methods for equilibrium computation

Participants: Blaise Faugeras, Holger Heumann.

A code which treats the quasi-static free-boundary equilibrium problem needs to solve nonlinear elliptic or parabolic problems with nonlinear source terms representing the current density profile vanishing outside the unknown free boundary of the plasma. The computational challenges in the design of such a code are: a problem setting in an unbounded domain with a nonlinearity due to the current density profile in the unknown plasma domain and the nonlinear magnetic permeability if the machine has ferromagnetic structures. In this project we focused on how the simulation on the unbounded domain can be reduced to computations on an interior bounded domain thanks to analytical Green's functions. The numerical solution on the interior domain is coupled through boundary conditions to the Green's function representation of the solution in the unbounded exterior domain. This approach is today fairly standard in many other application areas such as electromagnetics or elasticity and falls into the framework of the boundary element method (BEM). Most authors in the fusion literature deal with this question using the same method from von Hagenow and Lackner whereas the coupling can be conceived in different ways. In this project we implemented 3 different schemes in order to assess their performance. One of them, the classical Johnson-Nédélec FEM-BEM coupling (JNC) has never been tested before in a fusion equilibrium code.

5.8. A finite element method with overlapping meshes for free-boundary toroidal plasma equilibria in realistic geometry

Participants: Francesca Rapetti, Holger Heumann.

Existing finite element implementations for the computation of free-boundary axisymmetric plasma equilibria approximate the unknown poloidal flux function by standard lowest order continuous finite elements with discontinuous gradients. The location of critical points of the poloidal flux, that are of paramount importance in tokamak engineering, is constrained to nodes of the mesh, which leads to undesired jumps in transient problems. Moreover, recent numerical results for the self-consistent coupling of equilibrium with resistive diffusion and transport suggest the necessity of higher regularity when approximating the flux map.

In [23], we have proposed a mortar element method that employs two overlapping meshes. One mesh with Cartesian quadrilaterals covers the vacuum domain and one mesh with triangles discretizes the region outside the vacuum domain. The two meshes overlap in a narrow region around the vacuum domain. This approach gives the flexibility to achieve easily and at low cost higher order regularity for the approximation of the flux function in the domain covered by the plasma, while preserving accurate meshing of the geometric details exterior to the vacuum. The continuity of the numerical solution in the region of overlap is weakly enforced by a mortar-like projection. We have shown 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 [19], we have rather analysed the precision of the proposed approach, by varying the discretization parameters. We thus compute the approximation error between the computed and the synthetic solution of a model problem for the same method adopted in [1], by changing, for example, the local polynomial degree in the subdomains, the size of the overlap between the meshes, the local size of the mesh elements. Indeed, FE methods on composite meshes are widely used in practice, but their theoretical foundation is fairly limited in the literature. Therefore, we have reported in [2] some experimental convergence results for different discretization schemes involving composite meshes.

5.9. Circuit Equations

Participant: Holger Heumann.

We derived a new formulation to combine the circuit equations due to the poloidal field coil system with free boundary equilibrium calculations. The previous formulations based on a least squares formulation developed for and implemented in CEDRES++, was suffering from numerical instabilities. The new formulation was implemented in FEEQS.M and successfully validated.

5.10. Optimization of tokamak breakdown scenarios

Participants: Holger Heumann, Eric Nardon.

The standard method to initiate a plasma in a tokamak is to realize a so called Townsend avalanche by applying a high enough toroidal electric field (i.e. loop voltage) by means of a fast variation of the current in the poloidal field coils (in particular the central solenoid). For the avalanche to take place, the electrons need to be able to travel along the field lines over a long enough distance, so that they can gain an energy significantly larger than the ionization energy of the atoms. An empirical criterion for a successful breakdown is thus $EL_c > 70V$, where E is the toroidal electric field and L_c the connection length of the field lines. Hence, it is highly desirable to create a configuration in which the field length is as large as possible, or equivalently, in which the poloidal component of the field is as small as possible. We reformulated this task as a constrained optimization problem and used an implementation in FEEQS.M to find in an automated fashion such optimal configurations. Publication is in preparation.

5.11. High order C^0-continuous Galerkin schemes for high order PDEs

Participants: Sebastian Minjeaud, Richard Pasquetti.

We 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 linear or quadratic invariants, for high order partial differential equation on the basis of a (only) H^1-conformal Galerkin approximation, namely the Spectral Element Method. We address the Korteweg-de Vries equation but the proposed approach is *a priori* easily extensible to other partial differential equations and to multidimensional problems.

5.12. A MUSCL–scheme on staggered grids with kinetic–like fluxes for the barotropic Euler system

Participants: Julia Llobell, Thierry Goudon, Sebastian Minjeaud.

We set up a MUSCL version of the scheme introduced in [27] for solving the barotropic Euler equations. The scheme works on staggered grids, with numerical densities and velocities stored at dual locations, while the numerical fluxes are derived in the spirit of kinetic schemes. We have identified stability conditions for the second order method and have shown the ability of the scheme to capture the structure of complex flows with 2D simulations on MAC grids.

5.13. Stabilized SEM approximation of the 2D Saint-Venant system

Participant: Richard Pasquetti.

Following a study restricted to one space dimension, R. Pasquetti has developped a stabilized Spectral Element approximation of the two-dimensional Saint-Venant system. This stabilized SEM model uses the entropy viscosity method (EVM), that is a non linear viscous stabilization with viscosity proportional to the entropy production and bounded from above by a first order viscosity. We have especially focused on problems involving dry-wet transitions and proposed an elaborated variant of the EVM that allows to support the presence of dry zones. The algorithm has been tested against benchmarks problems, involving planar oscillations and axisymmetric oscillations in a paraboloid, for which exact solutions are known. The method was also checked successfully for flows combining dry-wet transitions and shocks. Part of this study was carried out in the National Center for Theoretical Science (Taipei, Taiwan). The work was presented at the ICOSAHOM 2016 congress (Rio, June 2016, see [16]).

5.14. Isoparametric mappings

Participant: Richard Pasquetti.

R. Pasquetti has carried out a numerical study to compare different isoparametric mappings for the approximation of non polygonal domains with high order triangular finite elements. For elliptic problems and Fekete-Gauss spectral elements, it turns out that isoparametric mappings based on PDEs (Laplace, linear elasticity) yield better results than those based on transfinite mappings. The results are summarized in a JCP Note (see [15]).

5.15. Full MHD numerical modelling with $C1$ finite element.

Participants: José Costa, Boniface Nkonga.

Many theoretical and numerical works in the field of tokamak modelling use specific approximations of the MHD model known as *reduced* MHD models. This is in particular the case of the Jorek software. The main objective of this work is therefore to extend the capability of this software to solve the full MHD equations while using the same finite element numerical method. This requires to design new stabilization strategies as well as appropriate projections of the momentum equation. This has been done during the thesis of José Costa [6] This work allowed a detailed study of the resistive internal kink instability as well as some preliminary results on X-point plasmas.

5.16. 2D Triangular Powell-Sabin Finite Elements

Participants: Giorgio Giorgiani, Hervé Guillard, Boniface Nkonga.

In order to avoid some mesh singularities when using quadrangular meshes for complex geometries and flux surfaces shapes, the use of triangular elements is a possible option that we are studying in view of its application to MHD modelling. It is not so easy to derive smooth finite element on triangle with reduced number of degree of freedom (ddl). The Bell reduced-quintic finite elements we have considered in the previous years have too much unknowns (6 per vertex). Powell-Sabin splines are piece-wise quadratic polynomials with a global C1-continuity and 3 unknowns per vertex, they have a local support, they form a convex partition of unity, they are stable, and they have a geometrically intuitive interpretation involving control triangles. In the previous years, we have developed the geometrical tools necessary to the construction of the Powell-Sabin splines and we are now beginning the study of the applicability of Powell-Sabin finite element for the numerical solution of PDE. We have used the Powell-Sabin starting from elliptic partial differential equations (including Grad-shafranov). We have applied these tools to solve hyperbolic 2D Euler equations with VMS stabilization. These results have been published in [11] and [18]

5.17. Massive gaz Injection

Participant: Boniface Nkonga.

The massive injection of impurity gas into a plasma has been proved to reduce forces and localized thermal loads caused by disruptions in tokamaks. This mitigation system is routinely used on JET to shut down plasmas with a locked mode. The DMV's injectors of JET have been modelled with all the 3D details (see Figure 1 , 2 and 3) . We have performed many 3D simulations and the predicted flight times are in accordance with experimental measurements. Moreover, the computations give also a clear domain for the application of 1D approximations and scaling.

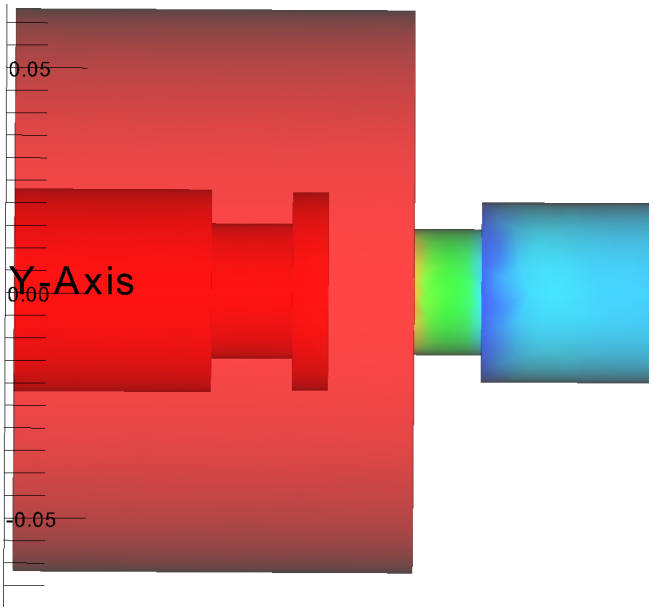


Figure 1. DMV resevoir of the JET Tokamak

5.18. A Multidimensional Analogue of the HLLI Riemann Solver for Conservative Hyperbolic Systems

Participants: Boniface Nkonga, Dinshaw Balsara.

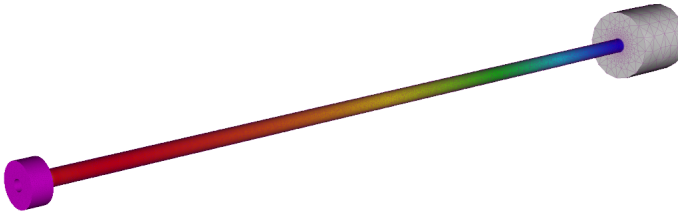


Figure 2. Reservoir, tube and plasma front of the Injection system of JET.

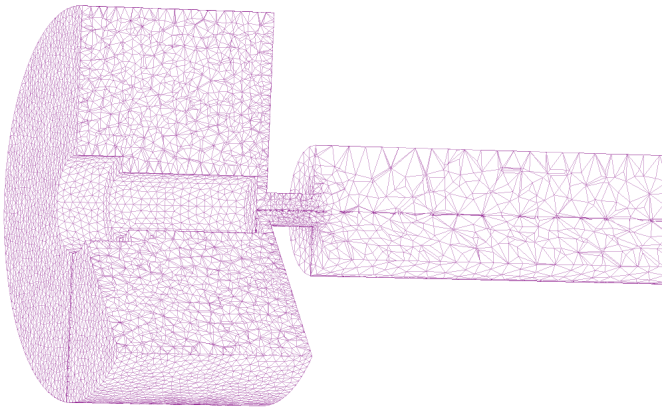


Figure 3. 3D Mesh of tetrahedral elements.

Just as the quality of a one-dimensional approximate Riemann solver is improved by the inclusion of internal sub-structure, the quality of a multidimensional Riemann solver is also similarly improved. Such multidimensional Riemann problems arise when multiple states come together at the vertex of a mesh. The interaction of the resulting one-dimensional Riemann problems gives rise to a strongly-interacting state. We wish to endow this strongly-interacting state with physically-motivated sub-structure. The fastest way of endowing such sub-structure consists of making a multidimensional extension of the HLLI Riemann solver for hyperbolic conservation laws. Presenting such a multidimensional analogue of the HLLI Riemann solver with linear sub-structure for use on structured meshes is the goal of this work. The multidimensional MuSIC Riemann solver documented here is universal in the sense that it can be applied to any hyperbolic conservation law.
The multidimensional Riemann solver is made to be consistent with constraints that emerge naturally from the Galerkin projection of the self-similar states within the wave model. When the full eigenstructure in both directions is used in the present Riemann solver, it becomes a complete Riemann solver in a multidimensional sense. I.e., all the intermediate waves are represented in the multidimensional wave model. The work also presents, for the very first time, an important analysis of the dissipation characteristics of multidimensional Riemann solvers. The present Riemann solver results in the most efficient implementation of a multidimensional Riemann solver with sub-structure. Because it preserves stationary linearly degenerate waves, it might also help with well-balancing. Implementation-related details are presented in pointwise fashion for the one-dimensional HLLI Riemann solver as well as the multidimensional MuSIC Riemann solver.
Several stringent test problems drawn from hydrodynamics, MHD and relativistic MHD are presented to show that the method works very well on structured meshes. Our results demonstrate the versatility of our method.

5.19. Modelling of plasma instabilities

Participants: Feng Liu, Boniface Nkonga, Guido Huijsmans, Alberto Loarte.

Non-linear simulations of MHD modes from 0 to 20 which include kink-peeling modes (KPM) and ballooning modes with different plasma equilibrium by varying both pedestal pressure and edge current have been studied further. The simulations indicated that sufficient high edge current is essential requirements for plasma saturate to edge harmonic oscillation (EHO), meanwhile the pedestal pressure is the key parameter for plasma saturating to ballooning mode. The influence of RMP (Resonant Magnetic Perturbations) on QH-mode (Quiescent High Confinement mode) has been re-evaluated by using the correct coil currents. Large number ergodic islands caused by RMP stabilize toroidal harmonics n=1, 2, 3, 4 modes in the edge of QH-mode plasma. ITER baseline scenario Q=10 plasma has been analyzed with respect to the access to a possible QH-mode regime. KPM is obtained at the edge plasma of ITER plasma for n=1 and n=1-5 modes (see Fig. 4).

5.20. Amoss : Comparison with experimental results and unreduced model on flat plane

Participants: B. Nkonga, H. Guillard, S. Gavrilyuck, Y-C. Tai, F. Yang, K.m. Shyue, C-Y Kuo.

The purpose of this work was the numerical study of the roll-waves that develop from a uniform unstable flow down an inclined rectangular channel. In particular, the formation of the roll-waves is studied by two different approaches. In the first approach, the roll-waves were produced in a long channel where a wave maker perturbed the free surface only at the channel inlet. The average discharge was fixed. In the second approach, the roll-waves were produced in a “periodic box” with a uniform flow velocity. The average depth of a perturbed free surface was the same as in the long channel. Formally, the “periodic box” and a long channel correspond to two different physical situations. However, the stationary profile formed for long time in these cases is the same. This allows us to use the “periodic box” as a simpler mathematical tool to study the asymptotic behavior of roll waves. In particular, the “periodic box” does not require a big space domain resolution. Several interesting phenomena were observed. First, it was proven that there exists Lmax such

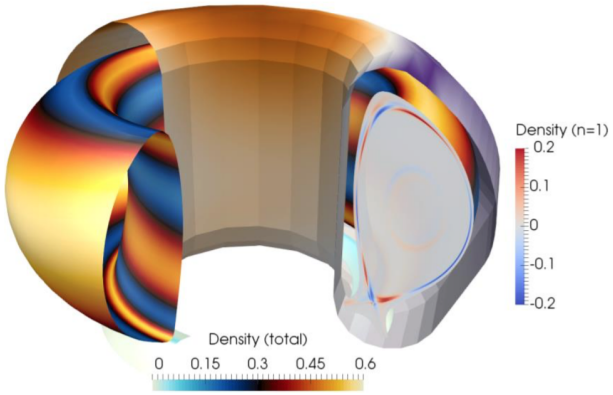


Figure 4. 3-D density structure at the separatrix and resistive wall potential of a n=1 saturated kink mode in ITER Q=10 scenario.

that any single roll wave of length $L > L_{\max}$ not stable. This can help to generalize the analytical results obtained by Liapidevskii (modulational stability study) and Baker et al. (the linear stability study) for the SV equations, to the case of the generalized models. The minimal length of periodic box for which a single roll wave is stable, was not observed. Second, a coarsening phenomenon was observed. When the inlet perturbation has two different frequencies, it produces the waves of the different wavelengths. The waves begin to interact. The short waves transfer their energy to the long waves, and finally we obtain the train of roll waves of a larger wavelength. A strong non-stationary modulation of the wave amplitude was observed. The formation of periodic roll wave train was shown for both long a channel and a "periodic box" for two sets of experimental parameters. In both cases, the free surface profile for the generalized models was found in a very good agreement with the experimental results. Finally, for a 2D simplified "Toy Model" we show that steady numerical solution corresponding to experimental data does not depend of transverse perturbations.

CLIME Project-Team

7. New Results

7.1. State estimation: analysis and forecast

One major objective of Clime is the conception of new methods of data assimilation in geophysical sciences. Clime is active on several challenging aspects: non-Gaussian assumptions, multiscale assimilation, minimax filtering, etc.

7.1.1. Assimilation of drifter data in the East Mediterranean Sea

Participants: Julien Brajard, Isabelle Herlin, Leila Issa [Lebanese American University, Lebanon], Laurent Mortier [LOCEAN], Daniel Hayes [Oceanography Centre, Cyprus], Milad Fakhri [CNRS, Lebanon], Pierre-Marie Poulain [Oceanography Institute of Trieste, Italy].

Surface velocity fields of the ocean in the Eastern Levantine Mediterranean are estimated by blending altimetry and surface drifters data. The method is based on a variational assimilation approach for which the velocity is corrected by matching real drifters positions with those predicted by a simple advection model, while taking into account the wind effect. The velocity correction is done in a time-continuous fashion by assimilating at once a whole trajectory of drifters with a temporal sliding window. Except for the wind component, a divergence-free regularization term was added to constrain the velocity field. Results show that, with few drifters, our method improves the estimated velocity in two typical situations: an eddy between the Lebanese coast and Cyprus, and velocities along the Lebanese coast. A description of these results is published in the Ocean Modelling journal.

7.1.2. State estimation for noise pollution

Participants: Raphaël Ventura, Vivien Mallet, Valérie Issarny [Mimove], Pierre-Guillaume Raverdy [SED], Fadwa Rebhi [Mimove], Cong Kinh Nguyen [Mimove].

70 million observations of ambient noise have been collected with the mobile application Ambiciti (previously, SoundCity). An important work was carried out on the calibration of the measurements. Over 100 mobile phones were calibrated against a sound level meter, at various noise intensities and frequencies, in order to test their response and devise a calibration strategy.

A data assimilation procedure has been put in place in order to select and assimilate the most reliable observations. Simulated noise maps have been improved with the observations, by computing the so-called best linear unbiased estimator (BLUE) with error covariance models suitable for noise pollution. The assimilation of mobile observation introduces new errors, like location errors, compared to the assimilation of the more common observations from fixed monitoring stations.

7.2. Image assimilation

Sequences of images, such as satellite acquisitions, display structures evolving in time. This information is recognized of major interest by forecasters (meteorologists, oceanographers, etc.) in order to improve the information provided by numerical models. However, the satellite images are mostly assimilated in geophysical models on a point-wise basis, discarding the space-time coherence visualized by the evolution of structures such as clouds. Assimilating image data in an optimal way is of major interest. This issue is twofold:

- from the model's viewpoint, the location of structures on the observations is used to control the state vector.
- from the image's viewpoint, a model of the dynamics and structures is built from the observations.

7.2.1. Estimation of motion and acceleration from image data

Participants: Dominique Béréziat [UPMC], Isabelle Herlin.

Image sequences allow visualizing dynamic systems and understanding their intrinsic characteristics. One first component of this dynamics is obtained by retrieving the velocity of the structures displayed on the sequence. This motion estimation issue has been extensively studied in the literature of image processing and computer vision. In this research, we step beyond the traditional optical flow methods and address the problem of recovering the acceleration from the whole temporal sequence, which has been poorly investigated, even if this is of major importance for some data types, such as fluid flow images. Acceleration is here viewed as the space-time function resulting from the forces applied to the studied system. To solve this issue, we propose a variational approach where a specific energy is designed to model both the motion and the acceleration fields. The contributions are twofold: first, we introduce a unified variational formulation of motion and acceleration under space-time constraints; second, we define the minimization scheme, which allows retrieving the estimations, and provide the full information on the discretization schemes. Experiments are conducted on synthetic and real image sequences, visualizing fluid-like flows, where direct and precise calculation of acceleration is of primary importance.

7.2.2. Rain nowcasting from radar image acquisitions

Participants: Isabelle Herlin, Étienne Huot.

This research concerns the design of an operational method for rainfall nowcasting that aims at prevention of flash floods. The nowcasting method includes two main components:

- a data assimilation method, based on radar images, estimates the state of the atmosphere: this is the estimation phase.
- a forecast method uses this estimation to extrapolate the state of the atmosphere in the future: this is the forecast phase.

Results are analyzed on space-time neighborhoods in order to prevent consequences of flash floods on previously defined zone.

Current research concerns the following issues:

- the use of object components in the state vector. The objective is to improve the description of the image data in order to get a better motion estimation and a more accurate localization of endangered regions.
- the extension of the estimation phase to a multiscale process.
- the merging with measures acquired by a network of pluviometers.

7.2.3. Ensemble Kalman filter based on the image structures

Participants: Dominique Béréziat [UPMC], Isabelle Herlin.

One major limitation of the motion estimation methods that are available in the literature concerns the availability of the uncertainty on the result. This is however assessed by a number of filtering methods, such as the ensemble Kalman filter (EnKF). Our research consequently concerns the use of a description of the displayed structures in an ensemble Kalman filter, which is applied for estimating motion on image acquisitions. Compared to the Kalman filter, EnKF does not require propagating in time the error covariance matrix associated to the estimation, resulting in reduced computational requirements. However, EnKF is also known for exhibiting a shrinking effect when taking into account the observations on the studied system at the analysis step. Various methods are available in the literature for correcting this effect, but they do not involve the structures displayed on the image sequence. We defined two alternative solutions to that shrinking effect: a dedicated localization function and an adaptive decomposition domain. These methods are both well suited for fluid flows images and applied on satellite images of the atmosphere.

7.3. Uncertainty quantification and risk assessment

The uncertainty quantification of environmental models raises a number of problems due to:

- the dimension of the inputs, which can easily be 10^5-10^8 at every time step;
- the dimension of the state vector, which is usually 10^5-10^7;
- the high computational cost required when integrating the model in time.

While uncertainty quantification is a very active field in general, its implementation and development for geosciences requires specific approaches that are investigated by Clime. The project-team tries to determine the best strategies for the generation of ensembles of simulations. In particular, this requires addressing the generation of large multimodel ensembles and the issue of dimension reduction and cost reduction. The dimension reduction consists in projecting the inputs and the state vector to low-dimensional subspaces. The cost reduction is carried out by emulation, i.e., the replacement of costly components with fast surrogates.

7.3.1. Sequential aggregation with uncertainty estimation

Participants: Jean Thorey, Vivien Mallet, Christophe Chaussin [EdF R&D].

In the context of ensemble forecasting, one goal is to combine an ensemble of forecasts in order to produce an improved probabilistic forecast. We previously designed a new approach to predict a probability density function or cumulative distribution function, from a weighted ensemble of forecasts. The procedure aims at forecasting the cumulative distribution function of the observation which is simply a Heaviside function centered at the observed value. Our forecast is the weighted empirical cumulative distribution function based on the ensemble of forecasts. Each forecast of the ensemble is attributed a weight which is updated whenever new observations become available. The performance of the forecast is given by the continuous ranked probability score (CRPS), which is the square of the two-norm of the discrepancy between the forecast and the observed cumulative distribution functions. The method guarantees that, in the long run, the forecast cumulative distribution function has a continuous ranked probability score at least as good as the best weighted empirical cumulative function with weights constant in time.

The CRPS computed from an ensemble of forecasts is subject to a bias. We proposed a new way to compute the CRPS in order to mitigate the bias and obtain better aggregation performance.

The work was applied to the forecast of photovoltaics production, both at EDF production sites and for global France production.

7.3.2. Sensitivity analysis of air quality simulations at urban scale

Participants: Vivien Mallet, Louis Philippe, Fabien Brocheton [Numtech], David Poulet [Numtech].

We carried out a sensitivity analysis of the urban air quality model Sirane. We carried out dimension reduction on both inputs and outputs of the air quality model. This designed a reduced-order model, which we then emulated. We sampled the (reduced) inputs to the reduced model, and emulated the response surface of the reduced outputs. A metamodel was derived by the combination of the dimension reduction and the statistical emulation. This metamodel performs as well as the original model, compared to field observations. It is also extremely fast, which allowed us to compute Sobol' indices and carry out a complete sensitivity analysis.

7.3.3. Sensitivity analysis of road traffic simulations and corresponding emissions

Participants: Ruiwei Chen [École des Ponts ParisTech], Vivien Mallet, Vincent Aguiléra [Cerema], Fabien Brocheton [Numtech], David Poulet [Numtech], Florian Cohn [Numtech].

This work deals with the simulation of road traffic at metropolitan scale. We compared state-of-the-art static traffic assignment and dynamic traffic assignment, which better represents congestion. The work was applied in Clermont-Ferrand and its surrounding region, for a time period of two years, and using about 400 traffic loop counters for evaluation. The dynamic model showed similar overall performance as the static model.

We developed an open source software for the computation of the emissions of traffic. It computes the emissions of the main air pollutants, according the vehicle fleet.

For both traffic assignment and pollutant emissions, we carry out sensitivity tests with respect to limit speed, roads capacities or fleet composition. A complete sensitivity analysis is out of reach with the complete, computational intensive, traffic assignment model. Hence further work has been engaged with the metamodeling of the traffic assignment model. Preliminary results are encouraging and tend to show that a very fast metamodel can perform as well as the complete model.

7.3.4. Ensemble variational data assimilation

Participants: Julien Brajard, Isabelle Herlin, Marc Bocquet [CEREA], Jérôme Sirven [LOCEAN], Olivier Talagrand [LMD, ENS], Sylvie Thiria [LOCEAN].

The general objective of ensemble data assimilation is to produce an ensemble of analysis from observations and a numerical model which is representative of the uncertainty of the system. In a bayesian framework, the ensemble represents a sampling of the state vector probability distribution conditioned to the available knowledge of the system, denoted the a-posteriori probability distribution.

Ensemble variational data assimilation (EnsVar) consists in producing such an ensemble by perturbating N times the observations according to their error law, and run a standard variationnal assimilation for each perturbation. An ensemble of N members is then produced. In the case of linear models, there is a theoretical guarantee that this ensemble is a sampling of the a-posteriori probability. But there is no theoretical result in the non-linear case.

The objective of this work is to study the ability of EnsVar to produce "good" ensemble (i.e. that sampled the a posteriori probablility) on a shallow-water model. Statistical properties of the ensemble are evaluated, and the sensitivity to the main features of the assimilation system (number, distribution of observations, size of the assimilation window, ...) are also studied.

COFFEE Project-Team

6. New Results

6.1. A few words on the results of the year

- Analysis of wave propagation in mechanics, partly in collaboration with phsyicsts [40], [24]
- Analysis of PDE system in chromatography [5] and in traffic flows modelling [30]
- Analysis of conservation laws, with many application like traffic flows, fluid mechanics, etc [29], [36], [37], [11], [18]
- Modeling of attractive dynamics between individuals, pattern formation, with the derivation, the analysis and simulations of hierarchies of mathematical models, from microscopic to macroscopic, [9], [17]
- Derivation and simulation of hydrodynamic models in biology (biofilms growth, intestinal gut), partly in collaboration with INRA, [4], [7], [16], [41], [2]
- Modeling and simulation of compositional multiphase flows in porous media, with many indisutrial collaborations with ANDRA< BRGM, EdF... [22], [32], [23], [33], [21], [34], [39], [42], [6], [20]
- Analysis of Finite Volume schemes in fluid mechanics [35], [15], [12], [38]
- Domain decomposition methods [43], [31]
- Many particles systems, effect of stochasticity [27], [1], [28], [8], [13], [10], [19], [3]

FLUMINANCE Project-Team

6. New Results

6.1. Fluid motion estimation

6.1.1. 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 has the great advantage to incorporate from the beginning an uncertainty on the 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. This estimator, which extend a local motion estimator previously proposed in the team, has shown to improve significantly the results of the corresponding deterministic estimator. This kind method is assessed in the context of river hydrologics applications through a collaboration with an Irstea Lyon research group (HHLY). This study is performed within the PhD thesis of Musaab Mohammed.

6.1.2. 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 the development of precise models relating the large-scale flow observations to the velocity, appropriate large-scale regularization strategies, and 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.

6.1.3. 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 keep on our investigation on the problem of efficient volume reconstruction. Our work takes place within the context of some modern optimization techniques. First, we focussed our attention on the family of proximal and splitting methods and showed that the standard techniques commonly adopted in the TomoPIV literature can be seen as particular cases of such methodologies. Recasting standard methodologies in a more general framework allowed us to propose extensions of the latter: i) we showed that the parcimony characterizing the sought volume can be accounted for without increasing the complexity of the algorithms (e.g., by including simple thresholding operations); ii) we emphasized that the speed of convergence of the standard reconstruction algorithms can be improved by using Nesterov's acceleration schemes; iii) we also proposed a totally novel way of reconstructing the volume by using the so-called "alternating direction of multipliers method" (ADMM). In 2016, this work has led to the publication of a contribution in the international journal IOP Measurement Science and Technology.

On top of this work, we also focussed on another crucial step of the volume reconstruction problem, namely the pruning of the model. The pruning task consists in identifying some positions in the volume of interest which cannot contains any particle. Removing this position from the problem can then potentially allow for a dramatic dimensionality reduction. This year, we provide a methodological answer to this problem through the prism of the so-called "screening" techniques which have been proposed in the community of machine learning. In 2016, this work led to the publication of one contribution in the proceedings of the international conference on acoustics, speech and signal processing (ICASSP'16).

6.1.4. Sparse-representation algorithms

Participant: Cédric Herzet.

The paradigm of sparse representations is a rather new concept which turns out to be central in many domains of signal processing. In particular, in the field of fluid motion estimation, sparse representation appears to be potentially useful at several levels: i) it provides a relevant model for the characterization of the velocity field in some scenarios; ii) it plays a crucial role in the recovery of volumes of particles in the 3D Tomo-PIV problem.

Unfortunately, the standard sparse representation problem is known to be NP hard. Therefore, heuristic procedures have to be devised to access to the solution of this problem. Among the popular methods available in the literature, one can mention orthogonal matching pursuit (OMP), orthogonal least squares (OLS) and the family of procedures based on the minimization of sparsity inducing norms. In order to assess and improve the performance of these algorithms, theoretical works have been undertaken in order to understand under which conditions these procedures can succeed in recovering the "true" sparse vector.

This year, we contributed to this research axis by deriving conditions of success for the algorithms mentioned above when the amplitudes of the nonzero coefficients in the sparse vector obey some decay. In a TomoPIV context, this decay corresponds to the fact that not all the particles in the fluid diffuse the same quantity of light (notably because of illumination or radius variation). In particular, we show that the standard coherence-based guarantees for OMP/OLS can be relaxed by an amount which depends on the decay of the nonzero coefficients. In 2016, our work has led to the publication of one paper in the journal IEEE Transactions on Information Theory.

We also investigated a new methodology to take sparsity into account into variational assimilation problems. We focussed on the problem of estimating of scalar transported by an unknown velocity field, when only low-resolution observations of the scalar are supposed to be available. The goal is to reconstruct both a high-resolution version of the scalar and the velocity field, assuming that these quantities admit a sparse decomposition in some proper frames. The associated optimization problem typically involves millions of variables and thus requires dedicated optimization procedures to be tractable. In 2016, we proposed a new assimilation scheme combining state-of-the-art optimization techniques (forward-backward propagation, ADMM, Attouch's procedure) to address this problem. Our algorithm is provably convergent while exhibiting a complexity per iteration evolving linearly with the problem's dimensions. This contribution has led to a journal publication in SIAM Journal on Imaging Science.

6.2. Tracking, Data assimilation and model-data coupling

6.2.1. Stochastic fluid flow dynamics under uncertainty

Participants: Pierre Derian, Etienne Mémin, Valentin Resseguier.

In this research axis we aim at devising Eulerian expressions for the description of fluid flow evolution laws under uncertainties. Such an uncertainty is modeled through the introduction of a random term that allows taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also uncertainties attached to non-uniform grid discretization. This model is mainly based on a stochastic version of the Reynolds transport theorem. Within this framework various simple expressions of the drift component can be exhibited for

different models of the random field carrying the uncertainties we have on the flow. We aim at using such a formalization within image-based data assimilation framework and to derive appropriate stochastic versions of geophysical flow dynamical modeling. This formalization has been published in the journal Geophysical and Astrophysical Fluid Dynamics [10]. Numerical simulation on divergence free wavelets basis of 3D viscous Taylor-Green vortex and Crow instability have been performed within a collaboration with Souleymane Kadri-Harouna. Besides, we explore in the context of Valentin Resseguier's PhD the extension of such framework to oceanic models and to satellite image data assimilation. This PhD thesis takes place within a fruitful collaboration with Bertrand Chapron (CERSAT/IFREMER). This year we have more deeply explored several uncertainty representations of classical geophysical models for ocean and atmosphere. This study have led to very promising stochastic representation for the Quasi Geostophic approximation (QG) with noises of different energy.

6.2.2. Free surface flows reconstruction and tracking

Participants: Dominique Heitz, Etienne Mémin.

We investigated the combined use of a Kinect depth sensor and of a stochastic data assimilation method to recover free-surface flows. More generally, we proposed a particle filter method to reconstruct the complete state of free-surface flows from a sequence of depth images only. The data assimilation scheme introduced accounts for model and observations errors. We evaluated the developed approach on two numerical test cases: a collapse of a water column as a toy-example and a flow in an suddenly expanding flume as a more realistic flow. The robustness of the method to simulated depth data quality and also to initial conditions was considered. We illustrated the interest of using two observations instead of one observation into the correction step. Then, the performance of the Kinect sensor to capture temporal sequences of depth observations was investigated. Finally, the efficiency of the algorithm was qualified for a wave in a real rectangular flat bottom tank. It was shown that for basic initial conditions, the particle filter rapidly and remarkably reconstructed velocity and height of the free surface flow based on noisy measures of the elevation

6.2.3. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: 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 will be experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we wish also to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models. The longer-term goal is to learn empirical subgrid models directly from image data. An accurate modeling of this term is essential for Large Eddies Simulation as it models all the non resolved motion scales and their interactions with the large scales.

We have pursued the first investigations on a 4DVar assimilation technique, integrating PIV data and Direct Numerical Simulation (DNS), to reconstruct two-dimensional turbulent flows. The problem we are dealing with consists in recovering a flow obeying Navier-Stokes equations, given some noisy and possibly incomplete PIV measurements of the flow. By modifying the initial and inflow conditions of the system, the proposed method reconstructs the flow on the basis of a DNS model and noisy measurements. The technique has been evaluated in the wake of a circular cylinder. It denoises the measurements and increases the spatiotemporal resolution of PIV time series. These results have been recently published in the Journal of Computational Physics [7]. Along the same line of studies the 3D case is ongoing. The goal consists here to reconstruct a 3D flow from a set of simultaneous time resolved 2D images of planar sections of the 3D volume. This work has been mainly conducted within the PhD of Cordelia Robinson. The development of the variational assimilation code has been initiated within a collaboration with A. Gronskis, S. Laizé (lecturer, Imperial College, UK) and

Eric Lamballais (institut P' Poitiers). A High Reynolds number simulation of the wake behind a cylinder has been recently performed within this collaboration. The 4DVar assimilation technique based on the numerical code Incompact3D is now implemented. We are currently trying to reconstruct a 3D turbulent flow from dual plane velocity observations. The control of subgrid parameterizations will be the main objective of the PhD of Pranav Chandramouli that is just starting.

6.2.4. Ensemble variational data assimilation of large scale fluid flow 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. Preliminary encouraging results have been obtained for simulation under uncertainty of 1D and 2D shallow water models.

6.2.5. Reduced-order models for flows representation from image data

Participants: Mamadou Diallo, Cédric Herzet, Etienne Mémin, Valentin Resseguier.

During the PhD thesis of Valentin Ressguier 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 the POD-Galerkin method. The pertinence of this reduced order model has been successfully assessed on several wake flows. This study has been published in two conference papers and one journal article.

On the other hand, we investigated the problem of reduced-model construction from partial observations. In this line of search, our contribution was twofold. We first proposed a Bayesian framework for the construction of reduced-order models from image data. Our framework enables to account for any prior information on the system to reduce and takes the uncertainties on the parameters of the model into account. Interestingly, the proposed approach reduces to some well-known model-reduction techniques when the observations are not partial (i.e., the observation operator can be inverted). Second, we provided a theoretical analysis of our methodology 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 the publications of one contribution in the proceedings of the international conference on acoustics, speech and signal processing (ICASSP'16). A journal version of these contributions has been submitted.

6.3. Analysis and modeling of turbulent flows

6.3.1. Singular and regular solutions to the Navier-Stokes equations (NSE) and relative turbulent models

Participant: Roger Lewandowski.

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 paid to the eddy viscosity term. For this regularized problem and when the eddy viscosity has enough regularity, we have 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 will be soon 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 an addition 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 is 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.

6.3.2. 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 started to study the impact of the introduction of a random modeling of the friction velocity on the classical wall law expression.

6.3.3. Hot-wire anemometry at low velocities

Participant: Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. Based on a market study the french patent application has been extended this year to a Patent Cooperation Treaty (PCT) application.

6.3.4. 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. These temporally resolved volumic observations will be used to assess the algorithms developped in the PhD of Ioana Barbu and in the postdoc of Kai Berger. Then this data will be used in 4DVar assimilation technique to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Cordelia Robinson.

6.4. Visual servoing approach for fluid flow control

6.4.1. Closed-loop control of a spatially developing shear layer

Participants: Christophe Collewet, Johan Carlier.

This study aims at controling 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. As in all our previous works in flow control, we propose a vision-based approach to control this flow. We investigate the use of an optimal control based on a reduced linearized state space model of the Navier-Stokes equations. A steady desired state was first considered leading to a linear time-invariant system. The main problem consists to maintain the flow in his desired state in presence of unknown perturbation. Different strategies have been evaluated for different types of actuators and different cost functions. Even if our control law is based on a linearized approach, its efficiency has been validated on a realistic numerical Navier-Stokes 3D solver. This work has been submitted to the 20th World Congress of the International Federation of Automatic Control (IFAC).

6.5. Reactive transport

6.5.1. Reactive transport in porous media

Participant: Jocelyne Erhel.

In many environmental applications, transport of solutes is coupled with chemical reactions, either kinetic or at equilibrium. These reactions involve not only solutes, but also sorbed species and minerals. The mathematical model is a coupled set of nonlinear partial algebraic differential equations. A classical approach is to discretize first in space then in time. Since the problem is rather stiff, explicit time discretization suffers from a drastic CFL-like condition. On the other hand, implicit schemes allow large timesteps during some periods of simulation. Implicit Euler scheme is often used for monotonicity properties. The Jacobian is computed from the transport operator and the chemical operator. We have designed such a global approach and implemented it in our software GRT3D. We have done numerical experiments on the benchmark MoMaS.

Publications: 2 conferences and one journal article [15], [20], [21]

Grant: H2MNO4

6.5.2. Reactive transport in fractured-porous media

Participants: Yvan Crenner, Benjamin Delfino, Jean-Raynald de Dreuzy, Jocelyne Erhel.

Even in small numbers, fractures must be carefully considered for the geological disposal of radioactive waste. They critically enhance diffusivity, speed up solute transport, extend mixing fronts and, in turn, modify the physicochemical conditions of reactivity around possible storage sites. Numerous studies addressing various applications (e.g. radioactive waste storage, CO2 sequestration, geothermal storage, hydrothermal alteration) have shown that fractures cannot be simply integrated within an equivalent porous medium. Our objective is to develop a reactive transport model based on the separation of the fracture and matrix domains, with diffusion conditions differing between the fracture and in the matrix, appropriate flow-rock interactions at equilibrium in the matrix and fracture-matrix exchange conditions at their interface.

This year, we developed a numerical model for a chemical system with several minerals, which is representative of a storage site.

Publications: 2 conferences [28], [27]

Grant: ANDRA

6.6. Linear solvers

6.6.1. Sparse linear solvers

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.

Grants and projects: EXA2CT 8.2.1 , EoCoE 8.2.2 , C2S@EXA 8.1.7

Publications: 3 conférences [22], [23], [39]

LEMON Team

7. New Results

7.1. Ocean modeling

Participants: Fabien Marche, Antoine Rousseau.

7.1.1. A first discrete formulation for Green-Naghdi equations on unstructured general meshes

We introduce in [17] the first numerical method available in the literature to approximate the solutions of the Green-Naghdi equations on fairly general unstructured meshes. The method relies on coupled elliptic and hyperbolic problems, the first one accounting for a dispersive correction of the free surface flow description provided by the second one, and on discontinuous polynomial approximations of arbitrary order and the construction of discrete differential operators suitable for such non-conforming approximations. It allows to handle general meshes and nonconforming interfaces. A nonlinear stability result is proved, together with the preservation at the discrete level of motionless steady states. Several test cases highlight the accuracy of this discrete formulation.

7.1.2. Quasi-hydrostatic ocean models

In [9], 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 obtain new terms in δ/ε, where the domain aspect ratio and the Rossby number are both small numbers. We provide here some rigorous crossed-asymptotics with regards to these parameters , prove some mathematical and physical results on the nontraditional models, and situate them among traditional ones. This was also published as lecture notes given by Antoine ROUSSEAU in 2014: see [8].

7.1.3. Interface conditions for ocean models

In [4] we are interested in the search of interface conditions to couple hydrostatic and nonhydrostatic ocean models. To this aim, we consider simplified systems and use a time discretization to handle linear equations. We recall the links between the two models (with the particular role of the aspect ratio $\delta = H/L$) and introduce an iterative method based on the Schwarz algorithm (widely used in domain decomposition methods). The convergence of this method depends strongly on the choice of interface conditions: this is why we look for exact absorbing conditions and their approximations in order to provide tractable and efficient coupling algorithms.
In [3] we present a study of optimized Schwarz domain decomposition methods for Navier-Stokes equations. Once discretized in time, optimal transparent boundary conditions are derived for the resulting Stokes equations, and a series of local approximations for these nonlocal conditions are proposed. Their convergence properties are studied, and numerical simulations are conducted on the test case of the driven cavity. It is shown that conditions involving one or two degrees of freedom can improve the convergence properties of the original algorithm.

7.2. Renewable energies

Participant: Antoine Rousseau.

7.2.1. Wind circulation around mills

In [5] we present a new methodology, together with numerical studies, related to a Lagrangian stochastic approach applied to the computation of the wind circulation around mills. We present our numerical method and numerical experiments in the case of non rotating and rotating actuator disc models. First, for validation purpose we compare some numerical experiments against wind tunnel measurements. Second we perform some numerical experiments at the atmospheric scale and present some features of our numerical method, in particular the computation of the probability distribution of the wind in the wake zone, as a byproduct of the fluid particle model and the associated PDF method.

7.3. Multiscale modeling for environmental issues

Participants: Mathieu Dartevelle, Carole Delenne, Vincent Guinot, Antoine Rousseau.

7.3.1. Upscaled modeling of a coastal lagoon in Camargue

In 2015, Sélim Cornet developed a numerical model for the hydrodynamics of Vaccares system in Camargue. The data and reference simulations (made with TELEMAC-2D) were provided by Tour du Valat (contact O. Boutron). Sélim's work consisted in the implementation and validation of the porosity shallow water model developped by Vincent GUINOT , in order to obtain accurate but inexpensive simulations of the Vaccares hydrosystem. In 2016, we identified inconsistencies in the porosity closure model. These modeling issues have been analysed and a new theoretical approach, including new energy principles in the derivation of the porosity model, are under investigation.

7.3.2. Feedback strategies for decontamination of water resources

In [2] we show how to couple systems of ODEs and PDEs to provide efficient feedback strategies for the biological decontamination of water resources. For natural resources, we impose not to introduce any bacteria in the resource and to treat it aside preserving a constant volume of the resource at any time. The feedback strategies are derived from the minimal time synthesis of the system of ODEs.

7.3.3. Dispersion in porous media

Solute dispersion in porous media is usually modelled using Fick's law or fractional variations of the solute dispersion equation. The Fickian model, however, is known to exhibit a number of drawbacks, such as poor scaling properties. This is also true for its fractional counterparts, that perform with limited success when compared to experimental data sets. In [46], a high-quality experimental device is built in the form of periodic heterogeneities (Model Heterogeneous Porous Medium) of length 15 cm. Placing up to 10 MHPM in series allows the scaling properties of the dispersion model to be analyzed. Besides providing a high quality experimental database, the results in [46] indicate that (i) previously identified scaling trends for the dispersion coefficient may easily be explained by experiment variability, (ii) there exists a linear transport model that allows the experimental behaviour to be reproduced at all scales, (iii) this model is not the advection-dispersion model (even fractional). More experiments have been performed this year with a different connexion between each MHPM. More experiments have been performed this year with a different connexion between each MHPM. The benchmarking of various numerical models is currently under process; it includes classical models such as Advection-Diffusion, Mobile-Immobile, Multi Rate ... as well as a proposed Purely Advective Multi Region model.

7.3.4. Modeling and identification for environmental applications

In collaboration with Mohsen Chebbi (ENIT, Tunis) and Salwa Toumi (ENIT, Tunis), we propose stochastic models of anaerobic membrane bioreactors [10]. These biotechnology processes are usually described as differential equations valid at large population scale. We propose model at different scales. At the microscopic scale, we consider a pure jump stochastic model that can be exactly simulated. However, when the size of the population is large that type of exact simulation is not feasible, hence we propose approximated simulation methods in discrete time, of the Poisson type or of the diffusive type. We establish the law of large numbers and the central limit theorem of the functional type.
We also consider different problems of simultaneous filtering and parameter estimation for hidden Markov models: in collaboration with Samuel Nyobe Som (University of Yaoundé 1) we study natural ressources examples; in collaboration with Oussama Hadj-Abdelkader (University of Tlemcen) we study applications in biotechnology. In both cases the fact that the frequency of data acquisition is slow enough to improve classical techniques.

7.4. Other results

Participants: Fabien Campillo, Carole Delenne, Antoine Rousseau.

7.4.1. Topography assessment from ordinal and continuous information

Hydrodynamic models in two dimensions require a precise knowledge of the domain topography, but data acquisition (field surveys, RADAR, etc.) remains difficult to set up at a large scale. Progress in remote sensing data now allows the automatic monitoring of water surfaces delineation from areal or satellite images (e.g. [48]); and flood dynamics from remote sensing data are known to be informative on floodplain topography for long. The idea is thus to combine sparse punctual information (obtained from ground survey) with continuous contour lines (obtained from image treatment) to better assess the domain topography. Two different approaches have been tested during Mathieu Dartevelle's internship (3 months): the first one is based on geostatistical considerations (kriging and conditional simulations) and the second one, deterministic, uses spline functions obtained from a minimisation process. The main challenge stands in the fact that, if the contour line is known to be an isovalue curve, its elevation is not known. First results have been presented in [12] but work is still needed especially to retrieve a precise estimation of curve elevation from very few data points. This work is done in collaboration with Jean-Stéphane Bailly (Lisah, AgroParisTech Montpellier).

7.4.2. Growth-fragmentation-death models

In collaboration with Coralie Fritsch (Inria Nancy) and Otso Ovaskainen (University of Helsinki), we propose a numerical approach that can be used to study the invasion fitness of a mutant in evolutionary models and to determine evolutionary singular strategies when the competitive exclusion principle holds [18]. Though the method is general, we illustrate this method with a mass-structured individual-based chemostat model. We assume that the mutations are rare and that the resident population is large, in which case the mutant population can be viewed, on a short time scale, as evolving in a constant environment. Both deterministic and stochastic models can be proposed to describe such a problem. We exploit a previously derived mathematical relationship between these models [7] to derive a general method for analyzing the invasion fitness of stochastic models.
In collaboration with Nicolas Champagnat and Coralie Fritsch (Inria Nancy), we studied the variations of the principal eigenvalue associated to a growth-fragmentation-death equation with respect to a parameter [16]. 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.

MAGIQUE-3D Project-Team

6. New Results

6.1. Seismic Imaging and Inverse Problems

6.1.1. Time-harmonic inverse problem

Participants: Hélène Barucq, Florian Faucher.

We study the seismic inverse problem for acoustic and elastic medium associated with the time-harmonic wave equation, and the underlying recovery of geophysical parameters. We employ Full Waveform Inversion (FWI) where the multi parameters reconstruction is based on iterative minimization techniques. This inverse problem shows a Lipschitz stability where the stability constant is related to the (conditional) lower bound of the Fréchet derivative, when assuming a piecewise constant representation of the parameters. We successively estimate the stability constant for different model partition in order to control the convergence of the scheme. Hence we define a multi-level (multi-scale, multi-frequency) algorithm where the natural progression of frequency is paired with the model partition. The method is implemented and numerical experiments are performed for elastic medium reconstruction, in particular for realistic geophysical situations.

6.1.2. Shape-reconstruction and parameter identification of an elastic object immersed in a fluid

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Julien Diaz, Rabia Djellouli.

We investigate the inversion of a series of parameters in the context of a 2D elasto-acoustic scattering problem. The inverse problem is solved by using a Newton-like method, where the shape of the scatterer is assumed to be Lipschitz-continuous. Herein, we want to recover the shape and the material parameters in the case of isotropic and anisotropic materials. Based on the different influences of these parameters on the far field pattern, the final goal is to propose an iterative algorithm to retrieve the parameters separately, by devoting some iterations to the reconstruction of the shape and the others to the determination of the parameters. On the other hand, due to the difficulties to retrieve the material parameters, the penetrability of scatters have been studied. The conclusion has been that the recovery of material parameters can be feasible, provided that the scattered waves are not completely reflected. The results of this work have been presented to the conference Inverse Problems for PDE in Bremen, Germany [24].

6.2. Mathematical modeling of multi-physics involving wave equations

6.2.1. A study of the numerical robustness of single-layer method with Fourier basis for multiple obstacle scattering in homogeneous media

Participants: Hélène Barucq, Juliette Chabassier, Ha Pham, Sébastien Tordeux.

We investigate efficient methods to solve direct and inverse problems for the propagation of acoustic wave in strongly inhomogeneous media in low-frequency regime. We start our investigation with inhomogeneities created by compactly-supported and non-overlapping obstacles. With a large number of small obstacles, optimized softwares based on Finite Element Method (FEM) 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 scattered field. We limit our numerical experiments to disc-shaped obstacles. We first compare our method with Montjoie (a FEM-based software); secondly, we investigate the efficiency of different solver types (direct and iterative) in solving the dense linear system generated by the method. We observe that the optimal choice depends on the distance between obstacles, their size and number, and applications.

6.2.2. Derivation and validation of impedance transmission conditions for the electric potential across a highly conductive casing

Participants: Hélène Barucq, Aralar Erdozain, David Pardo, Victor Péron.

Borehole resistivity measurements are a common procedure when trying to obtain a better characterization of the Earth's subsurface. The possible risk of having borehole collapses makes the employment of a casing very suitable for this type of scenarios. Such casing protects the borehole but it also highly complicates the resistivity measurements due to the thinness of the casing and the large contrasts between the conductivities of the casing and the rock formations.

This work is motivated by realistic configurations where the resistivity of the casing is proportional to the cube of the thickness of the casing. In this framework, our aim is to derive Impedance Transmission Conditions (ITCs) for the electromagnetic field across such a casing. As a first approach we derive ITCs for the electric potential. We consider a transmission problem for the static case of the electric potential, set in an axi-symmetric borehole shaped domain. This domain is composed of three different subdomains, the interior part of the borehole, the rock formations and the metallic casing.

In this framework, we address the issue of ITCs using two different approaches. The first one consists in deriving ITCs across the casing itself, whereas the second approach tackles the problem by deriving ITCs on an artificial interface located in the middle of the casing. We derive different models for the two considered approaches and we numerically assess them with a finite element method implementation. Then we perform a comparison on these models by showing the advantages and drawbacks of each model. Finally, we show an application to a borehole through-casing resistivity measurement scenario. This work delivers stability results and error estimates, leading to convergence of each approximate model. All the details regarding this work can be found in [43].and [10]. In addition it has been presented to the WONAPDE Conference [29].

6.2.3. Semi-analytical solutions for asymptotic models for the electric potential across a highly conductive casing

Participants: Hélène Barucq, Aralar Erdozain, Ignacio Muga, Victor Péron.

This work is performed in the framework of borehole through-casing resistivity measurements. A transmission problem for the electric potential is considered, where one part of the domain is a high-conductive casing. Numerical instabilities are created during the numerical simulations when such a casing is present in the configuration. Therefore, three different asymptotic models derived in [43] are considered, which are composed of impedance conditions specially designed to avoid the casing. These models correspond to approximations of orders one, two and four.

In this work, we employ analytical methods for the aforementioned asymptotic models, which provide a consistent solution to test and verify the numerical solutions (Finite Element Method). In addition, these methods are computationally cheaper than the purely numerical methods. The standard method we follow consists in employing cylindrical coordinates and assuming material homogeneity in the vertical and angular variables. The source term is represented as a Dirac distribution. Under these conditions, we represent the solution to our problem as an inverse Fourier integral in the vertical variable, and a Fourier series in the angular variable.

Numerical tests are carried out to compare with Finite Element solutions. Several difficulties have to be taken into account during the implementation of the semi-analytical solutions, like the treatment of the Dirac distribution and the presence of singularities when the Fourier variable tends to zero. These difficulties are also addressed in this work which is detailed in [10].

6.2.4. Numerical investigation of instabilities of Perfectly Matched Layers coupled with DG-schemes in elastodynamic

Participants: Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz, Simon Ettouati.

We observed long-term numerical instabilities when DG-schemes are coupled with PML in elastodynamic, even with isotropic media. To investigate the causes of this instabilities, we have led a series of numerical experiments with elasticus 5.1 . The conclusion was that the instabilities only appear in truly elastodynamic media (i.e. when the velocities of S waves is positive) and that different factors impact the stability : the heterogeneities of the domain, the choice of the fluxes, the boundary conditions, the use of unstructured meshes... In the best scenario, using a cartesian grid with periodic boundary conditions for an homogeneous medium and centered fluxes, we did not observe instabilities. However, changing only one element of the configuration made the instabilities appear. Our conclusion is that we need a very particular flux in the PML that should be able to handle the heterogeneities of the domain and the structure of the mesh. This flux should also be adapted to discretize the boundary condition. We are now working on the design of this flux.

6.2.5. Elasto-acoustic coupling

Participants: Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz, Simon Ettouati.

Last year, we developed a Discontinuous Galerkin Method for the elastoacoustic coupling in time domain. The proposed solution methodology in general and can be applied to any kind of fluxes. The method had been implemented in Elasticus 5.1 and we have transfered it into the Total plateform TMBM-DG 5.4 .

In [23], we have considered elastoacoustic coupling with curved interfaces and we have proposed a solution methodology based on Finite Element techniques, which allows for a flexible coupling between the fluid and the solid domain by using non-conforming meshes and curved elements. Differently from other non-conforming approaches proposed so far, our technique is relatively simpler and requires only a geometrical adjustment at the coupling interface at a preprocessing stage, so that no extra computations are necessary during the time evolution of the simulation. This work, has been achieved in collaboration with Angel Rodriguez Rozas, former post-doc of the team.

6.2.6. Atmospheric radiation boundary conditions for helioseismology

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé.

Modeling acoustic wave propagation inside a celestial body (as the Sun) prompts the question of imposing an adequate boundary condition. Classical atmosphere models suppose an exponential decay of the medium density and a constant wave celerity outside a given radius. This work proposes several radiation boundary conditions that mimic the presence of such an atmosphere and assesses their behavior numerically in radial and axisymmetric configurations.

6.2.7. Hybrid discontinuous finite element approximation for the elasto-acoustics.

Participants: Hélène Barucq, Henri Calandra, Julien Diaz, Elvira Shishenina.

Discontinuous finite element methods proved their accuracy and flexibility, but they are still criticized for the number of degrees of freedom which they use: it is much higher than the ones of the conventional methods based on continuous approximations.

Thus hybrid methods have been developed and their integration into the DIP is under way, both in the acoustic and elastic domains.

The global purpose of this work is to develop a new approach for solving wave equation in discontinuous function spaces. This will provide all propagators already developed in the CARBON platform. Possible directions in this research are for example the development of a Trefftz type approximation for elasto-acoustics, coupling with VEM, HDG.

Our current work is concentrated on using Trefftz method. The main idea of the method is that chosen basis functions of Trefftz approximation space are discrete local solutions of the initial equations to be solved.

The possible advantages of Trefftz type approximations compared to the standard ones are: 1) better orders of convergence; 2) flexibility in the choice of basis functions; 3) low dispersion; 4) incorporation of wave propagation directions in the discrete space; 5) adaptivity and local space-time mesh refinement.

The particularity of Trefftz methods is that in case of applying to time-dependent problems they require a space-time mesh.

We studied theory of application of the method to the coupled acoustic system, and implemented numerically Trefftz method to solve the first-order 1D acoustic wave propagation system. The obtained results were presented during annual workshop in Houston organized by Depth Imaging Partnership between Inria and Total.

6.3. Supercomputing for Helmholtz problems

6.3.1. Extend task-based node parallelism to cluster level: applications to geophysics

Participants: Emmanuel Agullo, Lionel Boillot, George Bosilca, Henri Calandra, Corentin Rossignon.

The context of this work is to replace static parallelism based on MPI + threads and/or CUDA by dynamic task-based parallelism on top of runtime systems. On a previous work, we demonstrated the speed-up of the new solution when applied to geophysics, at a node level. Moreover, this task paradigm proved its flexibility on several architectures such as ccNUMA big nodes or many-core Intel Xeon Phi co-processors.

We extended this principle to a set of nodes, eventually heterogeneous, in order to measure performance at a cluster level. Preliminary results on few homogeneous nodes were encouraging, ie still faster than pure MPI. Unfortunately, the geophysics algorithm being too repetitive, the load-balancing issue which can be removed within a node (i.e. between cores) comes back between nodes when they are numerous or few but heterogeneous. This is due to the work-stealing feature of the task paradigm which is by default enabled at the node level only.

To overcome this problem, we extended the work-stealing feature to cluster level. To do that we used the task identification by geometrical sub-meshes to detect candidates that can be exchanged between nodes. Then, we compared PAPI counters on these tasks to find the best choice. Finally, we use a separate task-based program to automatically do the main code task update. Preliminary results show clear improvement of load-balancing at cluster level.

This work has been presented to the conferences Rice Oil&Gas[34] and SIAM-PP (Parallel Processing) [35].

6.3.2. Numerical libraries for hybrid meshes in a discontinuous Galerkin context

Participants: Hélène Barucq, Lionel Boillot, Aurelien Citrain, Julien Diaz.

Elasticus team code 5.1 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 finalizing now the unstructured-goal.

The purpose is to use the h-adaptivity of discontinuous Galerkin method to easily encircle unstructured tetrehedra with hexahedra to form hybrid meshes (so in 3D). In addition, it would be interesting to couple numerical methods depending on the element types.

6.3.3. Code transfer: TMBM-DG/THBM into Total R&D environment

Participants: Lionel Boillot, Julien Diaz.

The goal of the DIP collaboration between Total and Inria is to transfer the validated research codes. At first, DIVA-DG has been created in conjunction with Total developers team. It concerns the time modeling of wave propagation. Then, we forked it into Elasticus code to focus on mathematical research at the Inria side. Finally, once validated, we managed its transfer into the recent Total R&D environment (so instead of DIVA template, we moved to TMBM template) to form the TMBM-DG 5.4 code. The entire code has been transferred now, including unit tests and full documentation.

In the meantime, another code emerged within the DIP collaboration, THBM, concerning the frequency modeling of wave propagation. The development is directly done since the beginning in the Total R&D environment. An important part is already validated while research still continues.

6.3.4. Hybridizable Discontinuous Galerkin methods for solving the elastic Helmholtz equations

Participants: Marie Bonnasse-Gahot, Henri Calandra, Julien Diaz, Stéphane Lanteri.

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. This results were presented in a submitted paper.

Moreover, as the HDG global matrix is very sparse, we focus on a suitable solver for this kind of matrix. We 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. We compared the performances of the two solvers when solving 3D elastic waves propagation over HDGm. These comparisons were presented at the MATHIAS 2016 conference and at the DIP Workshop [36], [37]

6.3.5. A Symmetric Trefftz-DG Formulation based on a Local Boundary Element Method for the Solution of the Helmholtz Equation.

Participants: Hélène Barucq, Abderrahmane Bendali, M'Barek Fares, Vanessa Mattesi, Sébastien Tordeux.

A general symmetric Trefftz Discontinuous Galerkin method is built in [12] for solving the Helmholtz equation with piecewise constant coefficients. The construction of the cor- responding local solutions to the Helmholtz equation is based on a boundary element method. A series of numerical experiments displays an excellent stability of the method relatively to the penalty parameters, and more importantly its outstanding ability to reduce the instabilities known as the "pollution effect" in the literature on numerical simulations of long-range wave propagation.

6.4. Hybrid time discretizations of high-order

6.4.1. High order time discretization for dissipative wave equations.

Participants: Juliette Chabassier, Julien Diaz, Anh-Tuan Ha, Sébastien Imperiale.

Magique-3D team is interested in numerical methods for wave propagation in realistic media, which are naturally dissipative in many application cases. In this internship, we wish to investigate several dissipation models, that lead to Partial Differential Equations with different structures. The simplest model is the scalar wave equation with homogeneous and constant damping $\frac{\partial^2 u}{\partial t^2} + R\frac{\partial u}{\partial t} - \Delta u = f$. In order to approach the complexity of the propagating medium and its geometry, high order finite elements in space are used. Once the spatial discretization is fixed, we get a differential equation of the kind $\frac{d^2 u_h}{dt^2} + B_h \frac{du_h}{dt} + A_h u_h = f_h$, where the mass matrix is the identity thanks to the mass lumping technique followed by a renormalization, B_h is the dissipation matrix and A_h the stiffness matrix. Classically, this equation is discretized in time with a centered and second order finite difference scheme known as the θ-scheme ($\theta > 0$)

$$\frac{u_h^{n+1} - 2u_h^n - u_h^{n-1}}{\Delta t^2} + B_h \frac{u_h^{n+1} - u_h^{n-1}}{2\Delta t} + A_h \left(\theta u_h^{n+1} + (1-2\theta)u_h^n + \theta u_h^{n-1}\right) = f_h^n \qquad (1)$$

In order to preserve the precision obtained with high order finite elements in space, we wish to design higher order time discretizations, while preserving some interesting mathematical properties as the dissipation of a discrete energy, and an efficiency close to the one observed for the second order scheme. More precisely, if $\theta = 0$ and B_h is diagonal, scheme (1) only requires the inversion of a diagonal matrix at each time step.

We want to use the technique of the modified equation, which consists in compensating the first term of the consistency error of a low order discretization, by adding a well chosen new term. If $\theta = 0$, this approach leads to the following fourth order accurate in time scheme

$$\left(I_h + \frac{\Delta t^2}{12} B_h\right) \frac{u_h^{n+1} - 2u_h^n - u_h^{n-1}}{\Delta t^2} + \left[B_h + \frac{\Delta t^2}{12}\left(B_h A_h - A_h B_h\right)\right] \frac{u_h^{n+1} - u_h^{n-1}}{2\Delta t} + A_h u_h^n - \frac{\Delta t^2}{12} A_h^2 u_h^n = \widetilde{f}_h^n$$

Even if B_h is diagonal, A_h and B_h do not commute in general. We propose to replace the matrix $B_h A_h - A_h B_h$, potentially hard to invert, by an approximated matrix, easy to invert, without deteriorating the consistency of the scheme.

An article is being written and will be submitted soon.

6.4.2. High order conservative explicit and implicit schemes for wave equations.

Participants: Juliette Chabassier, Sébastien Imperiale.

In 2016 we have studied the space/time convergence of a family of high order conservative explicit and implicit schemes for wave equations. An original proof of convergence has been proposed and provides an understanding of the lack of convergence of some schemes when the time step approaches its greatest admissible value for stability (CFL condition). An article has been submitted.

6.4.3. Efficient high order implicit time schemes for Maxwell's equations.

Participants: Hélène Barucq, Marc Duruflé, Mamadou N'Diaye.

The Padé approximant is well known to be one of the best approximation of an exponential function which is involved in the exact solution of the linear ODE (Ordinary Differential Equations):

$$y'(t) = Ay(t) + F(t)$$

where A is a given matrix (usually coming from finite element discretization) and F is a term source. The numerical solution can be constructed by approximating the exponential function using the diagonal Padé approximant:

$$R(z) = \frac{P_m(z)}{Q_m(z)}$$

The function $R(z)$ is a fraction involving two polynomial P_m and Q_m of same degree and approximating the exponential. The corresponding scheme is implicit and A-stable in the sense of Dahlquist. The associated stability function is the same as the stability function of the Gauss-Runge-Kutta schemes. However, Gauss-Runge-Kutta schemes can be used to handle non-linear ODEs, but they are too expensive to use in practice. The diagonal Padé schemes presented here can be seen as a simplification of Gauss-Runge-Kutta schemes in the case of linear ODE. We have proposed an efficient way to implement the diagonal Padé schemes with an accurate approximation of the source term to keep the correct order of accuracy.

The main drawback of Padé schemes is that the denominator $Q_m(z)$ has distinct roots. It implies that we have to solve distinct linear systems at each time step. As a result, we have also studied the case where the denominator has an unique real root γ :

$$R(z) = \frac{N(z)}{(1-\gamma z)^m}$$

The numerator N is then found to obtain the "best" approximation of the exponential under the constraint of the A-stability property of the underlying schemes. The obtained schemes have been called Linear Singly Diagonal Implicit Runge-Kutta schemes (Linear SDIRK) since they share the same property as SDIRK (a unique linear system to solve several times) but they can be applied only to linear ODEs. We provide a performance assessment of different implicit schemes (Padé schemes, SDIRK and Linear SDIRK). The comparison criteria are based on the amplitude and phase errors which are reliable gauges of accuracy when approximating waves problems. The Linear SDIRK schemes and the diagonal Padé schemes have been implemented in the code Montjoie. We have performed numerical experiments in 1-D and 2-D for Maxwell's equations to validate these schemes and compare their efficiency.

This work has been presented at the conference ICOSAHOM [28], the colloquium Inter'Actions en Mathématiques Lyon 2016 and the Mathias annual Total seminar [27].

6.4.4. Optimized high-order explicit Runge-Kutta-Nyström schemes.

Participants: Marc Duruflé, Mamadou N'Diaye.

In this work we propose a high order time integration explicit scheme to solve a second order derivative non-linear ordinary differential equation (ODE)

$$y'' = f(t, y)$$

To solve this family of ODEs, explicit one-step Runge-Kutta-Nyström have been proposed by Hairer et al. The stability condition (CFL) associated with these schemes have been studied for order 3, 4 and 5 by Chawla and Sharma. In this work, we have extended the stability studies for high order. We proposed optimal coefficients for Runge-Kutta-Nyström schemes of order 6, 7, 8 and 10 which have been obtained by optimizing the CFL. With the obtained optimal CFL, these schemes are well suited for stiff problems where the stability condition is restrictive. These schemes have been implemented in the code Montjoie.

Numerical experiments have been conducted in 1-D for the non-linear Maxwell's equation and show that obtained Runge-Kutta-Nyström schemes of order 7 is quite efficient. This work has been presented at the conference ICOSAHOM [48].

SERENA Team

6. New Results

6.1. Numerical algorithms for simulating diffusion processes in discontinuous media

Participant: Géraldine Pichot.

Grants: H2MN04 3

Software: SBM 5.2

Publications: [19]

We present several benchmark tests for Monte Carlo methods simulating diffusion in one-dimensional discontinuous media. These benchmark tests aim at studying the potential bias of the schemes and their impact on the estimation of micro- or macroscopic quantities (repartition of masses, fluxes, mean residence time,...). These benchmark tests are backed by a statistical analysis to filter out the bias from the unavoidable Monte Carlo error. We apply them on four different algorithms. The results of the numerical tests give a valuable insight of the fine behavior of these schemes, as well as rules to choose between them.

6.2. Locally space-time efficient estimates for parabolic problems

Participants: Martin Vohralík, Alexandre Ern, Iain Smears.

Grants: GATIPOR 8.3.1

Publications: [33]

In [33], we derive for the first time a posteriori error estimates for parabolic problems which are both globally reliable and locally space-time efficient. By this, one means that the error between a known approximate numerical solution and the unknown exact solution of a model parabolic PDE (the heat equation) is bounded from above on the whole space-time domain by a fully computable estimator, while this estimator does not overestimate significantly the error and localizes it both in space and in time. More precisely, the estimator also gives lower bounds on the error, up to a generic constant, and this on each time interval and in a small neighborhood of each space mesh element. We consider arbitrarily high-order conforming Galerkin spatial discretizations and arbitrarily high-order discontinuous Galerkin temporal discretizations, and the error is measured in a norm composed of the $L^2(H^1) \cap H^1(H^{-1})$-norm augmented by the temporal jumps of the numerical solution. The efficiency constant is robust with respect to (independent of) any mesh-size, time-step size, and the spatial and temporal polynomial degrees. The proposed estimators also have the practical advantage of not imposing any requirement on coarsening between the consecutive time steps.

STEEP Project-Team

7. New Results

7.1. Ecological accounting

Besides the publication of the article [2] on environmental pressures in supply chains in the leading journal in the field (*Journal of Industrial Ecology*), the most important result obtained on this front this year 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 teeams 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.

7.2. Modeling of human-mediated dispersal via road network in invasive spreads

In the case of ecosystem invasions, human-mediated dispersal often acts as a vector for many exotic species, both at the introduction and secondary spread stages. The introduction stage is mainly a consequence of human-mediated long distance dispersal and is known to happen at continental or global scales. Secondary spread, however occurs at smaller spatial and time scales (e.g. landscape), and can result from natural or human-mediated dispersal. Despite the importance of local goods and materials transportation (e.g. for landscaping, construction, or road-building) potentially promoting the spreading of invasive species, few studies have investigated short distance human-mediated dispersal. This lack of consideration seems to be the consequence of multiple factors:

- human-mediated dispersal is generally considered as a long distance dispersal process, more important for invasive species introduction than for secondary spread;
- it is difficult to qualify and quantify this mode of dispersal because of the multiplicity of potentially involved human activities;
- for organisms that can disperse naturally, it is complicated to distinguish between natural and human-mediated dispersal, as they may occur at similar scales.

Even though a range of methodologies are available for describing population spread by natural dispersal, only few models have been developed to describe and predict human-mediated dispersal consequences at small scales, and none of them take into account the topology of the transport infrastructure (roads, waterways). In this result, and in order to fill this gap and provide new insights into how invasion dynamics impact ecosystem services, we combined ecological (invasive species occurrence data) and geographical (transportationnetwork topology) data in a computer model to provide estimated frequencies and distances of materials transportations through the landscape. In this study (cf. [7]), we investigated the spreading pattern of Lasius neglectus, an invasive ant species originating from Turkey, which spread into Europe in the last decades. In this species, no mating or dispersal flights are performed, and its spread is therefore solely ensured by the transport of soil materials in which individuals are present. We built a numerical model enabling the estimation of multiple human-mediated dispersal parameters based on ground-truth sampling and a priori minimizing. After having built a model of the landscape-level spreading process that takes explicitly into account the topology of the road network, we localized the most probable sites of introduction, the number of jump events, as well as parameters of jump distances linked to the road network. Our model was also able to compute presence probability map, and can be used to calibrate sampling campaigns, explore invasion scenarios, and more generally perform invasion spread predictions. It could be applied to all the species that can be disseminated at local to regional scales by human activities through transportation networks.

7.3. 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 [9]. 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.

7.4. 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 [6]. 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).

We have also collaborated with the AIRSEA project-team towards applying novel sensitivity analysis tools to study the influence of the different parameter sets of a Tranus model [13]. The rationale is to then apply optimization methods to the most influential parameters. As a result, we were able to calibrate a real-life Tranus model such that results were of higher quality than with the baseline ad hoc approach, while reducing calibration time significantly.

TONUS Team

6. New Results

6.1. Time scheme for finite elements code for fluids models

Participants: Emmanuel Franck, Philippe Helluy, David Coulette, Ahmed Ratnani, Eric Sonnendrücker.

The finite element code JOREK use currently a classical implicit solver for reduced MHD model coupled with a block Jacobi preconditioning. For the future full MHD code we propose to change the solver in time to reduce the memory consumption and improve the robustness. During this year two directions have been followed. The first one is based on the classical physics-based preconditioning proposed by L. Chacon. Firstly, we have generalized this method by rewriting the preconditioning as a splitting scheme which separates the advection terms and the acoustic part and by generalizing the splitting algorithm. We obtain different solutions with different advantages. These different splitting schemes have been tested on simplified models and are currently tested on the Euler equations. The second direction is to use a relaxation scheme which allows to rewrite a nonlinear system as a linear hyperbolic system (larger that the previous one) and a nonlinear local source term. Using a splitting scheme we obtain a very simple method where in the first step we solve independent linear transport problems and in a second step we have some nonlinear projections. With a good parallelism and good solver for the transport subproblems the algorithm is very efficient compared to the classical one.

6.2. Preconditioning for elliptic solvers

Participants: Emmanuel Franck, Mariarosa Mazza, Ahmed Ratnani, Eric Sonnendrücker, Stefano Serra-Capizzano.

The different algorithms to discretize in time the MHD or to design preconditioning use solvers for a lot of elliptic operators like Laplacian. For high order finite elements like B-Splines the classical multi-grid methods are not very efficient. Indeed the number of iterations to converge increases strongly when the polynomial order increases. Using a theory called GLT, proposed by S. Serra-Capizzano, we have implemented and validated a smoother for multi-grid, able to obtain the convergence quasi independent of the polynomial degree. This method is also efficient as a preconditioning for mass matrices. We obtain at the end, very robust solvers for these simple problems and allows to perform the time algorithm for fluid models. The next step is to extend this method for more complex problems like vectorial elliptic problems.

6.3. Implicit Lattice Boltzmann scheme for fluid models

Participants: Emmanuel Franck, Philippe Helluy, David Coulette, Conrad Hillairet.

Many systems of conservation laws can be written under a lattice-kinetic form. A lattice-kinetic model is made of a finite set of transport equations coupled through a relaxation source term. Such representation is very useful:

- easy stability analysis, possibility to add second order terms in a natural way;
- can be solved by a splitting strategy;
- easy-to-implement implicit schemes, avoiding CFL constraint;
- high parallelism.

We have started to work on such approaches for solving the MHD equation inside a tokamak (postdoc of David Coulette). We have programmed a generic parallel lattice-kinetic solver in Kirsch, using the StarPU runtime. It presents a very good parallel efficiency. We have also started studying more theoretical aspects: stability of kinetic models, higher order time-integration, viscous terms modeling.

6.4. Hybrid computing

Participants: Philippe Helluy, Nhung Pham, Michel Massaro, Pierre Gerhard, David Coulette, Laura Mendoza, Conrad Hillairet.

In order to harness hybrid computers architecture, we have developed software and algorithms that are well adapted to CPU/GPU computing. For instance we have applied a task-graph approach for computing electromagnetic waves (https://hal.archives-ouvertes.fr/hal-01134222). We have also used an OpenCL-based GPU version of schnaps for computing a drift-kinetic plasma model (Nhung Pham's PhD). Recently, we have also developed a new implementation of the Discontinuous Galerkin solver into schnaps. We now use the StarPU runtime (http://starpu.gforge.inria.fr) for addressing automatically CPUs or GPUs available on the computational node. This development has been applied to the MHD equations (thesis of Michel Massaro). The new development will now be applied to kinetic acoustic simulations (Pierre Gerhard's PhD), gyrokinetic plasma simulations (Laura Mendoza's Postdoc) and implicit MHD simulations (Conrad Hillairet's PhD).

6.5. DG scheme for Drift-Kinetic equation

Participants: Laurent Navoret [correspondent], Philippe Helluy, Nhung Pham.

Using the discontinuous Galerkin solver of Schnaps, we have implemented a numerical scheme for the drift-kinetic model (in a cylinder geometry). The equation is written as an hyperbolic system after reduction in velocity (using spectral finite element). The code is parallelized on a multi-CPU or GPU architecture using OpenCL instructions. To solve the quasineutral equation (for the electric potential), the elliptic solver (already present in Schnaps) has been extended to be used slice by slice (of the cylinder). We have started by validating the code on the 2D guiding-center model and the diocotron instability test-case: we observe that the geometry approximation of the computational domain has a major impact on the precision of the numerical simulations.

6.6. Quasi-neutrality equation in a polar mesh

Participants: Michel Mehrenberger, Philippe Helluy, Guillaume Latu, Nicolas Crouseilles, Christophe Steiner.

In the quasi-neutrality equation in GYSELA, we are now able to treat correctly the inner radius thanks to a simple trick by taking the inner radius $\frac{\Delta r}{2}$. We also continue working on the gyro-average approximation. The new Padé method depends on a parameter ε. When setting ε to a large value, the solution is very similar to the classical Padé one, while taking small value for ε leads to a solution very near to the one obtained using the interpolation method (which approximates better the exact operator, but which can however lead to unstable results as it does not damp high modes). We can then prevent the scheme from instability, by setting large ε, but not too large in order to be more accurate than the classical Padé approximation. Further study in GYSELA is under discussion.

6.7. PICSL: Particle in Cell and Semi-Lagrangian schemes for two species plasma simulations

Participants: Michel Mehrenberger [correspondent], Sever Hirstoaga, Joackim Bernier, Yann Barsamian.

We have worked at CEMRACS 2016 on an algorithm that handles both the Particle in Cell method and the Semi-Lagrangian method in the context of a $2D \times 2D$ to handle the different scales associated with the ions and the electrons in Vlasov-Poisson simulation. Using PIC methods for the electrons allows to use easily specific numerical methods for fast dynamic. Numerical results are in accordance with the dispersion relation.

6.8. TARGET: TArgeting Realistic GEometry in Tokamak code gysela

Participants: Michel Mehrenberger, Nicolas Bouzat, Guillaume Latu, Camilla Bressan, Virginie Grandgirard.

We have worked at CEMRACS2016 on a new variant for the interpolation method to handle both mesh singularity at the origin and non circular geometry. It is based on a non uniform number of points for each closed flux line (intersection of the flux surfaces with the poloidal plane), which are concentric circles in the case of the circular geometry. This strategy, following previous works on curvilinear geometry and hexagonal meshes, should allow to generalize the work in [14] to non circular tokamaks.

6.9. Field aligned interpolation for gyrokinetics

Participants: Michel Mehrenberger, Maurizio Ottaviani, Yaman Güclü, Guillaume Latu, Eric Sonnendrücker.

A theoretical justification of the field align 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). A paper has been submitted [14].

6.10. High order implicit time splitting schemes for the BGK model

Participants: Michel Mehrenberger, Philippe Helluy, Laurent Navoret, David Coulette, Emmanuel Franck.

In the context of the Lattice Boltzmann or relaxation methods (6.1 -6.3), it is interesting to obtain a very high order implicit splitting. For this, we have considered a time splitting discretization of the BGK model with 3 velocities. First and second order schemes are studied before using Strang splitting coupled with a Semi Lagrangian or a Cranck-Nicholson DG scheme. Using complex time steps and composition methods, we obtain 4th order time step, unconditionally stable for the discrete BGK models. These results could be used with the Lattice Boltzmann method, the relaxation method and also the kinetic model.

6.11. Particle-In-Cell simulations for Vlasov-Poisson equations

Participants: Sever Hirstoaga, Yann Barsamian.

In the work [3], we implement in Selalib an efficient, regarding the memory access, Particle-In-Cell method which enables simulations with a large number of particles. Numerical results for classical one-dimensional Landau damping and two-dimensional Kelvin-Helmholtz test cases are exposed. The implementation also relies on a standard hybrid MPI/OpenMP parallelization. Code performance is assessed by the observed speedup and attained memory bandwidth. A convergence result is also illustrated by comparing the numerical solution of a four-dimensional Vlasov-Poisson system against the one for the guiding center model.

Then, we continued to optimize the code by analyzing different data structures for the particles (structure of arrays vs. arrays of structure) and for the grid fields (using space-filling curves like Morton, Hilbert etc.) with the aim of improving the cache reuse. In addition, we added the functionality of vectorization from the compiler and we obtained significant gain by testing the different data structures. We thus achieved to run PIC simulations processing 65 million particles/second on an Intel Haswell architecture, without hyper-threading. The hybrid parallelization through OpenMP/MPI gave satisfactory strong and weak scaling up to 8192 cores on GENCI's supercomputer Curie.

6.12. Kinetic modeling and simulation of edge tokamak plasmas and plasma-wall interactions

Participants: Sever Hirstoaga, David Coulette, Giovanni Manfredi.

We performed a full parallelization (over species and using 4D domain decomposition) of the 1D3V Multi-species Vlasov-Poisson finite-volumes code. The 4D code was then used to perform, by means of parametric studies, an analysis of the structure of the multi-scale boundary layer (the so-called Debye sheath and various pre-sheaths) for a magnetized-plasma in contact with an absorbing wall. This study allowed us to show, notably, that when the strong confining magnetic field is close to grazing incidence with respect to the absorbing surface, the boundary layer extends further into the plasma and as a result the magnitude of the electric field is lessened.
A second study was devoted to the dynamics of the propagation of the so-called "ELMs" (Edge-Localized-Modes) at the edge of Tokamak devices. The $1D1V$ collisionless model used in previous studies was extended by coupling a $1D1V$ kinetic model for the fast parallel propagation of the plasma disturbance along magnetic field lines with a fluid model in the directions perpendicular to the magnetic field. The coupling occurs by means of collision operators allowing for energy transfer between the parallel and perpendicular degrees of freedom. The initial asymptotic preserving scheme was extended to allow for adaptive time-stepping due to the introduction of the fluid transport equation. Using simulation results for various realistic collision rates we showed that collisional isotropization of the electronic population have a significant impact on the heat flux impacting the devices wall. The results were presented to the magnetic fusion community at the EPS conference in Leuven [9].

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, Elsa Rousseau, Nicolas Bajeux, Bapan Ghosh.

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [81]. 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, or through the introduction of biological control agents in deterministic [15] or stochastic [43], [33] fashion. 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 dynamics of plant pathogens [24].

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 and implemented on various models [67]. 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 [31]. This work is part of Stefano Casagranda's ongoing PhD thesis 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 procedure based on sensitivity analysis, to select the parameters to be estimated, to define their ranges and to set the values of the remaining parameters [72]. 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 parameter sets compatible with the data. So we based our fitting criterion on viral indicators rather than the whole viremia dynamics and we defined realistic data-based ranges for these indicators. We used a genetic algorithm for the minimisation. This ongoing work is part of Natacha Go's post-doctorate, supported by the MIHMES project, in collaboration with the Roslin Institute, Edinburgh, UK. It benefits from the resources and support of NEF computation cluster.

Parameter identification in compartmental systems. In collaboration with F. Dayan (R&D Manager, Dassault Systèmes), we work on practical problems of identifiability of parameters in linear pharmacokinetic models.

7.1.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Olivier Bernard, Valentina Baldazzi, Stefano Casagranda, Francis Mairet, Sofia Almeida, Claudia Lopez Zazueta, Lucie Chambon, Ivan Egorov.

7.1.2.1. Hybrid models analysis

Attractor computation using interconnected Boolean networks Following the work in [94] and [68], 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 [29]. This orbit is unique under some conditions on the parameters.

Piecewise linear representation of genetic regulatory networks The main goal was to develop a methodology for constructing piecewise linear and discrete models from a continuous model: given an initial partition of the state space, or grid, a piecewise constant vector field and diagram of transitions were computed based on the original ODE in the grid (M2 thesis of C. Kozia).

7.1.2.2. Continuous models analysis

A reduced model for the mammalian cell cycle This work focuses on identifying and analysing the main mechanisms underlying the cell cycle. A reduced two-dimensional model was proposed and calibrated against experimental data on cyclin B. As a validation, 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 in collaboration with F. Delaunay (and part of the PhD thesis of Sofia Almeida) has been submitted to a journal.

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. 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. We also study other models of the global cellular machinery. This is part of the PhD thesis of Stefano Casagranda, and done in collaboration with Inria IBIS project-team, in particular with D. Ropers.

Reduction of metabolic networks. We develop a dynamical reduction reduction for metabolic networks through Elementary Flux Modes and Quasi Steady State Approximation. The aim is, in the spirit of [1], to obtain a system of lower dimensions, with some accumulative variables. This is part of the PhD thesis of Claudia Lopez Zazueta.

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) [23]. We developed different versions of the problem, and consider a new problem where the aim is to optimize the production of a product (ANR project Reset).

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 [74].

7.1.2.4. Slow-Fast analysis of metabolic models

Metabolic modelling 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 [16], and show that Drum is a reasonable approximation, provided that growth rate stays low.

7.2. Fields of applications

7.2.1. Bioenergy

7.2.1.1. Modelling microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Ghjuvan Grimaud, Quentin Béchet, David Demory, Anaïs Bacquet, Jean-Philippe Steyer, Francis Mairet.

Experimental developments

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae and support model development. These experiments have been carried out in the Lagrangian simulator, under constant or periodic light and temperature, varying the total amount of light dose in the day [11]. The response in terms of storage carbon (triglycerides and carbohydrates) has been measured and correlated to the environment fluctuations.

Other experiments were carried out to reproduce the light signal percept by a cell in a raceway pond [71], derived from hydrodynamical studies [79]. 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 [70]. Experiments with coloured light demonstrated that the growth rate results from the absorbed light, whatever its wavelength.

A new methodology to measure cell viability has been set up. This approach is very promising to distinguish between net and gross growth rate [66]. It was used in the models to assess the impact of temperature on growth and mortality [20], [30].

On top of this, we carried out pilot experiments with solar light. We tested the impact of coloured film mimicking possible photovoltaic material. The collected data were used to calibrate models integrating the light spectrum in Ambre Veisseix's master thesis.

These works have been carried out in collaboration with A. Talec, S. Rabouille, and E. Pruvost (CNRS/UPMC -Oceanographic Laboratory of Villefranche-sur-Mer LOV).

Metabolism of carbon storage and lipid production

A macroscopic model for lipid production by oleaginous microalgae [7] has been previously proposed. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon [57], [56][16]. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*. A model was developed to represent heterotrophic growth on a mixture of acetate and butyrate [95]. A metabolic model was set up, on the basis of the DRUM framework [1], in order to simulate autotophic, heterotropic and mixotrophic growth, and to determine how to reduce substrate inhibition. The model was extended for other substrates such as glucose or glycerol in Anais Bacquet's master thesis.

Modelling the coupling between hydrodynamics and biology

The evolution of the biomass of microalgae in a raceway may be analyzed through an advection-diffusion-reaction Partial Differential Equations (PDE). First, the advection part corresponds to the transportation of the biomass through the raceway. Second, the diffusion coefficient allows to consider a Brownian motion for each particular trajectory of the particle. Finally, the reaction term corresponds to the biological dynamics. The optimization of the raceway was carried out by a vertical discretization of the raceway and an adjoint-based approach. In a similar way, the shape optimization was considered with the steady solutions of the Saint-Venant equations.

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 [63]. 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.

Modelling the photosynthesis response to fast fluctuating light

The impact of hydrodynamics on the light perceived by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [78]. The light signals that a cell experiences at the Lagrangian scale, depending on the fluid velocity, were then estimated. A Droop-Han model was used to assess the impact of light fluctuation on photosynthesis. A new model accounting for photoacclimation was also proposed [28]. Single cell trajectories were simulated, and the effect on photosynthesis efficiency was assessed using models of photosynthesis. These results were compared to experimental measurements where the high frequency light was reproduced.

Modelling 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. A simpler model was developed and used to identify the optimal working mode of a process based on photosynthetic biofilm growing on a conveyer belt, in Jérome Grenier's internship.

Modeling microalgae production processes

The integration of different models developed within BIOCORE [61], [65], [7] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems, on the basis of the dynamical model developed to describe microalgal growth under light and nitrogen limitations.

Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [36]. This model is used to predict lipid production in the perspective of large scale biofuel production [61].

In the framework of the ANR project Purple Sun, we developed a thermic 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).
Modelling thermal adaptation in microalgae

An extended statistical analysis was carried out on a database representing the temperature response of more than 200 microalgal species [12]. First the model proposed by [62] turned out to properly reproduce the temperature response. A model was then extracted to predict the observed link between the cardinal temperatures.

We have used Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. For a constant temperature, we have shown that the optimal temperature trait tends to equal the environment temperature [12]. We now use this method at the scale of the global ocean, validating our approach with experimental data sets from 194 species [75], [76].

Modelling 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.

7.2.1.2. Control and Optimization of microalgae production

On-line monitoring

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [73], [86]. Interval observers were designed for the estimation of the microalgae growth and lipid production within a production process [61] and validated experimentally [83].

Optimization of the bioenergy production systems

Based on simple microalgae models, analytical optimization strategies were proposed. We assessed strategies for optimal operation in continuous mode using the detailed model for raceways [88]. We first solved numerically an optimal control problem on a finite time horizon. Then, we re-analysed the optimization problem and derived a simplified sub-optimal strategy. These approaches were extended to outdoor cultivation, considering a possible variable culture depth. Assuming known weather forecasts considerably improved the control efficiency [21].

We also propose a nonlinear adaptive controller for light-limited microalgae culture, which regulates the light absorption factor (defined by the ratio between the incident light and the light at the bottom of the reactor).

Interactions between species

We had formerly proposed an adaptive controller which regulates the light at the bottom of the reactor [84]. 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 [64][39]. 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 proposed. First we modelled the adaptive dynamics for a population submitted to a variable temperature [12]. 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 [64].
Finally, in a more theoretical framework, we studied how to select as fast as possible a given species in a chemostat with two species at the initial instant. Using the Pontryagin maximum principle, we have shown that the optimal strategy is to maintain the substrate concentration to the value maximizing the difference between the growth rates of two species [58]. We now try to extend this result for n species with mutations.

7.2.2. Biological depollution

7.2.2.1. Control and optimization of bioprocesses for depollution

Participants: Olivier Bernard, Francis Mairet, Jean-Luc Gouzé.

We have considered the problem of global stabilization of an unstable bioreactor model (e.g. for anaerobic digestion), when the measurements are discrete and in finite number ("quantized"). These measurements define regions in the state space, wherein a constant dilution rate is applied We show that this quantized control may lead to global stabilization: trajectories have to follow some transitions between the regions, until the final region where they converge toward the reference equilibrium [82].

Although bioprocesses involve an important biodiversity, the design of bioprocess control laws are generally based on single-species models. In [26], we have proposed to define and study the multispecies robustness of bioprocess control laws: given a control law designed for one species, what happens when two or more species are present? We have illustrated our approach with a control law which regulates substrate concentration using measurement of growth activity. Depending on the properties of the additional species, the control law can lead to the correct objective, but also to an undesired monospecies equilibrium point, coexistence, or even a failure point. 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, Francis Mairet.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Phycover project is aiming at evaluating the potential of this process [93], [92].

We have proposed and analysed a three dimensional model which represent the coupling of a culture of microalgae limited by light and an anaerobic digester. We first prove the existence and attraction of periodic solutions. Applying Pontryagin's Maximum Principle, we have characterized optimal controls, including the computation of singular controls, in order to maximize methane production. Finally, we have determined numerically optimal trajectories by direct and indirect methods [59].

7.2.2.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer, Marjorie Alejandra Morales Arancibia.

In the sequel of the pioneering life cycle assessment (LCA) work of [80], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable.
The improvements due to technological breakthrough (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricity is produced on the plant [90]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage. This motivated the design of the purple sun ANR-project where electricity is produced by semi transparent photovoltaic panels [60] under which microalgae are growing. The LCA of such innovative processes where microalgae are grown under greenhouses has been carried out.

Finally, some work are aiming at normalising LCA for microalgae and proposing guidelines to make the LCA more easily comparable [69].

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 pests

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Nicolas Bajeux.

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 [6] [89], 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 (PhD of Nicolas Bajeux, [15]). Current research aims to understand the influence of different forms of stochasticity in the introduction process or the population dynamics on the efficacy of the introduction program [43], [33]. This last part of N. Bajeux's PhD si performed in collaboration with Vincent Calcagno (ISA).

Characteristics of space and the behavior and population dynamics of parasitoids

We tested the influence of the spatial heterogeneity of resource (hosts) distribution on the movements and fitness of individual parasitoids on a laboratory and a wild strain of the same species of *Trichogramma.* We showed that the level of resource aggregation has not the same influence on the different strains of the parasitoid, pointing out a behavioral adaptation of the laboratory strain [44]. This work is part of Victor Burte PhD Thesis (ISA, 2015-) and is done in close collaboration with V. Calcagno (ISA).

Connected research on the influence of space on the establishment of biological control agents is also being pursued both through computer simulations and laboratory experiments on *Trichogramma* [50]. This was the topic of the PhD thesis of Thibaut Morel Journel (ISA, defended in December 2015) [87] and is the present topic of Marjorie Haond (ISA, 2015-). In particular, we showed both theoretically and experimentally how landscape connectivity [27] or habitat richness [45], [37], [46] 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).

7.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Elsa Rousseau.

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 also 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 [41], [42]. A stochastic model was also produced to simulate the effect of drift on the virus epidemiological dynamics and on the durability of qualitative resistances [32], [40]. This was the topic of Elsa Rousseau's PhD thesis [14], and 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 [47]. This research is the main topic of Samuel Nilusmas PhD thesis (ISA, 2016-).

We extended the epidemiological model describing the phoma stem canker of oilseed rape, which aims at assessing the durability of crop resistance in the field and design efficient deployment strategies. We introduced a spatial structure based on real landscapes, as well as plant rotation strategies based on surveys conducted among farmers and cooperatives. We also performed a sensitivity analysis, to guide the model calibration. This ongoing work is part of (i) the K-Masstec project, which also incorporates experimental and field studies in collaboration with BIOGER (INRA Grignon); (ii) the GESTER project, with close collaborations with various INRA partners. It benefits from the resources and support of NEF computation cluster.

Eco-evolutionary dynamics of plant pathogens in seasonal environments

Understanding better pathogen evolution also requires to understand how closely related plant parasites may coexist. Such coexistence is widespread and is hardly explained through resource specialization. We showed that, in agricultural systems in temperate environments, the seasonal character of agrosystems is an important force promoting evolutionary diversification of plant pathogens [77]. The plant parasites reproduction mode may also strongly interact with seasonality. In this context, we investigated the special case of oak powdery mildew, an oak disease which is actually caused by a complex of two different species, combining original plant epidemic data with the semi-discrete seasonal plant epidemic model we introduced a few years ago [24]. This work has been done in collaboration with Frédéric Hamelin (Agrocampus Ouest), Marie Laure Desprez Loustau and Frederic Fabre (INRA Bordeaux).

7.2.3.2.1. Optimality/games in population dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Ivan Egorov.

Optimal resource allocation

Mycelium growth and sporulation is considered for phytopathogenic fungi. For biotrophic fungi, a flow of resource is uptaken by the fungus without killing its host; in that case, the life history traits (latence-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, and several spores in competition through a dynamic game. This work, in the framework of the ANR Funfit project, is done with Fabien Halkett of INRA Nancy.

Optimal foraging and residence times variations

We also investigated the problem in foraging theory of evaluating the expected harvest of an animal when conspecifics may arrive on the same patch of resource in a stochastic fashion, specificaly according to a Poisson process or a Bernoulli process [18].

With Marc Deschamps, similar questions were studied in theoretical economy in the context of a Cournot competition on a single market [17].

CARMEN Project-Team

7. New Results

7.1. Convergence analysis of a bidomain-bath model

M. Bendahmane and N. Chamakuri performed a convergence analysis for optimal control of a bidomain-bath model by using a finite-element scheme. The bidomain-bath model represents a commonly used experimental setup where a small piece of cardiac tissue is kept alive and studied for some time in a nutrient bath. The bidomain-bath model equations describe the cardiac bioelectric activity in the tissue and bath volumes where the control acts at the boundary of the tissue. The existence of the finite element scheme and convergence to a unique weak solution of the direct problem were established. The convergence proof was based on deriving a series of a-priori estimates and using a general L2-compactness criterion. Moreover, the well-posedness of the adjoint problem and the first order necessary optimality conditions were shown. Comparing to the direct problem, the convergence proof of the adjoint problem is based on using a general L1-compactness criterion. The model was used for a simulation of low-energy defibrillation.

7.2. An exponential Adams–Bashforth ODE solver for stiff problems

C. Douanla Lontsi, together with Y. Coudière and C. Pierre, obtained an important result on time integration of stiff differential problems. They considered Adams exponential integrators with general varying stabilizers. General stabilization brings flexibility and facilitates the integration of ODE systems and semilinear evolution PDEs coupled with ODE systems. They were able to prove the stability and convergence of this type of integrator by introducing a new framework that extends multistep linear methods. Dahlquist stability was numerically investigated. A(α)-stability was observed under a condition on the stabilizer, which is a singular property for explicit schemes. The method was numerically studied for two stiff models in electrophysiology. Its performance was compared with several classical methods. The authors concluded that for stiff ODE systems, it provides a cheaper way to compute accurate solutions at large time steps than implicit solvers.

7.3. Homogeneous Neumann condition on the torso for solving inverse problems

The electrical activity of the heart creates an electrical field in the body. This phenomenon is classically modelled in a quasistatic manner by Laplace's equation. The non-invasive electrocardiographic imaging (ECGI) problem consists in retrieving the best electrical map on the heart from given torso measurements. Classically, the solution is found as the best fit between data generated by a forward problem and the actual torso measurements, and it needs a regularization. Hence the inverse solution depends on the matrix of the forward problem, called the transfer matrix, and the choice of the regularization procedure. In 2006, a meshless method based on the method of fundamental solutions (MFS) was adapted by Y. Wang and Y. Rudy [54] to directly solve the inverse problem, combined with a 0-th order Tikhonov regularization. The MFS method is notably more robust than previous methods (e.g. BEM) to the uncertainties introduced by the segmentation of the geometries. In the MFS, the potential is expressed as summation of the fundamental solution of the Laplace equation over a discrete set of virtual point sources placed outside of the domain of interest. The inverse solution is searched as the set of sources that best fit the boundary conditions on the torso, up to the regularization term. This formulation yields a linear system, which matrix depends on the torso and heart geometries, and the boundary conditions at the torso surface. The regularization parameter also heavily depends on the properties of the transfer matrix. The boundary conditions considered in [54] are: i) the Dirichlet conditions, meaning that the potentials at the torso surface are fitted to the recorded ones, ii) homogeneous Neumann conditions (HNC) meaning that the normal flux of current is minimized.

Numerically, the HNC requires to build accurate directions at each measurement location of the body surface, which is a first difficulty. In addition, the body is cut at the top and the bottom where no-flux conditions are probably not relevant. Lastly, the matrix coefficients related to the HNC appears to be much smaller than the ones from the Dirichlet condition, due to the distance between the torso and the actual electrical source (the heart).

J. Chamorro-Servent, Y. Coudière and R. Dubois studied the effect of the HNC on the matrix. They showed that enforcing the Neumann condition has a negligible effect on the solution of the inverse problem. Reconstructed potentials and activation time maps were built for in-silico data. No major differences were found between the standard MFS and the MFS removing the HNC in terms of potentials and activation times. In addition, removing the HNC reduces the ill-conditioning of the problem and the computational burden: the normal at the torso surface is not required anymore, and the problem size is divided by 2. The results of this work were presented as a poster in CinC 2016, and collected in a proceeding for the same conference by J. Chamorro-Servent et al. [18].

7.4. Adaptive placement of the pseudo-boundaries improves the conditioning of the inverse problem

In order to complete the investigaton concerning the MFS technique from [54], J. Chamorro-Servent, Y. Coudière and R. Dubois also studied the effect of the location of the virtual sources of the MFS method on the solutions of the inverse problem. Specifically, the regularization term spoils the biophysical content of the solution, and the regularization parameter must be chosen as small as possible. But the problem must be regularized enough to overcome its sensitivity to: i) noise on the measured potentials, ii) uncertainty in the location of measurement sites with respect to the surface on which the sources are distributed, iii) errors of segmentation of the geometries, iv) influence of cardiac motion, etc.

The regularization parameter can be studied in view of the singular values of the matrix, or for given measurments, the discrete Picart condition as defined by Hansen [47].

In the MFS problem, explained in section 7.3 , the virtual sources are placed by inflating and deflating the heart and torso surfaces with respect to the heart's geometric center. However, for some heart-torso geometries, this geometrical center is a poor reference. Furthermore, it has been proved in other fields that the placement of the virtual sources influences the ill-posedness of the MFS problem. However, this has not been tested for the ECGI problem.

J. Chamorro-Servent, R. Dubois and Y. Coudière proposed a new method of placement of these virtual sources based on the minimal distance of each point considered on the heart surface to the torso electrodes. The singular value analysis and the discrete Picard condition were used to optimize the location of these sources. The new distribution of sources was compared with the standard one for a set of experimental data. These data consist of simultaneous acquisition of the cardiac (on a Langendorff perfusion of the heart) and body surface potentials, in a controlled experimental environment.

The results presented by J. Chamorro-Servent et al. at CinC2016 [24] showed that the new distribution of sources made the inverse problem less ill-posed and therefore, less sensitive to the regularization parameter chosen. This improved the reconstructed potentials on the heart surface, especially when artefact (as for example the baseline) or noise were present.

Further results from the combination of the works described here and in section 7.3 were presented in a poster in the Liryc workshop of October 2016 [33] by J. Chamorro-Servent et al. A journal manuscript is currently under preparation (to submit in 2017).

7.5. Reduced sodium current in the lateral ventricular wall induces J waves in the ECG

"J waves," a particular abnormal waveform in electrocardiogram (ECG) leads, are associated with a higher risk for ventricular fibrillation. M. Potse has performed a series of simulations to investigate three possible

mechanisms that could explain such waves and the associated arrhythmia risk. Out of these, a reduced sodium current in the lateral area of the left ventricular wall turned out to be the most powerful to cause J waves. The lateral area is particular because it is normally late activated, and a further delay due to regionally reduced sodium current can lead to J waves in the ECG. If the same occurs elsewhere in the heart, the resulting J waves would be masked by other ECG peaks. The simulations were supported, as far as possible, by experiments performed at the University of Amsterdam. The results have been published in the journal Frontiers in Physiology, and further refinements have recently been shown in a poster at the Annual workshop of IHU Liryc [14], [43].

7.6. Atrial fibrillation due to complex geometry

Atrial fibrillation (AF), a situation in which the electrical activation of the atria proceeds chaotically, is believed to be due to abnormal tissue structure (for example fibrosis), which slows propagation, and abnormalities in ionic currents, which make the action potential shorter. In collaboration with the Center for Computational Medicine in Cardiology in Lugano, Switzerland, we performed series of simulations in which we tried to reproduce these effects [20]. Rapid stimulation of the atria caused AF in some of the simulations, with a likelihood related to the severity of fibrosis. However, we also observed a 30 % likelihood of AF initiation in a model with no fibrosis at all. In these cases, the complex structure of our highly realistic models alone in combination with the rapid-pacing protocol sufficed to create situations of conditional propagation block, which led to a reentrant arrhythmia. These results may shed a new light on the course of new-onset AF. A manuscript on this topic is under preparation.

DRACULA Project-Team

6. New Results

6.1. Mathematical modeling of memory CD8 T cell ontogeny and quantitative predictions

Primary immune responses generate both short-term effector and long-term protective memory cells from naive CD8 T cells. The delineation of the genealogy linking those cell types has been complicated by the lack of molecular markers allowing to discriminate effector from memory cells at the peak of the response. Coupling transcriptomics and phenotypic analyses, and in collaboration with immunologists from Lyon (Jacqueline Marvel's team, Centre International de Recherche en Infectiologie), we identified a novel marker combination that allows to track nascent memory cells within the effector phase [13]. We then used mathematical models based upon our previous description of the dynamics of T cell immune response ([35], [45]) to investigate potential differentiation pathways. We thereby could describe the dynamics of population-size evolutions to test potential progeny links and we could demonstrate that most cells follow a linear naive-early effector-late effector-memory pathway. Of interest for vaccine design, our mathematical model also allows long-term prediction of memory cell numbers from early experimental measurements. Alltogether, 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 (vaccines) in terms of memory cell numbers generated.

6.2. Multiscale model of the CD8 T cell immune response

We presented in [43] the first multiscale model of CD8 T cell activation in a lymph node. We now described in [14] an update of this modeling approach. CD8 T cell dynamics are described using a cellular Potts model (hence cells are discrete interacting objects), whereas intracellular regulation is associated with a continuous system of nonlinear ordinary differential equations. We focused our study on describing the role of Interleukin 2 (IL2) secretion. One major result was the demonstration of the full relevance of a bona fide multiscale description of the process: the observed (all or none) emergent behavior at the cell population scale could not have been straightforwardly deduced from the simple examination of (seemingly tenuous) differences in the cellular or molecular levels in separation.

6.3. Moving the Boundaries of Granulopoiesis Modelling

The human blood cell production system usually remains extremely robust, in terms of cell number or function, with little signs of decline in old age. To achieve robustness, circulating blood cells rely on a formidable production machinery, the hematopoietic system, located in the bone marrow. All circulating blood cells-red blood cells, white blood cells and platelets-are renewed on a daily basis. The hematopoietic system produces an estimated $1e12$ cells per day. This is a significant fraction of the $3.7e13$ cells in an adult. Robustness is partly due to the short time scales at which cell populations are able to return to equilibrium, combined with large cell numbers and renewal rates. White blood cells (WBCs), among which neutrophils are most prevalent, are the body's first line, innate immune system. Upon infection, WBCs are mobilized from the bone marrow, to increase their number in circulation and fight off pathogen within hours. The 26 billion circulating neutrophils in human have a mean residence time of only 11h in the blood. After their release from the bone marrow, they quickly disappear in the peripheral tissues and are destroyed in the spleen, liver and bone marrow. In addition to the high renewal rate of circulating blood cells, a large number of mature neutrophils, ten times or more the circulating number, is kept in a bone marrow reserve, ready for entering circulation. This high renewal rate and mobilization capability, however, come at a cost. The blood system is an easy target for chemotherapeutic drugs, whose main way of acting is by killing proliferating

cells. White blood cells and end especially neutrophils, with their fast turnover, are particularly vulnerable to chemotherapy. Chemotherapy can induce neutropenia-a state of low absolute neutrophil count (ANC)-in cancer patients, which puts them at risk of infection. Homeostatic regulation of white blood cells is mainly controlled by the cytokine Granulocye-Colony Stimulating Factor (G-CSF). G-CSF promotes survival of white blood cell precursors and their differentiation into mature cells. The identification of this protein in the 1980's, and the subsequent development of human recombinant forms of G-CSF paved the way to the treatment of chemotherapy-induced neutropenia. G-CSF therapy as also been successful at treating congenital and other forms of neutropenia. Today, G-CSF is used as an adjuvant in several anti-cancer treatment protocols. The aim of the adjuvant therapy is to minimize the length of the neutropenic episodes. However, exogenous G-CSF administration interferes with white blood cell production regulation. What should be a straightforward effect–administer G-CSF to cause the ANC to increase–turns to be more complicated than that. For instance, it was observed that early timing of G-CSF administration could lead to prolonged neutropenic phase. Thus, in order to take advantage of the full potential of G-CSF, a detailed understanding of the physiological interaction between neutrophils and exogenous G-CSF is necessary. In this issue of the Bulletin (see [7]), Craig and colleagues present a physiological model of neutrophil production that includes a detailed modelling of the kinetics of G-CSF.

6.4. Bone marrow infiltration by multiple myeloma causes anemia by reversible disruption of erythropoiesis

Multiple myeloma (MM) infiltrates bone marrow and causes anemia by disrupting erythropoiesis, but the effects of marrow infiltration on anemia are difficult to quantify. Marrow biopsies of newly diagnosed MM patients were analyzed before and after four 28-day cycles of nonerythrotoxic remission induction chemotherapy. Complete blood cell counts and serum paraprotein concentrations were measured at diagnosis and before each chemotherapy cycle. At diagnosis, marrow area infiltrated by myeloma correlated negatively with hemoglobin, erythrocytes, and marrow erythroid cells. After successful chemotherapy, patients with less than 30% myeloma infiltration at diagnosis had no change in these parameters, whereas patients with more than 30% myeloma infiltration at diagnosis increased all three parameters. Clinical data were used to develop mathematical models of the effects of myeloma infiltration on the marrow niches of terminal erythropoiesis, the erythroblastic islands (EBIs) (see [12]). A hybrid discrete-continuous model of erythropoiesis based on EBI structure/function was extended to sections of marrow containing multiple EBIs. In the model, myeloma cells can kill erythroid cells by physically destroying EBIs and by producing proapoptotic cytokines. Following chemotherapy, changes in serum paraproteins as measures of myeloma cells and changes in erythrocyte numbers as measures of marrow erythroid cells allowed modeling of myeloma cell death and erythroid cell recovery, respectively. Simulations of marrow infiltration by myeloma and treatment with nonerythrotoxic chemotherapy demonstrate that myeloma-mediated destruction and subsequent reestablishment of EBIs and expansion of erythroid cell populations in EBIs following chemotherapy provide explanations for anemia development and its therapy-mediated recovery in MM patients.

6.5. Mathematical modelling of hematopoiesis dynamics with growth factor-dependent coefficients

In [4] and [5], we propose and analyze an age-structured partial differential model for hematopoietic stem cell dynamics, in which proliferation, differentiation and apoptosis are regulated by growth factor concentrations. By integrating the age-structured system over the age and using the characteristics method, we reduce it to a delay differential system (with discret delay [4] and distribute delay [5]). We investigate the existence and stability of the steady states of the reduced delay differential system. By constructing a Lyapunov function, the trivial steady state, describing cell's dying out, is proven to be globally asymptotically stable when it is the only equilibrium of the system. The asymptotic stability of the positive steady state, the most biologically meaningful one, is analyzed using the characteristic equation. This study may be helpful in understanding the uncontrolled proliferation of blood cells in some hematological disorders. This study may be helpful in understanding the behavior of hematopoietic cells in some hematological disorders.

6.6. Mathematical modelling of Chronic Myleoid Leukemia (CML)

Firstly, an analysis of a reduced version of our model has been performed by A. Besse et al. (manuscript in revision). It allows to analyze the structure of the steady states and their stability. Typically, the situation is as follows. There are 4 steady states: 0 (unstable) a low one (stable) an intermediate (unstable) and a high (stable).

Secondly, considering another framework of modelling [37], it was observed by A. Besse et al. (see also the thesis of A. Besse) that, under the assumptions of the models, the long term response might be non monotonous with respect to the dose. In words, when the disease load has been reduced enough, it might be more efficient (it is not a question of toxicity) to reduce the dose. This comes from a balance between quiescence induction and apoptosis effects of the drug.

6.7. Hybrid Modelling in Biology

The paper [19] presents a general review on hybrid modelling which is about to become ubiquitous in biological and medical modelling. Hybrid modelling is classically defined as the coupling of a continuous approach with a discrete one, in order to model a complex phenomenon that cannot be described in a standard homogeneous way mainly due to its inherent multiscale nature. In fact, hybrid modelling can be more than that since any types of coupled formalisms qualify as being hybrid. The paper [19], first presents the evolution and current context of this modelling approach. It then proposes a classification of the models through three different types that relate to the nature and level of coupling of the formalisms used.

6.8. Design and study of a new model describing the effect of radiotherapy on healthy cells

This new project started in January 2016 between a start up Neolys Diagnostics, an Inserm team from Lyon and some members of the Dracula team (Léon Matar Tine and Laurent Pujo-Menjouet) (see [11]). We recruited a student to start a PhD (Aurélien Canet) paid for one half by Neolys and the other half by the labex Milyon. The objective of this collaboration is to use deterministic models (as a first step) to describe the dynamics of ATM proteins in the cytoplasm moving to the nucleus. Once there, they recognize and repair damaged DNA (due to nuclear radiations) and to give solid mechanistic explanations of the phenomenological linear quadratic model used until now by biologists and clinicians. Next step is then to use data provided by the Inserm team to calibrate our model and use it for clinical tests by Neolys (to detect radiosensitive persons (3 different groups) and prevent individual from creating cancer induced by nuclear radiations).

6.9. Contribution to the interaction between Alzheimer's disease and prion with the analysis of a mathematical model arising from in vitro experiments

Alzheimer's disease (AD) is a fatal incurable disease leading to progressive neuron destruction. AD is caused in part by the accumulation of $A\beta$ monomers inside the brain, which have the faculty to aggregate into oligomers and fibrils. Oligomers are the most toxic structures as they can interact with neurons via membrane receptors, including PrPc proteins. This interaction leads to the misconformation of $PrPc$ into pathogenic oligomeric prions, $PrPol$. The objective of our collaboration with the Inra team lead by Human Rezaei (Jouy en Josas), is to design and study a brand new model describing in vitro $A\beta$ polymerization process (see [25]). We include interactions between oligomers and $PrPc$ that induces the misconformation of PrPc into PrPol. The model consists of nine equations, including size structured transport equations, ordinary differential equations and delayed differential equations. Our collaboration is only at its beginning and we applied for an ANR grant highlighting this interdisciplinary work.

6.10. Methods of Blood Flow Modelling

The paper [9] is devoted to recent developments in blood flow modelling. It begins with the discussion of blood rheology and its non-Newtonian properties. After that it presents some modelling methods where blood is considered as a heterogeneous fluid composed of plasma and blood cells. Namely, it describes the method of Dissipative Particle Dynamics and presents some results of blood flow modelling. The last part of this paper deals with one-dimensional global models of blood circulation. It explains the main ideas of this approach and presents some examples of its application.

6.11. Anomalous diffusion as an age-structured renewal process

Continuous-time random walks (CTRW) are one of the main mechanisms that are recurrently evoked to explain the emergence of subdiffusion in cells. CTRW were introduced fifty years ago as a generalisation of random walks, where the residence time (the time between two consecutive jumps) is a random variable. If the expectation of the residence time is defined, for instance when it is dirac-distributed or decays exponentially fast, one recovers "normal" Brownian motion. However, when the residence time expectation diverges, the CTRW describes a subdiffusive behavior. The classical approach to CTRW yields a non-Markovian (mean-field) transport equation, which is a serious obstacle when one wants to couple subdiffusion with (bio)chemical reaction. In [8], we took an alternative approach to CTRW that maintains the Markovian property of the transport equation at the price of a supplementary independent variable. We associate each random walker with an age a, that is the time elapsed since its last jump and describe the subdiffusive CTRW using an age-structured partial differential equations with age renewal upon each walker jump. In the spatially-homogeneous (zero-dimensional) case, we follow the evolution in time of the age distribution. An approach inspired by relative entropy techniques allows us to obtain quantitative explicit rates for the convergence of the age distribution to a self-similar profile, which corresponds to convergence to a stationary profile for the rescaled variables. An important difficulty arises from the fact that the equation in self-similar variables is not autonomous and we do not have a specific analytical solution. Therefore, in order to quantify the latter convergence, we estimate attraction to a time-dependent "pseudo-equilibrium", which in turn converges to the stationary profile.

6.12. Doubly nonlocal reaction-diffusion equations and the emergence of species

The paper [6] is devoted to a reaction-diffusion equation with doubly nonlocal nonlinearity arising in various applications in population dynamics. One of the integral terms corresponds to the nonlocal consumption of resources while another one describes reproduction with different phenotypes. Linear stability analysis of the homogeneous in space stationary solution is carried out. Existence of travelling waves is proved in the case of narrow kernels of the integrals. Periodic travelling waves are observed in numerical simulations. Existence of stationary solutions in the form of pulses is shown, and transition from periodic waves to pulses is studied. In the applications to the speciation theory, the results of this work signify that new species can emerge only if they do not have common offsprings. Thus, it is shown how Darwin's definition of species as groups of morphologically similar individuals is related to Mayr's definition as groups of individuals that can breed only among themselves.

6.13. Existence of very weak global solutions to cross diffusion models

The entropy structure has been used in [26] to derived a very general theorem for existence for cross diffusion models. The theory is based on the interplay between the entropy structure which gives some compactness in space (gradient control) and the duality strucutre identified by Michel Pierre for general parabolic systems, which gives integrability. We derive a very general results under very general structural hypothesis (exsitence of an entropy which is compatible with reaction terms and relevance of the duality strucutre). The key is the construction of implicit solutions of the semi discete version (time is discretized) which happens to verify all the strucutres and are very regular. Moreover, we give a simple condition for mutliple case (more than 3

species) for building examples with an entropy structure based on the detailed balance structure proposed in [34].

M3DISIM Project-Team

7. New Results

7.1. Mathematical and Mechanical modeling

7.1.1. A 3D contact-mechanics model of the heart and thorax for seismocardiography

Participants: Alexandre Laurin [correspondant], Sébastien Imperiale, Dominique Chapelle, Philippe Moireau.

The current interpretation of seismocardiogram fiducial points depends on their phenomenological association with the timing of events on simultaneous echocardiograms. Signal processing methods can be devised to acquire these timings automatically (see [21] and [22]). So far, no causal framework has been tested to explain this timing, nor their direction and amplitude. This work attempted to adapt a comprehensive 3D cardiac model to interact through contact with a model of the thoracic cage. The heart model was designed to represent multi-scale, multi-physics physiological processes such as the electrical activation, the mechanical contraction, as well as the system circulation. The objective is to link observed acceleration of the sternum to the underlying physiology, and offer a potential mechanical explanation for them. The modelling chain necessary to go from the heart model to a simulated SCG has been successfully implemented (see Figure 1). Furthermore, the complexity of the thoracic model has been substantially reduced, without deteriorating results, to improve the portability of the entire process. Once the relevant parameters of in-vivo thoraces will have been precisely identified, it will be possible to compute heart forces and the various cardiac events that cause them directly from SCG measurements. The subsequent aim is to apply the model to ageing and pathological physiologies.

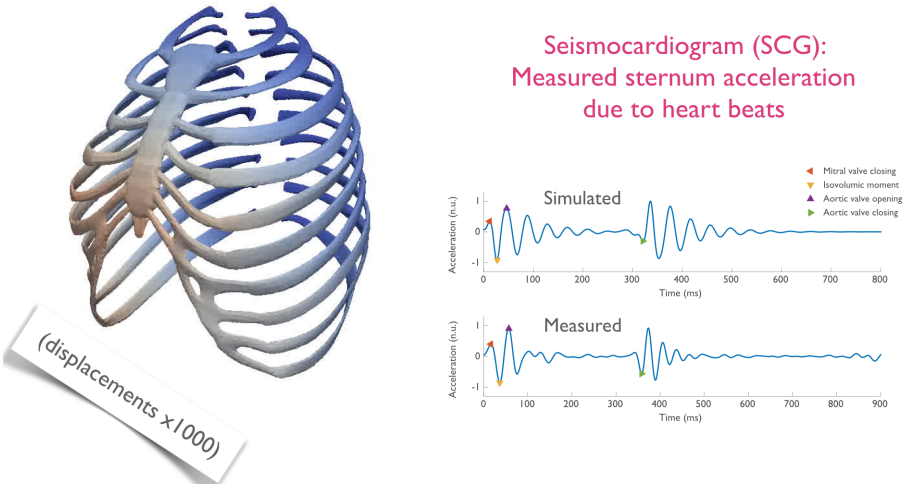


Figure 1. Simulation of the thorax deformation due to a heart beat and of the corresponding seismocardiogram

7.1.2. Multi-scale modeling of muscle contraction

Participants: François Kimmig [correspondant], Dominique Chapelle, Matthieu Caruel.

This works aims at proposing a multi-scale model of the muscular contraction that can be used in the context of cardiac simulation. The modeling will be based on the stochastic equations that describe muscular contraction at the molecular level. Asymptotic counterparts of the stochastic model will be considered in order to provide pertinent simplified models. The modeling elements will be confronted with experiments that will be performed on cardiac muscle cells by collaborators in the team of Professor Vincenzo Lombardi at the University of Florence.

In the framework of this collaboration, a chemicomechanical model is being implemented into CardiacLab, a simulation environment developed by the team. It will enrich the range of modeling tools of the team for the active contribution of muscle cells to the cardiac behavior.

7.1.3. Multiphysics and multiscale modelling, data-model fusion and integration of organ physiology in the clinic: ventricular cardiac mechanics

Participants: Radomir Chabiniok [correspondant], Philippe Moireau, Dominique Chapelle, Maxime Sermesant [Team Asclepios].

With heart and cardiovascular diseases continually challenging healthcare systems worldwide, translating basic research on cardiac (patho)physiology into clinical care is essential. Exacerbating this already extensive challenge is the complexity of the heart, relying on its hierarchical structure and function to maintain cardiovascular flow. Computational modelling has been proposed and actively pursued as a tool for accelerating research and translation. Allowing exploration of the relationships between physics, multiscale mechanisms and function, computational modelling provides a platform for improving our understanding of the heart. Further integration of experimental and clinical data through data assimilation and parameter estimation techniques is bringing computational models closer to use in routine clinical practice. This work published in [17] reviews developments in computational cardiac modelling and how their integration with medical imaging data is providing new pathways for translational cardiac modelling.

7.1.4. Eidolon: visualization and computational framework for multi-modal biomedical data analysis

Participant: Radomir Chabiniok [correspondant].

Biomedical research, combining multi-modal image and geometry data, presents unique challenges for data visualization, processing, and quantitative analysis. Medical imaging provides rich information, from anatomical to deformation, but extracting this to a coherent picture across image modalities with preserved quality is not trivial. Addressing these challenges and integrating visualization with image and quantitative analysis results in Eidolon, a platform which can adapt to rapidly changing research workflows. In the paper [26] we outline Eidolon, a software environment aimed at addressing these challenges, and discuss the novel integration of visualization and analysis components. These capabilities are demonstrated through the example of cardiac strain analysis, showing the Eidolon supports and enhances the workflow.

7.1.5. Mathematical and numerical modeling of elastic waves propagation in the heart

Participants: Federica Caforio [correspondant], Dominique Chapelle, Sébastien Imperiale.

The objective of this work is to develop a rigorous mathematical and numerical background for the extension and dissemination of elastography imaging modalities, applied to the cardiac setting. The problems treated concern the topics of mathematical modelling, numerical analysis and scientific computing. More precisely, the plan is to define a linearised model for the propagation of elastic waves in the heart, to study approximations of these models and define adapted numerical methods for the discretisation of the resulting partial differential equations.

7.2. Numerical Methods

7.2.1. Effective and energy-preserving time discretization for a general nonlinear poromechanical formulation

Participants: Bruno Burtschell, Dominique Chapelle [correspondant], Philippe Moireau.

In this work, we consider a general nonlinear poromechanical model, formulated based on fundamental thermodynamics principle, suitable for representing the coupling of rapid internal fluid flows with large deformations of the solid, and compatible with a wide class of constitutive behavior. The objective of the present work is to propose for this model a time discretization scheme of the partitioned type, to allow the use of existing time schemes - and possibly separate solvers - for each component of the model, i.e. for the fluid and the solid. To that purpose, we adapt and extend an earlier proposed approach devised for fluid-structure interaction in an Arbitrary Lagrangian-Eulerian framework. We then establish an energy estimate for the resulting time scheme, in a form that is consistent with the underlying energy principle in the poromechanical formulation, up to some numerical dissipation effects and some perturbations that we have carefully identified and assessed. In addition, we provide some numerical illustrations of our numerical strategy with test problems that present typical features of large strains and rapid fluid flows, and also a case of singular transition related to total drainage. An example of challenging application envisioned for this model and associated numerical coupling scheme concerns the perfusion of the heart. This work has resulted in the publication [15].

7.2.2. Delayed feedback control method for calculating space-time periodic solutions of viscoelastic problems

Participants: Ustim Khristenko, Patrick Le Tallec.

We are interested in fast techniques for calculating a periodic solution to viscoelastic evolution problems with a space-time periodic condition. In order to avoid the invertion of very large matrices, such a solution is often computed as an asymptotic limit of the initial value problem with arbitrary initial data. We have developed a control method, accelerating the convergence to the periodic state. The main idea is to modify our problem by introducing a feedback control term, based on a periodicity error.

First, an abstract evolution problem has been studied. From the analytic solution of the modified (controlled) problem, an efficient control has been constructed, optimizing the spectrum of the problem. The proposed control term can be mechanically interpreted, and its efficiency increases with the relaxation time.

In order to confirm numerically the theoretical results, a finite element simulation has been carried out on a full 3D model for a steady rolling of a viscoelastic tyre with periodic sculpture. It has demonstrated that the controlled solution converges indeed faster than the non-controlled one, and that the efficiency of the method increases with the problem's relaxation time, that is when the memory of the underlying problem is large.

7.2.3. Construction and analysis of an adapted spectral finite element method to convective acoustic equations

Participant: Sébastien Imperiale [correspondant].

This work addresses the construction of a non spurious mixed spectral finite element (FE) method to problems in the field of computational aeroacous-tics. Based on a computational scheme for the conservation equations of linear acoustics, the extension towards convected wave propagation is investigated. In aeroacoustic applications, the mean flow effects can have a significant impact on the generated sound field even for smaller Mach numbers. For those convective terms, the initial spectral FE discretization leads to non-physical, spurious solutions. Therefore, a regularization procedure is proposed and qualitatively investigated by means of discrete eigenvalues analysis of the discrete operator in space. A study of convergence and an application of the proposed scheme to simulate the flow induced sound generation in the process of human phonation underlines stability and validity. This work has resulted in the publication [19].

7.2.4. Space/time convergence analysis of a class of conservative schemes for linear wave equations

Participants: Juliette Chabassier [MAGIQUE 3D team], Sébastien Imperiale [correspondant].

This work concerns the space/time convergence analysis of conservative two-steps time discretizations for linear wave equations. Explicit and implicit, second and fourth order schemes are considered, while the space discretization is given and satisfies minimal hypotheses. The convergence analysis is done using energy techniques and holds if the time step is upper-bounded by a quantity depending on space discretization parameters. In addition to showing the convergence for recently introduced fourth order schemes, the novelty of this work consists in the independency of the convergence estimates with respect to the difference between the time step and its greatest admissible value. This work has resulted in the publication [16].

7.3. Inverse Problems

7.3.1. Front observer for data assimilation of electroanatomical mapping data for a numerical atrial model

Participants: Antoine Gérard [Carmen team], Annabelle Collin [Monc team], Jason Bayer [Carmen team], Philippe Moireau [correspondant], Yves Coudière [Carmen team].

The purpose of our work is to personalize an atrial model of the propagation of the action potential, based on electrical catheter data with the help of the data assimilation approach introduced in [Collin & Al, Journal of Computational Physics 2015]. The originality of the approach is to indroduce a Luenberger observer of a surface atrial model of the propagation which can pursue - like in classical Kalman filtering approach - the actual activation front reconstructed from catheter data. Moreover, this approach may account for the breakthrough of new activation fronts at anytime with an additional topological gradient term. In the present work, we adapt this approach to the bilayer surface atrial model of th propagation of action potentials [Labarthe & Al, Europace 2014], and evaluated for the first time on a real patient's dataset. First, the model was geometrically fit to the patient's data. A fiber architecture was added to the geometry. Then an initial electrophysiological state was guessed, and the model was run with the Luenberger filter for some catheter data acquired during a CARTO procedure. All along the simulation, the filter corrects the action potential so as to track CARTO local activation times, while preserving a biophysical behavior. With this technique, we are able to reconstruct smooth activation maps over the whole atrial surfaces. This promising technique may also allow to reconstruct velocity fields and directions, phase map and possibly give information on repolarization. This work results from a collaborative project carried out during a training session at CEMRACS 2016 in Marseille, Luminy. This work has resulted in the publication [28].

7.3.2. Iterative observer-based state and parameter estimation for linear systems

Participant: Atte Aalto [correspondant].

In this work we propose an iterative method for joint state and parameter estimation using measurements on a finite time interval for systems that are backward output stabilizable. Since this time interval is fixed, errors in initial state may have a big impact on the parameter estimate. We propose to use the back and forth nudging (BFN) method for estimating the system's initial state and a Gauss–Newton step between BFN iterations for estimating the system parameters. Taking advantage of results on the optimality of the BFN method, we show that for systems with skew-adjoint generators, the initial state and parameter estimate minimizing an output error cost functional is an attractive fixed point for the proposed method. We treat both linear source estimation and bilinear parameter estimation problems.

7.3.3. Estimation from moments measurements for amyloid depolymerisation

Participants: Aurora Armiento [Mamba team], Marie Doumic [Mamba team], Philippe Moireau [correspondant].

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation, a key feature for amyloid diseases. This study aims at setting this framework problem when the experimental measurements consist in the time-dynamics of a moment of the population (*i.e.* for instance the total polymerised mass, as in Thioflavine T measurements, or the second moment measured by Static Light Scattering). We propose a general methodology, and we solve the problem theoretically and numerically in the case of a depolymerising system. We then apply our method to experimental data of degrading oligomers, and conclude that smaller aggregates of ovPrP protein should be more stable than larger ones. This has an important biological implication, since it is commonly admitted that small oligomers constitute the most cytotoxic species during prion misfolding process. This work has resulted in the publication [14].

7.3.4. Analysis of an observers strategy for initial state reconstruction in unbounded domains

Participants: Antoine Tonnoir [correspondant], Sonia Fliss [Poems team], Sébastien Imperiale, Philippe Moireau.

In this work, we are interested in the problem of recovering a compactly supported initial state of the wave equation in unbounded domain (such as the whole plane, a waveguide...). To this purpose, we assume that the velocity is known in a bounded observation region surrounding the support of the initial state. We consider an iterative algorithm of reconstruction based on back and forth nudging and prove the exponential convergence of this algorithm and its robustness with respect to noisy measures, at the continuous level. We also study the effect of the discretization process on the convergence of the algorithm.

7.4. Experiments and Clinical applications

7.4.1. Characterization of mechanical properties of soft tissues

Participants: Jean-Marc Allain [correspondant], Jean-Sebastien Affagard, Maeva Lopez Poncelas.

Soft tissues - such as skin - have complex mechanical properties: large strains, anisotropy, etc.. Identifying constitutive properties incorporating microstructure effects is very important for applications in medicine (surgery and other therapies) and industry (anti-ageing cosmetics, etc.). This characterization, however, requires complex experiments. We have developed a novel biaxial traction experimental method for mice skin, relying on a sensitivity analysis for determining optimal experimental parameters, including in particular sample size and most informative loading paths. This protocol has already been used on multiple samples, and 3 distinct constitutive laws of increasing complexity have been characterized (Master's internship of Maeva Lopez).

Another originality in our approach is to place our setup under a microscope to monitor the microstructure evolution during the test. These rich measurements allow detailed comparisons of classical models (such as Holzapfel's) with our data.

7.4.2. Non-invasive model-based assessment of passive left-ventricular myocardial stiffness in healthy subjects and in patients with non-ischemic dilated cardiomyopathy

Participant: Radomir Chabiniok [correspondant].

Patient-specific modelling has emerged as a tool for studying heart function, demonstrating the potential to provide non-invasive estimates of tissue passive stiffness. However, reliable use of model-derived stiffness requires sufficient model accuracy and unique estimation of model parameters. In this work we present personalised models of cardiac mechanics, focusing on improving model accuracy, while ensuring unique parametrisation. The influence of principal model uncertainties on accuracy and parameter identifiability was systematically assessed in a group of patients with dilated cardiomyopathy and healthy volunteers. For all cases, we examined three circumferentially symmetric fibre distributions and two epicardial boundary conditions. Our results demonstrated the ability of data-derived boundary conditions to improve model accuracy and highlighted the influence of the assumed fibre distribution on both model fidelity and stiffness estimates. The model personalisation pipeline – based strictly on non-invasive data – produced unique

parameter estimates and satisfactory model errors for all cases, supporting the selected model assumptions. The thorough analysis performed enabled the comparison of passive parameters between volunteers and dilated cardiomyopathy patients, illustrating elevated stiffness in diseased hearts.

7.4.3. Age-related changes in intraventricular kinetic energy: a physiological or pathological adaptation

Participant: Radomir Chabiniok [correspondant].

Aging has important deleterious effects on the cardiovascular system. In this work we sought to compare intraventricular kinetic energy (KE) in healthy subjects of varying ages with subjects with ventricular dysfunction to understand if changes in energetic momentum may predispose individuals to heart failure. Four-dimensional flow MRI was acquired in 35 healthy subjects (age: 1 -67 yr) and 10 patients with left ventricular (LV) dysfunction (age: 28-79 yr). Healthy subjects were divided into age quartiles (1st quartile: 16 yr, 2nd quartile: 17-32 yr, 3rd quartile: 33-48 yr, and 4th quartile: 49 - 64 yr). KE was measured in the LV throughout the cardiac cycle and indexed to ventricular volume. In healthy subjects, two large peaks corresponding to systole and early diastole occurred during the cardiac cycle. A third smaller peak was seen during late diastole in eight adults. Systolic KE (P 0.182) and ejection fraction (P 0.921) were preserved through all age groups. Older adults showed a lower early peak diastolic KE compared with children (P 0.0001) and young adults (P 0.025). Subjects with LV dysfunction had reduced ejection fraction (P 0.001) and compared with older healthy adults exhibited a similar early peak diastolic KE (P 0.142) but with the addition of an elevated KE in diastasis (P 0.029). In healthy individuals, peak diastolic KE progressively decreases with age, whereas systolic peaks remain constant. Peak diastolic KE in the oldest subjects is comparable to those with LV dysfunction. Unique age-related changes in ventricular diastolic energetics might be physiological or herald subclinical pathology. This work has resulted in the publication [24].

7.4.4. Patient-specific computational analysis of ventricular mechanics in pulmonary arterial hypertension

Participant: Martin Genet [correspondant].

Patient-specific biventricular computational models associated with a normal subject and a pulmonary arterial hypertension (PAH) patient were developed to investigate the disease effects on ventricular mechanics. These models were developed using geometry reconstructed from magnetic resonance (MR) images, and constitutive descriptors of passive and active mechanics in cardiac tissues. Model parameter values associated with ventricular mechanical properties and myofiber architecture were obtained by fitting the models with measured pressure–volume loops and circumferential strain calculated from MR images using a hyperelastic warping method. Results show that the peak right ventricle (RV) pressure was substantially higher in the PAH patient (65 mmHg versus 20 mmHg), who also has a significantly reduced ejection fraction (EF) in both ventricles (left ventricle (LV): 39% versus 66% and RV: 18% versus 64%). Peak systolic circumferential strain was comparatively lower in both the left ventricle (LV) and RV free wall (RVFW) of the PAH patient (LV: -6.8% versus -13.2% and RVFW: -2.1% versus -9.4%). Passive stiffness, contractility, and myofiber stress in the PAH patient were all found to be substantially increased in both ventricles, whereas septum wall in the PAH patient possessed a smaller curvature than that in the LV free wall. Simulations using the PAH model revealed an approximately linear relationship between the septum curvature and the transseptal pressure gradient at both early-diastole and end-systole. These findings suggest that PAH can induce LV remodeling, and septum curvature measurements may be useful in quantifying transseptal pressure gradient in PAH patients. This work has resulted in the publication [25].

MAMBA Project-Team

7. New Results

7.1. Cancer

Participants: Luis Lopes Neves de Almeida, Group Emmanuel Barillot [Institut Curie], Catherine Bonnet [DISCO team, Saclay], Thibault Bourgeron, Group Kai Breuhahn [Hospital of University of Heidelberg, Pathology], Rebecca Chisholm, Jean Clairambault, François Delhommeau [Haematology department, St Antoine Hospital, Paris], Marie Doumic, Dirk Drasdo, Walid Djema [DISCO team, Saclay], Julie Favre [EPFL, Lausanne], Olivier Fercoq [Télécom ParisTech], Ghassen Haddad [ENIT, Tunis], Shalla Hanson [Department of mathematics, Duke University, Durham, NC], Pierre Hirsch [Haematology department, St Antoine Hospital, Paris], Groups Invade, Lungsysii, Hicham Janati [ENSAE, Paris], Tim Johann, Group Klingmueller [German Cancer Center, Heidelberg], Michal Kowalczyk [Univ. Santiago de Chile], Annette Larsen [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Tommaso Lorenzi [University of St Andrews, Scotland], Alexander Lorz, Frédéric Mazenc [DISCO team, Saclay], Benoît Perthame, Camille Pouchol, Andrada Quillas Maran, Fernando Quirós [Univ. Autónoma de Madrid], Michèle Sabbah [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Min Tang [Jiaotong University, Shanghai], Teresa Teixeira [IBCP], Emmanuel Trélat [LJLL, UPMC], Paul Van Liedekerke, Oliver Sedlaczek [German Cancer Center (DKFZ) and Hospital of University of Heidelberg, Germany], François Vallette [INSERM, CRCNA, Nantes], Nicolas Vauchelet, Irène Vignon-Clementel [REO], Zhou Xu [IBCP], Yi Yin.

7.1.1. Senescence and telomere shortening

In many animals, aging tissues accumulate senescent cells, a process which underlies the loss of regeneration capacity of organs and is ultimately detrimental to the organism. Senescence is also required to protect organisms from unlimited proliferation that may arise from numerous stimuli or deregulations. Due to these opposing effects in aging and cancer, senescence is considered antagonistic pleiotropic; senescence is beneficial to protect from cancer in the young organism, but becomes detrimental late in life. Therefore, understanding the mechanisms of cellular senescence may lead to the development of global therapies to debilities specific for the aged, as well age-associated diseases and cancer. These are major public health issues in France, and other western aging countries.

Replicative senescence, induced by telomere shortening, exhibits considerable asynchrony and heterogeneity, the origins of which remain unclear. In [19], following [61], we formally study how telomere shortening mechanisms impact on senescence kinetics and define two regimes of senescence, depending on the initial telomere length variance. We provide analytical solutions to the model, highlighting a non-linear relationship between senescence onset and initial telomere length distribution. This study reveals the complexity of the collective behaviour of telomeres as they shorten, leading to senescence heterogeneity.

7.1.2. Stability analysis of a delay differential model of healthy and leukaemic haematopoiesis

The collaboration with the DISCO team (Inria Saclay, C. Bonnet, F. Mazenc and their PhD student W. Djema), supported by the collaboration with the team of haematologists led by F. Delhommeau at St. Antoine Hospital in Paris, has been continued, with common research work underway. A new model describing the coexistence between ordinary and mutated haematopoietic stem cells was introduced and analysed in [32]. Interpreting theoretical conditions found to guarantee the survival of healthy cells while eradicating unhealthy ones leads us to propose possibly innovative therapies obtained by combining the infusion of different drugs (Flt-3 inhibitors such as quizartinib, cytosine arabinoside, anthracyclines).

7.1.3. Interactions between tumour cell populations and their cellular micro-environments

This is the main object of study, together with the consideration of phenotype and genotype heterogeneity in cancer cell populations (see *Highlights of the Year*), of the *THE ITMO Cancer* national call 2016, to which two (out of three) of the submitted projects involving our team, which themselves were two out of the six laureate projects at the national level, have been successfully funded. The two projects, EcoAML and MoGlImaging have been launched in November 2016.

7.1.4. Evolution and cancer; drug resistance in cancer cell populations

We have continued to develop our phenotypically based models of drug-induced drug resistance in cancer cell populations, representing their Darwinian or Lamarckian evolution under drug pressure by integro-differential equations. The properties of phenotype-structured PDEs are explored in theoretical articles with examples [25], [78]. We will also use them in the two projects laureates to the THE (*Tumour Heterogeneity in its Ecosystem*) ITMO Cancer call of 2016 (see *Highlights of the Year*), EcoAML and MoGlImaging, to help predict early evolution towards leukaemogenesis (EcoAML, leader F. Delhommeau, St. Antoine Hospital, Paris) and emergence of resistance to temozolomide in glioblastoma cell populations (MoGlImaging, leader E. Moyal, Toulouse, F. Vallette, Nantes, being our main work correspondent). In this version, mutualistic exchanges between the cancer cell population and its supporting stroma will be represented as impinging on the phenotypic variables that describe the relevant heterogeneity at stake in the two cell populations.

With F. Vallette, we have co-supervised Hicham Janati's ENSAE 2nd year (M1) internship on the investigation of cancer resistance in a Glioblastoma cell line [46] with gene expression data coming from F. Vallette's lab in Nantes. This internship represents for us a first step in the quest for relevant (most likely multidimensional) phenotypes, based on bioinformatic and biostatistic methods to process experimental dynamic gene expression data, to interactively identify our physiologically structured models of heterogeneity and its evolution in cancer cell populations. The task ahead is immense, but our commitment in the THE consortium (see *Highlights of the Year*) with biologists providing us with such data (F. Vallette, F. Delhommeau) gives us good expectations to be successful with it in a close future. Following Hicham Janati's internship [46], Julie Favre (M1 student at EPFL) has been hired in a new internship to set the practical grounds for the interactive collaboration (begun with the THE program in November 2016) between our team and F. Delhommeau's team on model-based processing of gene expression data produced by a heterogeneous leukaemic cell population and by its surrounding stromal cell population.

The evolution towards drug-induced drug resistance in cancer cell populations may be described by methods of adaptive dynamics for continuous phenotype-structured populations, as such cell populations are fundamentally phenotypically, if not genetically, heterogeneous. In [11], [40], we review the bases of heterogeneity and drug resistance in cancer, its assessment by biological experiments and by mathematical modelling and methods of optimal control that may be applied to represent and optimise combined delivery of cytotoxic and cytostatic drugs, see below "Optimal control and drug resistance" [52].

7.1.5. Therapy optimisation

PK-PD: optimisation with respect to unwanted side effects. A previous pharmacokinetics-pharmacodynamics (PK-PD) model for the action of anticancer drugs at the molecular level, coupled with an age-structured linear model of the cell division cycle, has been updated in [12] (introduced in a special issue on PK-PD [9]) with optimisation of the combined delivery of 3 different drugs (5-fluorouracil, oxaliplatin, leucovorin). This is joint work with Olivier Fercoq, Télécom ParisTech. It represents the coalescence of two distinct types of models, both studied in previous years in our team: molecular ODE-based models of the action of anticancer drugs, and optimisation (using a Uzawa-like algorithm applied to the first eigenvalues of the two growing populations, minimising the cancer eigenvalue - objective - while maintaining the healthy eigenvalue above a reference threshold - constraint -, supposed to be linked to the state of health of the patient) of the control of linear growth models based on age-structured transport equations for the cell division cycle in the two populations separately.

Optimal control and drug resistance. In the framework of Camille Pouchol's PhD thesis, co-supervised at LJLL by E. Trélat and J. Clairambault, analysing the behaviour of healthy and cancer cell populations structured in a continuous resistance phenotype to a cytotoxic drug, and exposed to cytostatic and cytotoxic chemotherapies, we have firstly established, in an asymptotic analysis using a Lyapunov functional inspired from works by P.-E. Jabin and G. Raoul [77], results of convergence and concentration for constant drug concentrations [52] (following [84]). In a second part of this work, we have derived from them analytical conditions of optimality for the delivery of the drugs in a general class of controls. A numerical example of the optimal strategy is illustrated on Figure 1 , where the phenotype x continuously ranges from totally sensitive ($x = 0$) to totally resistant ($x = 1$), and healthy and cancer cells are represented by densities of cells $n_H(t, x)$, $n_C(t, x)$. The simulations confirm that the optimal strategy consists of letting the cancer cell population become more and more homogeneous around a sensitive phenotype, and then to use the maximal amount of drugs. This proposed strategy may be related with the "drug holiday" practiced in the clinic of cancers. We also show *en passant* the clearly detrimental effect of delivering cytotoxic drugs at high *constant* doses, as they inevitably induce the emergence of a thriving resistant subpopulation, which is illustrated on Figure 2 .

7.1.6. Lung and breast cancer

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumour treatment assessment, reflecting Brownian movement of water molecules in tissues. Since densely packed cells restrict molecule mobility, tumour tissues produce less attenuated DWI signals than normal tissues. However, no general quantitative relation between DWI data and the cell density has been established. In order to link low-resolution clinical cross-sectional data with high-resolution histological information, we have developed an image processing and analysis chain, which was used to study the correlation between the diffusion coefficient (D value) estimated from DWI and tumour cellularity from serial histological slides of a resected non-small cell lung cancer (NSCLC) tumour. Colour deconvolution followed by cell nuclei segmentation was performed on digitised histological images to determine local and cell-type specific 2D (two-dimensional) densities. From these the 3D (three-dimensional) cell density was inferred by a model-based sampling technique, which is necessary for the calculation of local and global 3D tumour cell count. Next, DWI sequence information was overlaid with high-resolution CT data and the resected histology using prominent anatomical hallmarks for co-registration of histology tissue blocks and non-invasive imaging modalities for data. The integration of cell numbers information and DWI data derived from different tumour areas revealed a clear negative correlation between cell density and D value. Importantly, spatial tumour density can be quantitatively calculated based on DWI data to estimate tumour heterogeneity [55].
In a followup we currently study to what extent the relation between cellularity and DWI - diffusion coefficient can be inferred from biopsies instead of tumour serial sections. Moreover, we are studying the relation between DWI and tumour microvasculature [33].

7.1.7. Biomechanically mediated growth control of cancer cells

Mechanical feedback has been identified as a key regulator of tissue growth, by which external signals are transduced into a complex intracellular molecular machinery. Using multiscale computational modeling of multicellular growth in two largely different experimental settings with the same tumour cell line we demonstrated that the cellular growth response on external mechanical stress may nevertheless be surprisingly quantitatively predictable. Our computational model represents each cell as an individual unit capable of migration, growth, division, and death and is parameterised by measurable biophysical and bio-kinetic parameters. A cell cycle progression function depending on cell compression was established by comparisons of computer simulations with experiments of spheroids growing in an alginate elastic capsule. After a calibration step with free growing spheroids growing in a liquid suspension to capture the different growth conditions, the model using the same cell cycle progression function can predict the mechanical stress response of spheroid growth in a completely different experimental technique using Dextran, where stress is exerted by osmotic pressure. Our findings suggest that the stress response of cell growth may be highly reproducible even in otherwise different environments. This encourages the idea that robust functional modules may

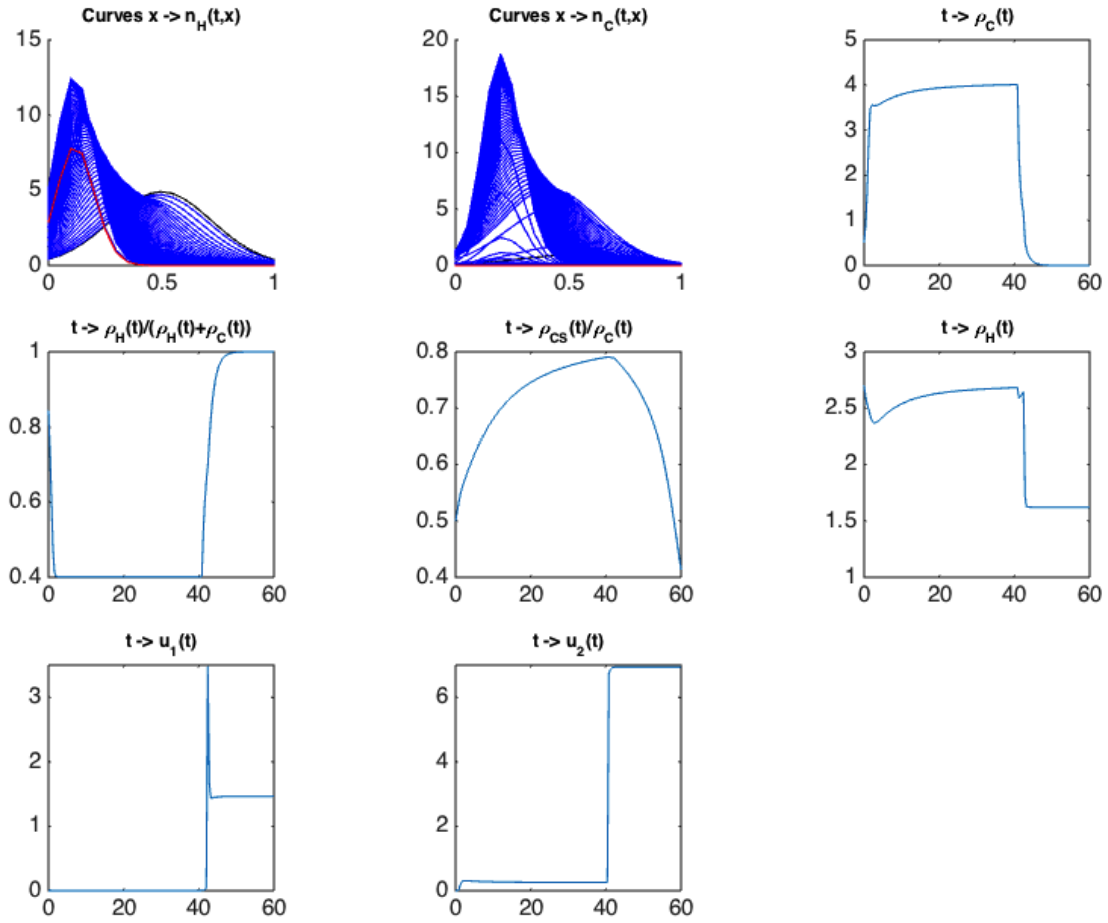


Figure 1. Simulation of the optimal control problem in time horizon $T = 60$. Top, left and middle: time evolution of $x \mapsto n_H(t,x)$, number of healthy cells with drug resistance expression phenotype x, and of $x \mapsto n_C(t,x)$, number of cancer cells with the same phenotype x. The initial conditions are in black, the final ones in red. Top right (resp., centre right): evolution with time of the total number of cancer cells $\rho_C(t) = \int_0^1 n_C(t,x)dx$ (resp., of healthy cells $\rho_H(t) = \int_0^1 n_H(t,x)dx$). Centre left (resp., centre middle), evolution with time of the ratio of healthy cells to total cells (resp., of sensitive cancer cells defined by the weighted integral $\rho_{CS}(t) = \int_0^1 (1-x)n_C(t,x)dx$ to the total cancer cell population). Bottom, left and middle: evolution with time of the optimal drug infusions of cytotoxic (u_1) and cytostatic (u_2) drugs. One can check on this simulation the quasi-bang-bang character of the optimal control.

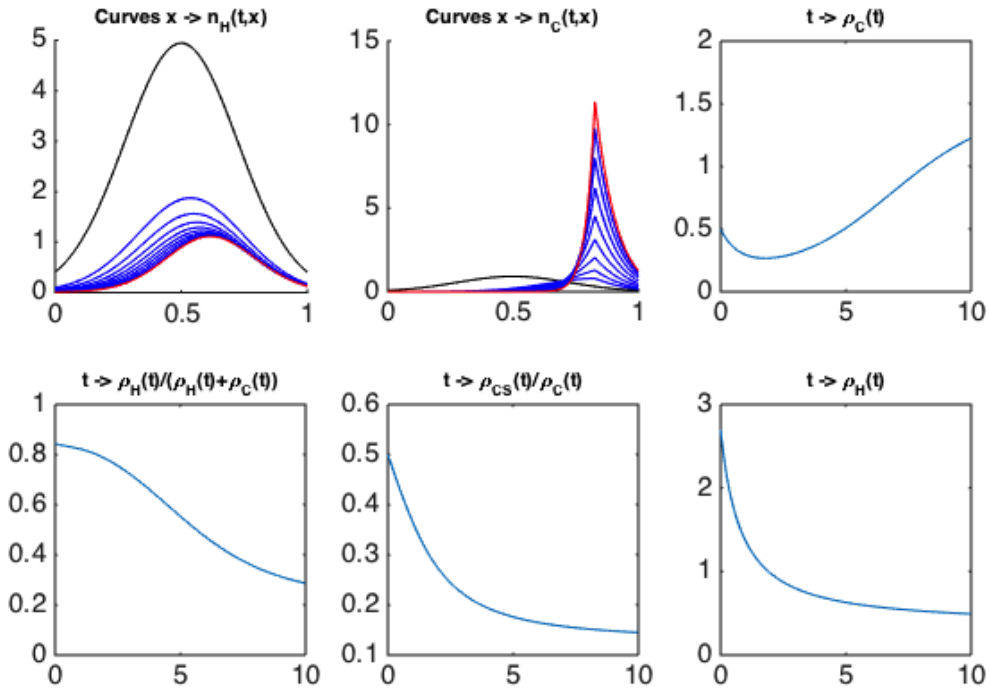


Figure 2. Comparative evolution with constant high drug doses. Catastrophic deleterious effects of the treatment on the concentration of the drug resistance phenotype x in the cancer cell population (top middle), and on the cell population numbers $\rho_C, \rho_H, \rho_{CS}$. Simulation with $u_1(t) = \mathrm{Cst} = 3.5$, $u_2(t) = \mathrm{Cst} = 2$, in time horizon $T = 10$.

be identified, thus helping us to understand complex cell behaviours such as cell growth and division in relation to mechanical stress. The model analysis further elucidates the relation between applied pressure, cell compressibility and cell density. Moreover, the model developments within this paper points a way of how to handle the so far open issue of high compression within the popular so-called Centre-Based Models, in which force between cells is modelled as forces between cell centres [54].

7.2. Epidemiology

Participants: Luis Lopes Neves de Almeida, M. Soledad Aronna [FGV, Rio de Janeiro], Pierre-Alexandre Bliman, Flávio C. Coelho [FGV, Rio de Janeiro], Martin Strugarek, Nicolas Vauchelet, Jorge Zubelli [IMPA, Rio de Janeiro].

7.2.1. Establishing Wolbachia by feedback

The releases of *Wolbachia*-positive mosquitoes are usually completed on an open-loop approach, that is, with a schedule computed once for all before the beginning of the experiment. Using the fact that measurements are achieved and available during the whole release process, we applied feedback control technique to devise an introduction protocol which is proved to guarantee that the population converges to a stable equilibrium where the totality of mosquitoes carry *Wolbachia*. A major advantage of feedback compared to open-loop approaches is its ability to cope with the uncertainties in the model dynamics (typically in the modelling of the life stages and the population structure), in the parameters (population size, mortality, reproductive rates, etc.), and in the size of the population to be treated.

7.2.2. Travelling waves in the problem of infestation by Wolbachia

As described above, a new method of control of dengue fever consists in releasing Wolbachia-infected mosquitos in the field, in the aim to replace the whole existing population by a population unable to transmit Dengue fever. In the study of the feasibility of such a strategy, an important issue concerns the spacial propagation of the mosquitoes. More precisely, releasing infected mosquitoes in a given domain (which can be a part of the city of Rio de Janeiro), the hope is to invade the whole area. The study of this propagation phenomena falls into the study of existence of traveling wave. In a recent paper [30], the authors have proposed a mathematical model to study such phenomena and they have simplified it to recover a well-know simple bistable system for which existence of traveling wave is known. The study of the probability of success of spacial invasiveness has been performed in [53].

7.3. Aggregation Kinetics

Participants: Aurora Armiento, Tom Banks [CRSC, NCSU, Raleigh, USA], Etienne Bernard, Thibault Bourgeron, José Antonio Carrillo [Imperial College, London, United Kingdom], Marie Doumic, Dirk Drasdo, Miguel Escobedo [Universidad del País Vasco, Bilbao, Spain], Sarah Eugène, Pierre Gabriel [Université Paris-Dauphine], Marc Hoffmann [Ceremade, Université Paris-Dauphine], François James [MAPMO, Université d'Orléans], Nathalie Krell [Université de Rennes 1], Frédéric Lagoutière [Département de mathématiques d'Orsay], Philippe Moireau [Inria Paris Saclay, M3DISIM project-team], Benoît Perthame, Stéphanie Prigent, Human Rezaei [VIM, INRA Jouy-en-Josas], Lydia Robert [Laboratoire Jean Perrin, UPMC], Philippe Robert [Inria Paris, RAP project-team], Maria Teresa Teixeira [IBCP, Paris], Joan Torrent [INRA, Jouy-en-josas], Magali Tournus [Ecole Centrale de Marseille], Nicolas Vauchelet, Min Tang [Jiaotong University, Shanghai], Zhou Xu [IBCP, Paris], Wei-Feng Xue [University of Kent, United Kingdom], Yi Yin.

7.3.1. Heterogeneity as an intrinsic feature in biological dynamics

Variability in nucleated polymerisation The kinetics of amyloid assembly show an exponential growth phase preceded by a lag phase, variable in duration as seen in bulk experiments and experiments that mimic the small volumes of cells. Sarah Eugène's Ph.D, defended in September 2016, was devoted to the study of the origins and the properties of the observed variability in the lag phase of amyloid assembly currently not accounted for by deterministic nucleation dependent mechanisms. In [20], we formulated a new stochastic minimal model

that is capable of describing the characteristics of amyloid growth curves despite its simplicity. We then solve the stochastic differential equations of our model and give mathematical proof of a central limit theorem for the sample growth trajectories of the nucleated aggregation process. These results give an asymptotic description for our simple model, from which closed form analytical results capable of describing and predicting the variability of nucleated amyloid assembly were derived. We also demonstrate the application of our results to inform experiments in a conceptually friendly and clear fashion. Our model offers a new perspective and paves the way for a new and efficient approach on extracting vital information regarding the key initial events of amyloid formation.

However, this first model does not explain completely the variability observed in the experiments. In [17], we thus investigated extensions to take into account other mechanisms of the polymerisation process that may have an impact on fluctuations.The first variant consists in introducing a preliminary conformation step to take into account the biological fact that, before being polymerised, a monomer has two states, regular or misfolded. Only misfolded monomers can be polymerised so that the fluctuations of the number of misfolded monomers can be also a source of variability of the number of polymerised monomers. The second variant represents the reaction rate of spontaneous formation of a polymer as of the order of α, with α some positive constant. First and second order results for the starting instant of nucleation are derived from these limit theorems. The proofs of the results rely on a study of a stochastic averaging principle for a model related to an Ehrenfest urn model, and also on a scaling analysis of a population model.

Image and statistical analysis of protein fibrils Protein fibrils present an important structural diversity, not only their length, but also their width, whether they present branches or not, etc. These structures may reveal the presence of different types of aggregates, possibly formed out of different polymerisation pathways. To analyse this diversity of shapes and structures, we developed an image analysis software, based on the expertise acquired by Y. Yin during her PhD for the image analysis of vessels. This software is able to track fibrils and measure their length, number of branches, and variable widths, even with poor quality images and crossing fibrils. This done, it allows us to perform a statistical analysis of the fibrils, to elucidate the main structuring features (Figure 3).

7.3.2. Inverse Problems and Data Assimilation Applied to Protein Aggregation and other settings

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation, a key feature for amyloid diseases. A. Armiento's Ph.D was devoted to the question of adapting data assimilation strategies to the specific context and difficulties of protein aggregation. In [6], we settled a framework problem when the experimental measurements consist in the time-dynamics of a moment of the population (*i.e.*, for instance the total polymerised mass, as in Thioflavine T measurements, or the second moment measured by Static Light Scattering). We propose a general methodology, and we solve the problem theoretically and numerically in the case of a depolymerising system. We then apply our method to experimental data of degrading oligomers, and conclude that smaller aggregates of ovPrP protein should be more stable than larger ones. This has an important biological implication, since it is commonly admitted that small oligomers constitute the most cytotoxic species during prion misfolding process.

7.3.3. Time asymptotics for growth-fragmentation equations

The long-term dynamics of fragmentation and growth-fragmentation equations has constantly been for the MAMBA (and for the ex-BANG) team an important research field. Thanks to these common efforts, these equations are now well understood. However, there remain some interesting open questions. In particular, if the generic long-time behaviour for the linear equation is known - given by a (generally exponential) trend towards a steady exponential growth described by the positive eigenvector linked to the dominant eigenvalue, see [83] for most recent results - critical cases are not yet fully understood.

With Miguel Escobedo, we focused on an important critical case, when the fragmentation is constant and the growth rate is either null or linear [16]. Using the Mellin transform of the equation, we determine the long time behaviour of the solutions and the speed of convergence, which may be either exponential or at most

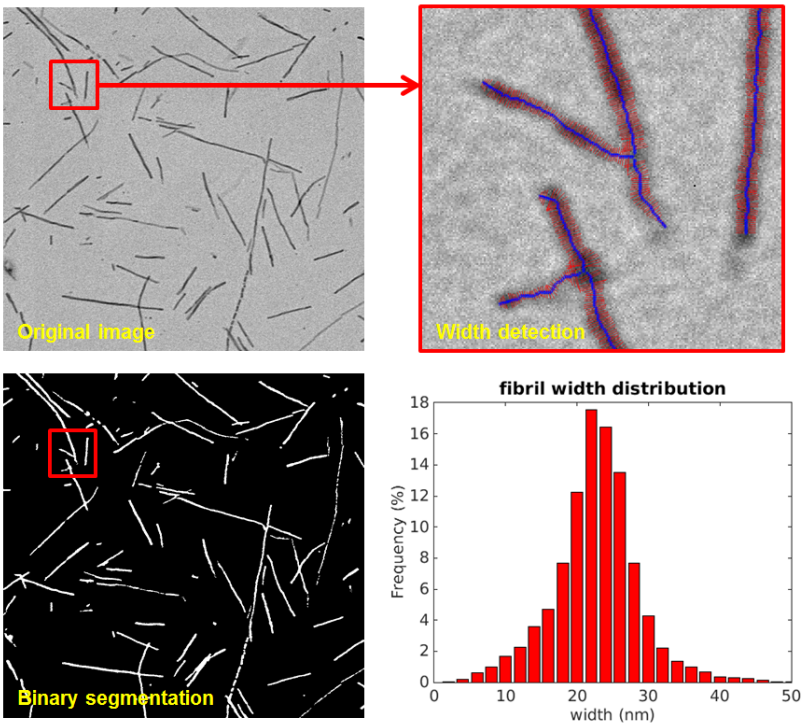


Figure 3. Fibril analysis yielding from original data via segmentation and analysis to the fibril length distribution.

polynomial according to the subdomain of $(t, x) \in \mathbb{R}_+^2$ which is considered. Our results show in particular the strong dependence of this asymptotic behaviour with respect to the initial data, in contrast to the generic results. Following our study, J. Bertoin and A. Watson proposed a complementary probabilistic analysis of related models [57]. These results exemplify the continuing need for further analysis of these interesting equations.

With E. Bernard and P. Gabriel, in [36], we investigated the "idealised" mitotic case, when the growth is exponential and the division results in two exactly equal parts. This case exhibits a lack of dissipativity, and the solutions appear to have a periodic limit cycle. We were nonetheless able to prove an entropy inequality, and to express the limit as an explicit oscillatory function, analytically given by the projection of the initial state on the space generated by the countable set of the dominant eigenvectors of the operator.

7.3.4. Cell aggregation by chemotaxis

Bacterial chemotaxis is now a well-known phenomenon. In particular, it has been established that the motion of bacteria is due to the alternation of straight swims in a given direction with tumble phases. More precisely, when bacteria notice that they do not go in a favorable direction, they may change their direction. Well established models are now available. In particular, the use of such systems allows to recover successfully the behaviour observed in biological experiments (see e.g. [18]). The bacterial response to changes in their environment can be described by an internal variable. In a recent work [29], it has been established that a well-known kinetic model can be obtained from such a model incorporating an internal variable.

When, the frequency of tumbling is high, the motion is mainly driven by tumbling and models reduce to describe aggregation phenomena. From a mathematical point of view, the study of such model is challenging since classical solution may not exist for any time. Then a notion of weak measure solution should be introduced [10]. Numerical investigation of such solutions has been performed in [21], [48].

7.4. Modelling of the liver

Participants: François Bertaux [Imperial College, London], Noémie Boissier, Dirk Drasdo, Géraldine Cellière, Adrian Friebel, Group Heinzle [Univ. Saarbruecken, Germany], Group Hengstler [IfADo, Germany], Stefan Hoehme, Tim Johann, Irène Reo [Vignon-Clementel], Paul Van Liedekerke, Eric Vibert [Hopital Paul Brousse], Group Zerial [Max-Planck Inst. for Molecular Genetics, Dresden, Germany], Groups Iflow, Notox, Vln.

7.4.1. Ammonia detoxification after drug-induced damage

Overdosing acetaminophen (APAP) is the main reason for acute liver failure in the US and UK. Overdose of APAP destroys the hepatocytes localised in the center of each liver lobule (pericentral damage), the repetitive functional and anatomical tissue units of liver. The Human has about 1 million of such lobules. As a consequence, the blood is not sufficiently detoxified from ammonia, which is toxic to the body and can lead to encephalopathy. In France about 1000 cases occur with ammonia toxicity each year. In recent papers we demonstrated by an integrated model that the widely accepted scheme of key reactions for ammonia detoxification is insufficient to explain ammonia detoxification after pericentral lobule damage and predicts a missing ammonia sink [71]. The integrated model couples ODEs representing the consensus reactions in the spatial temporal liver lobule regeneration model. This finding has triggered new experiments leading to the identification of a widely ignored but fundamentally important ammonia sink mechanism. We could show by testing a number of different mechanisms within novel models that this sink mechanism was the only one able to explain the data [74] (and Geraldine Cellière's PhD thesis [3], 2016). The reaction turned out to have the potential to be used in therapeutics by injection of a molecular cocktail triggering it.
In a follow-up work, the ammonia detoxifying reactions have been integrated into each hepatocyte of the previously established tissue-level liver lobule model of regeneration. The final multi-level model simulates blood flow, transport of metabolites and detoxification of ammonia in every hepatocyte of a regenerating lobule. This multi-level model could validate the missing ammonia sink found in the integrated model in ref. [74] but yields differences to the integrated model if the ammonia sink mechanism is integrated. Still by

reparameterisation, adding the ammonia sink mechanism, the model is able to explain the data but the results clearly show that spatio-temporal modelling can give results different from pure compartment modelling. In the case of quantitative modelling in pharmacology or toxicology this can be fundamental. We were able to analyse and generalise these findings.

7.4.2. Predicting in vivo drug toxicity from in vitro data

In vitro experiments on APAP (aka paracetamol, acetaminophen) have been used to calibrate a model of APAP drug toxicity using in vitro data, modifying this model to predict in vivo toxicity. This procedure is aimed at as a general pathway from cosmetic and pharmaceutical companies to eliminate or at least reduce animal experiments and, in perspective, permit a better prediction of drug toxicity in the Human. Three critical differences between in vitro and in vivo were stepwise integrated in the model calibrated with in vitro toxicity data to study their impact on in vivo toxicity predictions. (1) The temporal drug exposure profile, (2) the temporal concentration profile of a class of key enzymes, CYP enzymes. Only in hepatocytes in which CYP enzymes are present is APAP metabolised and can downstream cell death occur. (3) The liver architecture represents critical differences in the spatial distribution of the drug. The results are in preparation for publication (Géraldine Cellière's PhD thesis 2016, Cellière et. al., in preparation).

7.4.3. Liver cancer

The aggressiveness of a tumour may be reflected by its micro-architecture. To gain a deeper understanding of the mechanisms controlling the spatial organisation of tumors at early stages after tumour initiation, we used an agent-based spatio-temporal model previously established to simulate features of liver regeneration [76]. This model was further developed to simulate scenarios in early tumour development, when individual initiated hepatocytes gain increased proliferation capacity [37]. The model simulations were performed in realistic liver microarchitectures obtained from 3D reconstruction of confocal laser scanning micrographs. Interestingly, the here established model predicted that initially initiated hepatocytes arrange in elongated patterns. Only when the tumour progresses to cell numbers of approximately $4,000$ does it adopt spherical structures. This model prediction was validated by the analysis of initiated cells in a rat liver tumour initiation study using single doses of 250 mg/kg of the genotoxic carcinogen N-nitrosomorpholine (NNM). Indeed, small clusters of GST-P positive cells induced by NNM were elongated, almost columnar, while larger GDT-P positive foci of approximately the size of liver lobules, adopted spherical shapes. Simulation of numerous possible mechanisms demonstrated that only hepatocyte-sinusoidal-alignment (HSA), a previously discovered order mechanism involved in the coordination of liver tissue architecture, could explain the experimentally observed initial deviation from spherical shape. The present study demonstrates that the architecture of small hepatocellular tumour cell clusters early after initiation is still controlled by physiological control mechanisms. However, this coordinating influence is lost when the tumour grows to approximately $4,000$ cells, leading to further growth in spherical shape (Figure 4). Our findings stress the potential importance of organ micro-architecture in understanding tumour phenotypes.

7.5. Miscellaneous

Participants: M. Soledad Aronna [FGV, RIo de Janeiro], Bettina d'Avila Barros [FGV, Rio de Janeiro], Pierre-Alexandre Bliman, Noémie Boissier, Géraldine Cellière, Flávio C. Coelho [FGV, Rio de Janeiro], Marie Doumic, Benoît Perthame, Tales Rands Amazonas [FGV, Rio de Janeiro], Group Reo [Inria Paris - Rocquencourt], Edouard Ribes [SANOFI], Martin Strugarek, Nathan Toubiana [École Polytechnique], Paul Van Liedekerke, Nicolas Vauchelet, Jorge Zubelli [IMPA, Rio de Janeiro].

7.5.1. Diffusive waves generated by a travelling wave

Observations in developmental biology show that calcic waves, generated after fertilisation within the egg cell endoplasmic reticulum, propagate within the egg cell. This motivates to explore in which circumstances a travelling wave solution of a reaction-diffusion equation can generate a travelling wave for the diffusion equation. For this purpose, we construct analytical solutions for a system composed of a reaction-diffusion equation coupled with a purely diffusive equation. We consider both the monostable (of the Fisher-KPP type)

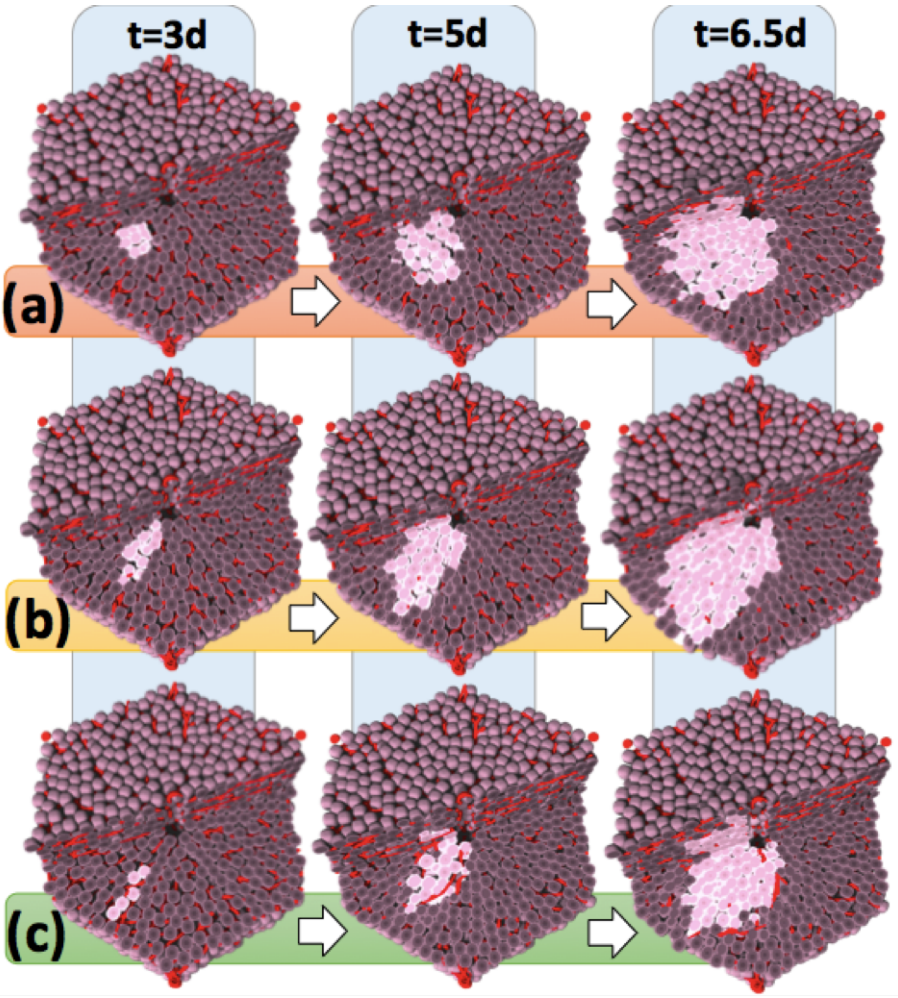


Figure 4. Scenarios of tumour growth in a single liver lobule in (a) absence of hepatocyte-sinusoidal-alignment (HSA), (b) presence of HSA, and (c) presence of HSA with elevated tangential friction impeding hepatocyte movement perpendicular to the columns formed along the sinusoids [35]. The images represent snapshots 3, 5 and 6.5 days after initiation, defined as the time point when a transformed hepatocyte adopts an increased proliferation rate. Notice that HSA (b, c) clearly causes early asymmetry of tumour cell assemblies (leftmost image column at 3 days) while with increasing tumour size this asymmetry is increasingly lost (right panel at 6.5 days). A one-cell thick column could be found if the movement perpendicular to the sinusoids was impeded by elevated shear forces, e.g., from tight junctions. This predicted evolutionary scenario reproduces the experimentally observed scenario.

and bistable cases. We use a piecewise linear reaction term so as to build explicit solutions, which leads us to compute exponential tails, the exponents of which are roots of second, third or fourth-order polynomials. These rise conditions on the coefficients for existence of a travelling wave of the diffusion equation. The question of positivity and monotonicity is only partially answered. See [49].

7.5.2. Dealing with uncertainty in modelling

Interval observers for time-varying uncertain epidemiological models. SIR models constitute an elementary class of deterministic models of evolution of epidemics. We examine here the issue of state estimation for such models, subjected to seasonal variations and uncertainties in the transmission rates. Direct or indirect (through a vector) transmission is considered. In both cases, the measurement is assumed to consist of the number of new infectives per unit time, that is the information usually provided by the public health systems. We construct classes of interval observers with estimate-dependent gain, and provide corresponding asymptotic error bounds.

7.5.3. Modelling strategic workforce planning with structured population equations

We initiated a promising collaboration with the human resource department of Sanofi (E. Ribes), aiming at proposing a unified modelling of workforce planning based on structured population equations. Strategic Workforce Planning is a company process providing best in class, economically sound, workforce management policies and goals. Despite the abundance of literature on the subject, this is a notorious challenge in terms of implementation. Reasons span from the youth of the field itself to broader data integration concerns that arise from gathering information from financial, human resource and business excellence systems. In [43], we set the first stones to a simple yet robust quantitative framework for Strategic Workforce Planning exercises. Firstly, a method based on structured equations is detailed. It is then used to answer two main workforce-related questions: how to optimally hire to keep labour costs flat? How to build an experience-constrained workforce at a minimal cost? Further developments are in progress.

MONC Project-Team

7. New Results

7.1. Free boundary problem for cell protrusion formations: theoretical and numerical aspects

Authors: Olivier Gallinato, Masahito Ohta, Clair Poignard, Takashi Suzuki

In this paper, a free boundary problem for cell protrusion formation is studied theoretically and numerically. The cell membrane is precisely described thanks to a level set function, whose motion is due to specific signalling pathways. The aim is to model the chemical interactions between the cell and its environment, in the process of invadopodia or pseudopodia formation. The model consists of Laplace equation with Dirichlet condition inside the cell coupled to Laplace equation with Neumann condition in the outer domain. The actin polymerization is accounted for as the gradient of the inner signal, which drives the motion of the interface. We prove the well-posedness of our free boundary problem under a sign condition on the datum. This criterion ensures the consistency of the model, and provides conditions to focus on for any enrichment of the model. We then propose a new first order Cartesian finite-difference method to solve the problem. We eventually exhibit the main biological features that can be accounted for by the model: the formation of thin and elongated protrusions as for invadopodia, or larger protrusion as for pseudopodia, depending on the source term in the equation. The model provides the theoretical and numerical grounds for single cell migration modeling, whose formulation is valid in 2D and 3D. In particular, specific chemical reactions that occured at the cell membrane could be precisely described in forthcoming works. Journal: Journal of Mathematical Biology, Springer Verlag (Germany), 2016, <10.1007/s00285-016-1080-7> lien hal: https://hal.inria.fr/hal-01412264v1

7.2. Mathematical model for transport of DNA plasmids from the external medium up to the nucleus by electroporation

Authors: Michael Leguèbe, M Notarangelo, Monika Twarogowska, Roberto Natalini, Clair Poignard

This work is devoted to modelling gastrointestinal stromal tumour metastases to the liver, their growth and resistance to therapies. More precisely, resistance to two standard treatments based on tyrosine kinase inhibitors (imatinib and sunitinib) is observed clinically. Using observations from medical images (CT scans), we build a spatial model consisting in a set of non-linear partial differential equations. After calibration of its parameters with clinical data, this model reproduces qualitatively and quantitatively the spatial tumour evolution of one specific patient. Important features of the growth such as the appearance of spatial heterogeneities and the therapeutical failures may be explained by our model. We then investigate numerically the possibility of optimizing the treatment in terms of progression-free survival time and minimum tumour size reachable by varying the dose of the first treatment. We find that according to our model, the progression-free survival time reaches a plateau with respect to this dose. We also demonstrate numerically that the spatial structure of the tumour may provide much more insights on the cancer cell activities than the standard RECIST criteria, which only consists in the measurement of the tumour diameter. Finally, we discuss on the non-predictivity of the model using only CT scans, in the sense that the early behaviour of the lesion is not sufficient to predict the response to the treatment. Journal: Mathematical Medicine and Biology, Oxford University Press (OUP), 2016, <10.1093/imammb/dqw002> lien hal: https://hal.inria.fr/hal-01380292

7.3. Free boundary problem for cell protrusion formations: theoretical and numerical aspects

Authors: Olivier Gallinato, Masahito Ohta, Clair Poignard, Takashi Suzuki

In this paper, a free boundary problem for cell protrusion formation is studied theoretically and numerically. The cell membrane is precisely described thanks to a level set function, whose motion is due to specific signalling pathways. The aim is to model the chemical interactions between the cell and its environment, in the process of invadopodia or pseudopodia formation. The model consists of Laplace equation with Dirichlet condition inside the cell coupled to Laplace equation with Neumann condition in the outer domain. The actin polymerization is accounted for as the gradient of the inner signal, which drives the motion of the interface. We prove the well-posedness of our free boundary problem under a sign condition on the datum. This criterion ensures the consistency of the model, and provides conditions to focus on for any enrichment of the model. We then propose a new first order Cartesian finite-difference method to solve the problem. We eventually exhibit the main biological features that can be accounted for by the model: the formation of thin and elongated protrusions as for invadopodia, or larger protrusion as for pseudopodia, depending on the source term in the equation. The model provides the theoretical and numerical grounds for single cell migration modeling, whose formulation is valid in 2D and 3D. In particular, specific chemical reactions that occured at the cell membrane could be precisely described in forthcoming works. Journal: Journal of Mathematical Biology, Springer Verlag (Germany), 2016, <10.1007/s00285-016-1080-7> lien hal: https://hal.inria.fr/hal-01412264v1

7.4. Spatial modelling of tumour drug resistance: the case of GIST liver metastases Mathematical Medicine and Biology Advance

Authors: Guillaume Lefebvre, François Cornelis, Patricio Cumsille, Thierry Colin, Clair Poignard, Olivier Saut

This work is devoted to modelling gastrointestinal stromal tumour metastases to the liver, their growth and resistance to therapies. More precisely, resistance to two standard treatments based on tyrosine kinase inhibitors (imatinib and sunitinib) is observed clinically. Using observations from medical images (CT scans), we build a spatial model consisting in a set of non-linear partial differential equations. After calibration of its parameters with clinical data, this model reproduces qualitatively and quantitatively the spatial tumour evolution of one specific patient. Important features of the growth such as the appearance of spatial heterogeneities and the therapeutical failures may be explained by our model. We then investigate numerically the possibility of optimizing the treatment in terms of progression-free survival time and minimum tumour size reachable by varying the dose of the first treatment. We find that according to our model, the progression-free survival time reaches a plateau with respect to this dose. We also demonstrate numerically that the spatial structure of the tumour may provide much more insights on the cancer cell activities than the standard RECIST criteria, which only consists in the measurement of the tumour diameter. Finally, we discuss on the non-predictivity of the model using only CT scans, in the sense that the early behaviour of the lesion is not sufficient to predict the response to the treatment. Journal: Mathematical Medicine and Biology, Oxford University Press (OUP), 2016, <10.1093/imammb/dqw002> lien hal: https://hal.inria.fr/hal-01380292

7.5. Mathematical modeling of cancer immunotherapy and synergy with radiotherapy

Team participant: S. Benzekry Other participants: R. Serre, N. André, J. Ciccolini, D. Barbolosi (SMARTc, Inserm, Marseille, FR), L. Padovani, X. Muracciole (Radiotherapy Unit, La Timone Hospital, Marseille, FR), F. Barlési (Multidisciplinary Oncology and Therapeutic Innovations Unit, AP-HM, Marseille, FR) and C. Meille (Roche Pharmaceutics, Basel, Switzerland)

Combining radiotherapy with immune checkpoint blockade may offer considerable therapeutic impact if the immunosuppressive nature of the tumor microenvironment (TME) can be relieved. In this study, we used mathematical models, which can illustrate the potential synergism between immune checkpoint inhibitors and radiotherapy. A discrete-time pharmacodynamic model of the combination of radiotherapy with inhibitors of the PD1–PDL1 axis and/or the CTLA4 pathway is described. This mathematical framework describes how a growing tumor first elicits and then inhibits an antitumor immune response. This antitumor immune response is described by a primary and a secondary (or memory) response. The primary immune response appears first

and is inhibited by the PD1–PDL1 axis, whereas the secondary immune response happens next and is inhibited by the CTLA4 pathway. The effects of irradiation are described by a modified version of the linear-quadratic model. This modeling offers an explanation for the reported biphasic relationship between the size of a tumor and its immunogenicity, as measured by the abscopal effect (an off-target immune response). Furthermore, it explains why discontinuing immunotherapy may result in either tumor recurrence or a durably sustained response. Finally, it describes how synchronizing immunotherapy and radiotherapy can produce synergies. The ability of the model to forecast pharmacodynamic endpoints was validated retrospectively by checking that it could describe data from experimental studies, which investigated the combination of radiotherapy with immune checkpoint inhibitors. In summary, a model such as this could be further used as a simulation tool to facilitate decision making about optimal scheduling of immunotherapy with radiotherapy and perhaps other types of anticancer therapies.

7.6. Non-Standard Radiotherapy Fractionations Delay the Time to Malignant Transformation of Low-Grade Gliomas

Team participant: S. Benzekry. Other participants: A. Henares-Molina, V.M. Perez-Garcia and A. Martinez-Gonzalez (Môlab, Universidad de Castilla-La Mancha, Ciudad Real, Spain) P.C. Lara (Radiation Oncology, Las Palmas University Hospital, Las Palmas, Spain), M. Garcia-Rojo (Pathology department, Jerez de la Frontera Hospital, Jerez de la Frontera, Spain)

Grade II gliomas are slowly growing primary brain tumors that affect mostly young patients. Cytotoxic therapies (radiotherapy and/or chemotherapy) are used initially only for patients having a bad prognosis. These therapies are planned following the "maximum dose in minimum time" principle, i. e. the same schedule used for high-grade brain tumors in spite of their very different behavior. These tumors transform after a variable time into high-grade tumors, what decreases significantly the patient's life expectancy. In this paper we study mathematical models describing the growth of grade II gliomas in response to radiotherapy. We find that protracted metronomic fractionations, *i.e.* therapeutical schedules enlarging the time interval between low-dose radiotherapy fractions, may lead to a better tumor control without an increase in toxicity. Other non-standard fractionations such as protracted or hypoprotracted schemes may also be beneficial. The potential survival improvement depends on the tumor proliferation rate and can be even of the order of years. A conservative metronomic scheme, still being a suboptimal treatment, delays the time to malignant progression of at least one year when compared to the standard scheme.

7.7. Model-driven optimization of antiangiogenics + cytotoxics combination in breast cancer mice treated with bevacizumab and paclitaxel

Team participant: S. Benzekry. Other participants: S. Mollard (CRUK, Cambridge, UK), J. Ciccolini, D-C Imbs, R. El Cheikh, D. Barbolosi (SMARTc, Inserm, Marseille, FR)

Bevacizumab is the first-in-class antiangiogenic drug administrated concomitantly with cytotoxics. Several reports have shown that antiangiogenics could induce a transient phase of vascular normalization, thus ensuring a better drug delivery provided that cytotoxics administration is delayed. However, determining this best sequence is challenging. We have developed a simple mathematical model describing the impact of antiangiogenics on tumor vasculature. A 3.4 days delay between bevacizumab and paclitaxel was first proposed by the model as an optimal sequence. To test its relevance, 84 mice were orthotopically xenografted with human MDA-231Luc+ breast cancer cells. Two different sets of experiments were performed, based upon different bevacizumab dosing (10 or 20 mg/kg) and inter-cycle intervals (7 or 10 days), comprising several combinations with paclitaxel. Results showed that scheduling bevacizumab administration 3 days before paclitaxel improved antitumor efficacy (48% reduction in tumor growth as compared with concomitant dosing, $p<0.05$) while reducing metastatic spreading. Additionally, bevacizumab alone could lead to more aggressive metastatic disease with shorter survival in animals. Our phenomenological model was able to fit e perietal data a d provided insight o the underlying d a i s of as ulature's a ilit to deliver the cytotoxic agent. Final simulations suggested a new, data-informed optimal sequence of 2.4 days. Our data suggest that concomitant

dosing between bevacizumab and paclitaxel could be a sub-optimal strategy at bedside. In addition, this proof of concept study suggests that mathematical modelling could help to identify the best sequence among a variety of possible alternate treatment modalities, thus refining the way experimental or clinical studies are conducted.

7.8. Dynamics of concomitant resistance: data, theories and mathematical modeling

Team participant: S. Benzekry Other participants: C. Lamont, L. Hlatky, P. Hahnfeldt (Center of Cancer and Systems Biology, Boston, USA)

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, the two models of competition for nutrients and indirect angiogenesis-regulated inhibition were not able to explain the growth behavior as well as a third model based on direct systemic inhibition. The superior model offers a depiction of concomitant resistance that provides an improved theoretical basis for tumor growth control that may also find utility in therapeutic planning to avoid post-surgery metastatic acceleration.

7.9. Modeling spontaneous metastasis following surgery: an in vivo-in silico approach

Team participant: S. Benzekry. Other participants: A. Tracz, M. Mastri, R. Corbelli and J. Ebos (Roswell Park Cancer Institute, Buffalo, USA) D. Barbolosi (SMARTc, Inserm, Marseille, FR)

Rapid improvements in the detection and tracking of early-stage tumor progression aim to guide decisions regarding cancer treatments as well as predict metastatic recurrence in patients following surgery. Mathematical models may have the potential to further assist in estimating metastatic risk, particularly when paired with in vivo tumor data that faithfully represent all stages of disease progression. Herein we describe mathematical analysis that uses data from mouse models of spontaneous metastasis developing after surgical removal of orthotopically implanted primary tumors. Both presurgical (primary tumor) and postsurgical (metastatic) growth was quantified using bioluminescence and was then used to generate a mathematical formalism based on general laws of the disease (*i.e.* dissemination and growth). The model was able to fit and predict pre-/post-surgical data at the level of the individual as well as the population. Our approach also enabled retrospective analysis of clinical data describing the probability of metastatic relapse as a function of primary tumor size. In these data-based models, inter-individual variability was quantified by a key parameter of intrinsic metastatic potential. Critically, our analysis identified a highly nonlinear relationship between primary tumor size and postsurgical survival, suggesting possible threshold limits for the utility of tumor size as a predictor of metastatic recurrence. These findings represent a novel use of clinically relevant models to assess the impact of surgery on metastatic potential and may guide optimal timing of treatments in neoadjuvant (presurgical) and adjuvant (postsurgical) settings to maximize patient benefit.

7.10. Computational Trials: Unraveling Motility Phenotypes, Progression Patterns, and Treatment Options for Glioblastoma Multiforme

Team participants: Thierry Colin, Olivier Saut. Other participants: Fabio Raman, Elizabeth Scribner, Olivier Saut, Cornelia Wenger, Hassan Fathallah-Shaykh.

Glioblastoma multiforme is a malignant brain tumor with poor prognosis and high morbidity due to its invasiveness. Hypoxia-driven motility and concentration-driven motility are two mechanisms of glioblastoma multiforme invasion in the brain. The use of anti-angiogenic drugs has uncovered new progression patterns of glioblastoma multiforme associated with significant differences in overall survival. Here, we apply a mathematical model of glioblas- toma multiforme growth and invasion in humans and design computational trials using agents that target angiogenesis, tumor replication rates, or motility. The findings link highly- dispersive, moderately-dispersive, and hypoxia-driven tumors to the patterns observed in glioblastoma multiforme

treated by anti-angiogenesis, consisting of progression by Expand- ing FLAIR, Expanding FLAIR + Necrosis, and Expanding Necrosis, respectively. Further- more, replication rate-reducing strategies (e.g. Tumor Treating Fields) appear to be effective in highly-dispersive and moderately-dispersive tumors but not in hypoxia-driven tumors. The latter may respond to motility-reducing agents. In a population computational trial, with all three phenotypes, a correlation was observed between the efficacy of the rate- reducing agent and the prolongation of overall survival times. This research highlights the potential applications of computational trials and supports new hypotheses on glioblastoma multiforme phenotypes and treatment options.

N

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.

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. Symmetric coupling of multiple timescale systems with mixed-mode oscillations

Participants: Soledad Fernández García, Alexandre Vidal, Fabrizio de Vico Fallani [EPI Aramis], Frédérique Clément.

We have analyzed a six-dimensional slow-fast system consisting of two coupled identical oscillators. Each oscillator is a three-dimensional system consisting of a FitzHugh-Nagumo system with an additional variable representing the calcium concentration. Individually, each three-dimensional subsystem possesses an attractive Mixed-Mode oscillations limit cycle, displaying small oscillations due to the presence of a folded saddle-node type II singularity for a certain range of the parameters values. We have considered a linear coupling through the fast variable in the slow equation and study the synchronization patterns of two identical systems with identical coupling parameter. Apart from stable in-phase and stable anti-phase synchronization patterns, the system presents almost-in-phase synchronization, oscillation death of one of the oscillators and total oscillation death, intertwined with complex transitions involving period doubling cascade, period adding phenomena and chaos. We have pointed out the role of Mixed-Mode oscillations in the birth of the different patterns and the transitions from one regime to another.

Part of these results have been presented as a contributed talk to the SIAM conference on life science https://www.siam.org/meetings/ls16/: (A Study of the Synchronization Between Two Coupled Neuron Models Generating Mixed-Mode Oscillations. A. Vidal, S.Fernández García, F. Clément, F. De Vico Fallani). MS48 Applications of Multiple Time Scale Dynamics in Biological Systems.

7.1.3. 3D-Explosion of cycles and spike-adding in the Hindmarsh-Rose model

Participants: Lucile Megret, Mathieu Desroches [Sophia], Jean-Pierre Françoise, Maciej Krupa [Sophia].

We have considered slow-fast systems that feature bursting oscillations, the minimal configuration being two fast variables and one slow variable. In the Hindmarsh-Rose model, as the slow variable z evolves, the fast dynamics undergoes several bifurcations (two Hopf bifurcations, two homoclinic bifurcations, two focus-node and two saddle-node bifurcations). We have focused on the existence of a sequence of 3D-candidate limit periodic sets of a new type. Numerical simulations have shown that it generates for the full 3D-dynamics and (the small parameter) " small enough a 3D-explosion of cycles. We have discussed the relation between this 3D-explosion and the spike-adding. We have also emphasized another new phenomenon induced by the slow-crossing of a saddle-node bifurcation with solutions which after coming close to the fold point, continue to follow it along its non-hyperbolic center manifold. We have shown how this phenomenon is also involved in the spike-adding mechanism taking place in square-wave bursters such as the Hindmarsh-Rose system.
Part of these results have been presented at the "36e Séminaire de la Sociéte Francophone de Biologie théorique", St-Flour (France), June 12-15 2016.

7.1.4. Wild oscillations in a nonlinear neuron model with resets

Participants: Jonathan Rubin [University of Pittsburgh], Justyna Signerska-Rynkowska, Jonathan Touboul, Alexandre Vidal.

In a series of two studies, 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.

The first study [30] 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 [31] 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.

7.1.5. Canard Explosions in delay differential equations

Participants: Jonathan Touboul, Maciej Krupa [Sophia].

We have analyzed in [21] canard explosions in delayed differential equations with a one-dimensional slow manifold. This study is applied to explore the dynamics of the van der Pol slow-fast system with delayed self-coupling. In the absence of delays, this system provides a canonical example of a canard explosion. We have shown that as the delay is increased a family of "classical" canard explosions ends as a Bogdanov-Takens bifurcation occurs at the folds points of the S-shaped critical manifold.

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 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 [29].

These results have been partly presented in a workshop on "Asymptotic behavior of systems of PDEs arising in physics and biology : theoretical and numerical points of view" (ABPDE II), Lille, June 15-17, 2016.

7.2.2. Analysis of the asymptotic behavior of a model for the morphogenesis in ovarian follicles

Participants: Frédérique Clément, Frédérique Robin, Romain Yvinec [INRA].

We have designed and analyzed a simplified version of our multiscale model for the morphogenesis of ovarian follicles [6]. We have formulated both a stochastic model, in the framework of branching processes, and a deterministic one, in the framework of nonconservative transport equations. The simplifications result in linear models, in which the oocyte growth is uncoupled from the proliferation of the surrounding follicular cells. The cell population is distributed into concentric layers around the oocyte, and structured according to the cell age. Cells are subject to the process of cell division, which resets their age and allow them to possibly move to the adjacent outer layer. Since there is no symmetry in the cell displacements (the only allowed cell motion is centrifugal), we have faced the problem of the model irreducibility. To study the asymptotic behavior, we thus had to adapt the classical results based on entropy or the computation of stochastic moments. We have proved that there is, as expected, an exponential asymptotic growth led by a Malthus parameter, which can be computed analytically in the simplest (Markovian) case, or numerically. Interestingly, the value of this global parameter merges with one of the local Malthus-like parameters defined on the layer level. In both the deterministic and stochastic cases, we could derive accurate information on the time-varying mean cell number per layer and we also got additional information on the asymptotic age distribution.

This work has been undergone in the framework of the master thesis of Frédérique Robin (M2 Mathématiques du Vivant, Université Paris-Saclay), and pursued as a PhD subject. Preliminary results have been the matter of a presentation during the "Journées INRA-Inria" held in Mallemort (France) on October 6-7th: F. Clément, F. Robin, R Yvinec. Dynamiques de populations cellulaires structurées individus-centrées : Morphogenèse des follicules ovariens.

7.2.3. Numerical study of a mathematical model for the dynamics of progenitor cell populations in the mouse cerebral cortex

Participants: Marie Postel, Alice Karam [IBPS], Frédérique Clément, Sylvie Schneider-Maunoury [IBPS].

We have studied numerically our multi-scale mathematical 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 already acquired in the framework of the NeuroMathMod project (Sorbonne-Universités Émergence call with IBPS, Institut de Biologie Paris Seine). 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.

7.3. Macroscopic limits of stochastic neural networks and neural fields

7.3.1. Limit theorems and effective dynamics

Participants: Jonathan Touboul, Philippe Robert [EPI RAP], Cristobal Quiñinao [IMT], Stéphane Mischler [CEREMADE].

We have pursued our investigations on the dynamics of large-scale neural networks modeling the brain, in two main directions:

We have studied in [26] the mean-field limit and stationary distributions of a pulse-coupled network modeling the dynamics of a large neuronal assemblies. Our model takes into account explicitly the intrinsic randomness of firing times, contrasting with the classical integrate-and-fire model. The ergodicity properties of the Markov process associated with finite networks have been investigated. We have derived the limit in distribution of the sample path of the state of a neuron of the network when its size gets large. The invariant distributions of this limiting stochastic process have been analyzed as well as their stability properties. We have shown that the system undergoes transitions as a function of the averaged connectivity parameter, and can support trivial states (where the network activity dies out, which is also the unique stationary state of finite networks in some cases) and self-sustained activity when connectivity level is sufficiently large, both being possibly stable.

We have investigated in [23] existence and uniqueness of solutions of a McKean-Vlasov evolution PDE representing the macroscopic behavior of interacting Fitzhugh-Nagumo neurons. This equation is hypoelliptic, nonlocal and has unbounded coefficients. We have proven the existence of a solution to the evolution equation and non trivial stationary solutions. Moreover, we have demonstrated the uniqueness of the stationary solution in the weakly nonlinear regime. Eventually, using a semigroup factorisation method, we have shown exponential nonlinear stability in the small connectivity regime.

7.3.2. Spectrum of random matrices

Participants: Jonathan Touboul, Gilles Wainrib [ENS], Luis Carlos Garcia Del Molino [New-York University], Khashayar Pakdaman [IJM].

We have considered in [20] the ensemble of Real Ginibre matrices with a positive fraction $\alpha > 0$ of real eigenvalues. We have demonstrated a large deviation principle for the joint eigenvalue density of such matrices and we have introduced a two phase log-gas whose stationary distribution coincides with the spectral measure of the ensemble. Using these tools we have provided an asymptotic expansion for the probability $p^n_{\alpha n}$ that an $n \times n$ Ginibre matrix has $k = \alpha n$ real eigenvalues and we have characterized the spectral measures of these matrices.

7.4. Modeling of brain development and brain functions

7.4.1. Organization of the visual cortex

Participants: Jonathan Touboul, Jérôme Ribot [CIRB], Alberto Romagnoni [ENS], Daniel Bennequin [IMG-PRG], Chantal Milleret [CIRB].

In the early visual cortex, information is processed within functional maps whose layout is thought to underlie visual perception. However, the precise organization of these functional maps as well as their interrelationships remains unresolved. We have investigated using new data acquisition and analysis as well as mathematical modeling, the inter-relationship between different visual maps in cat visual cortex.

We have shown in [25] that spatial frequency representation in cat areas 17 and 18 exhibits singularities around which the map organizes like an electric dipole potential. These singularities are precisely co-located with singularities of the orientation map: the pinwheel centers. We have first shown, using high resolution optical imaging, that a large majority (around 80%) of pinwheel centers exhibit in their neighborhood semi-global extrema in the spatial frequency map. These extrema create a sharp gradient that was confirmed with electrophysiological recordings. Based on an analogy with electromagnetism, a mathematical model of a dipolar structure has been proposed and accurately fitted to optical imaging data for two third of pinwheel centers with semi-global extrema. We have concluded that more than half of orientation pinwheel centers form spatial frequency dipoles in cat early visual cortex.

We have demonstrated mathematically in [27] that two natural principles, local exhaustivity of representation and parsimony, would constrain the orientation and spatial frequency maps to display co-located singularities around which the orientation is organized as a pinwheel and spatial frequency as a dipole. We have further focused on the theoretical implications of this structure. Using a computational model, we have shown that this architecture allows a trade-off in the local perception of orientation and spatial frequency, but this would occur for sharper selectivity than the tuning width reported in the literature. We therefore re-examined physiological data and have shown that indeed the spatial frequency selectivity substantially sharpens near maps singularities, bringing to the prediction that the system tends to optimize balanced detection between different attributes.

7.4.2. Modeling the timing of neurogenesis and control of the neuron pool : Enhanced abventricular proliferation compensates cell death in the embryonic cerebral cortex

Participants: Betty Freret-Hodara [IJM], Yi Cui, Amélie Griveau [IJM], Lisa Vigier [IJM], Yoko Arai [IJM], Jonathan Touboul, Alessandra Pierani [IJM].

Loss of neurons in the neocortex is generally thought to result in a final reduction of cerebral volume. Yet, little is known on how the developing cerebral cortex copes with death of early-born neurons. We have tackled this issue by taking advantage of a transgenic mouse model in which, from early embryonic stages to mid-corticogenesis, abundant apoptosis is induced in the postmitotic compartment. Unexpectedly, the thickness of the mutant cortical plate at E18.5 was normal, due to an overproduction of upper layer neurons at E14.5. We have developed and simulated a mathematical model to investigate theoretically the recovering capacity of the system and found that a minor increase in the probability of proliferative divisions of intermediate progenitors (IPs) is a powerful compensation lever. Combined with our experimental observations, these results illustrate the remarkable plasticity of neocortical progenitors to adapt to major embryonic insults via the modulation of abventricular divisions thereby ensuring the production of an appropriate number of neurons.

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, Faisal Amlani, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Mikel Landajuela Larma, Damiano Lombardi, Marina Vidrascu.

In [15] a simplified fluid-structure interaction method is proposed in order to deal with the simulation of fluids in elastic pipes. The motivation of this work is the modeling of the blood flow in arterioles. The structure is modeled by a non-linear Koiter shell, without bending. In addition, the presence of active elastic fibers is considered. The structure is lumped into the boundary condition of the fluid problem leading to a generalized Robin boundary condition. A finite elements discretization is proposed and several numerical test cases are presented to assess the properties of the method.

In [45] a reduced order modeling method is investigated to deal with 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. An online update is also introduced in order to guarantee stability and robustness. Several 3D fluid-fluid and fluid-structure couplings are presented as numerical experiments.

In [44] two new numerical methods for incompressible fluid/thin-walled structure interaction problems using unfitted meshes are proposed. The spatial discretization is based on different variants of Nitsche's method with cut elements. The degree of fluid-solid splitting (semi-implicit or explicit) is given by the order in which the space and time discretizations are performed. For the semi-implicit schemes, energy-based stability and a priori error estimates are derived and which guarantee the unconditional stability and optimal accuracy in the energy-norm of one the methods. Stability and a priori error estimates are also derived for one of the explicit schemes. Numerical experiments in a benchmark illustrate the performance of the different methods proposed.

7.2. Numerical methods for biological flows

Participants: Chloé Audebert, Jean-Frédéric Gerbeau, Céline Grandmont, Sanjay Pant, Marc Thiriet, Irene Vignon-Clementel.

In [16], we present a new approach for the outflow boundary conditions of Navier-Stokes equations in hemodynamics that consists in adding a 3D artificial part where the Navier-Stokes equations are modified to obtain an equivalent energy balance to a standard coupling with a 3-element Windkessel model. We investigate theoretically the stability of the system and compare it to previously introduced methods. We compare these coupling methods for numerical simulations of blood flow in three patient-specific models, which represent different flow regimes in the pulmonary and systemic circulations.

In [36], we highlight and present solutions to several challenges of the UKF method, a data-assimilation method, pertinent to reduced models of cardiovascular haemodynamics. These include methods to a) avoid ill-conditioning of covariance matrix; b) handle a variety of measurement types; c) include a variety of prior knowledge in the method; and d) incorporate measurements acquired at different heart-rates, a common situation in the clinic where patient-state differs between various clinical acquisitions.

In [18], we introduce a kinetic scheme to solve the 1D Euler equations of hemodynamics, which solution on several benchmark tests for both arterial and venous wall laws compares well with the literature. In particular, it is shown that it has a good behavior when the section area of a vessel is close to zero, which is an important property for collapsible or clamped vessels. The application to liver surgery shows that a closed-loop model of the global circulation, including 0D and 1D equations, is able to reproduce the change of waveforms observed after different levels of hepatectomy.

In [17], we explain with a 0D closed-loop lumped model the hemodynamics changes observed during partial hepatectomy in pigs [22]. The typical increase of portal pressure, increase of liver pressure loss, slight decrease of portal flow and major decrease in arterial flow are quantitatively captured by the model for a 75% hepatectomy. The different post-operative states, observed in experiments, are reproduced with the proposed model. Thus, an explanation for inter-subjects post-operative variability is proposed. This work needs to be translated to humans, in which liver flow modulation is a subject of surgery research [39].

In [24], we propose a computational approach for efficient design study of a reducer stent to be percutaneously implanted in enlarged right ventricular outflow tracts (RVOT) of repaired Tetralogy of Fallot. Hemodynamics of different designs are simulated in the stented RVOT via a reduce order model based on proper orthogonal decomposition on a reference device configuration. To validate the approach, forces exerted on the valve and on the reducer are monitored, varying with geometrical parameters, and compared with the results of full CFD simulations.

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 [38], a proof of concept study, a constant shear stress hypothesis and structured pulmonary trees are used to derive adaptive outflow boundary conditions for 3D-0D postoperative blood flow simulations. This strategy provides better predictions of pulmonary flow distribution than the conventional strategy of maintaining outflow boundary conditions.

In [26] the effect of inserted needle on the subcutaneous interstitial flow is studied. The goal is to describe the physical stress affecting cells during acupuncture needling. The convective Brinkman equations are considered to describe the flow through a fibrous medium. Three-dimensional simulations are carried out by employing an ALE finite element model. Numerical studies illustrate the acute physical stress developed by the implantation of a needle.

In [32], a fully three-dimensional blood flow simulation through a complete rigid macrovascular circuit, namely the intracranial venous network, instead of a reduced order simulation and partial vascular network is presented. The biomechanical modeling step is carefully analyzed and leads to the description of the flow governed by the dimensionless Navier-Stokes equations for an incompressible viscous fluid. The equations are then numerically solved with a free finite element software using five meshes of a realistic geometry obtained from medical images to prove the feasibility of the pipeline. Some features of the intracranial venous circuit in the supine position such as asymmetric behavior in merging regions are discussed.

7.3. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi, Fabien Raphel, Eliott Tixier.

In [51] the variability of phenomena described by parametric partial differential equations is studied. In particular, given population statistics on a system observables, 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 and compare it to other existing techniques.

In [50] a reduced order modeling method is proposed in order to speed-up the solution of reaction diffusion equations. It is based on the Approximated Lax Pair method, the discretisation is carried out by adopting an empirical interpolation framework in order to deal with non-polynomial nonlinearities. Some numerical examples on the FKPP equations as well as the equations in electrophysiology are proposed.

We published in [25] a discussion about the Comprehensive in vitro Proarrhythmia Assay (CiPA), which is a nonclinical Safety Pharmacology paradigm for discovering electrophysiological mechanisms that are likely to confer proarrhythmic liability to drug candidates intended for human use. In particular, we presented the use of mathematical modeling in Safety Pharmacology to better understand the electric signals acquired by multielectrode arrays.

7.4. Lung and respiration modeling

Participants: Laurent Boudin, Muriel Boulakia, Céline Grandmont, Nicolas Pozin, Irene Vignon-Clementel.

In [46], we proved the existence of global weak solutions to the incompressible Navier-Stokes-Vlasov system in a three-dimensional time-dependent domain with absorption boundary conditions for the kinetic part. This model arises from the study of respiratory aerosol in the human airways. The proof is based on a regularization and approximation strategy designed for our time-dependent framework.

In [52] we develop a lung-ventilation model. The parenchyma is described as an elastic homogenized media, irrigated by the tracheo-bronchial tree, a nonlinear resistive pipe network. Both are strongly coupled, and an efficient algorithm that takes advantage of the tree dyadic structure is proposed. This framework is used with different types of boundary conditions, including a nonlinear Robin model of the surrounding lung structures, to exhibit global and local coupling effects, for various ventilations. The model is also compared to a more classical exit-compartment (0D) approach.

In [34], we present a new framework that is designed to simulate ventilation and particle fate throughout the respiration cycle, both difficult to dynamically image. The flow and the particle transport and deposition models in the main bronchi are coupled to 1D models that account for the distal lobar lung structures. This enables modeling of inspiration as well as expiration. This leads to differentiated particle deposition over time, and between lobes and generations. Strong agreement to previously collected regional rat experimental data is shown, as the 1D models account for lobe-dependent morphology.

7.5. Miscellaneous

Participants: Laurent Boudin, Jean-Frédéric Gerbeau, Damiano Lombardi, Sanjay Pant, Marina Vidrascu, Irene Vignon-Clementel.

In [47], we derive the Maxwell-Stefan formalism from the Boltzmann equation for mixtures with general cross-sections. The derivation uses the Hilbert asymptotic method for systems at low Knudsen and Mach numbers. We also formally prove that the Maxwell-Stefan coefficients can be linked to the direct linearized Boltzmann operator for mixtures. That allows to compute the values of the Maxwell-Stefan diffusion coefficients with explicit and simple formulae with respect to the cross-sections. We also justify the specific ansatz we use thanks to the so-called moment method.

In [19] we give a presentation of the mathematical and numerical treatment of plate dynamics problems including rotational inertia. The presence of rotational inertia in the equation of motion makes the study of such problems interesting. We employ HCT finite elements for space discretization and the Newmark method for time discretization in FreeFEM++, and test such methods in some significant cases: a circular plate clamped all over its lateral surface, a rectangular plate simply supported all over its lateral surface, and an L-shaped clamped plate.

In [31] we investigated a modified k-nearest neighbors method to assess the differential entropy of a probability density distribution given a set of samples. Instead of considering a classical Kozachenko-Leonenko approximation, an improved parametric gaussian representation is proposed. The method aims at improving the performances of the classical estimator when considering the probability density distribution of model observations, which are featured by a strong anisotropy or functional dependency.

In [49] a dynamical adaptive tensor method is proposed to build parsimonius discretisations for systems whose domain can be naturally decomposed as a product of sets. A modified Proper Generalised Decomposition step is introduced, that allows to project the equations residual on a tensorised space. Contrary to the majority of the methods proposed, the tensor rank is adapted to guarantee a chosen precision. The method is applied to the Vlasov-Poisson system of equations. In order to preserve the hamiltonian structure of the problem, a symplectic integrator is proposed. The convergence of the method is proved and several high-dimensional test-cases are presented in order to validate the approach.