RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2018

Section New Results

Edition: 2019-03-07

ABS Project-Team

5. New Results

5.1. Modeling interfaces and contacts

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

5.1.1. Origin of Public Memory B Cell Clones in Fish After Antiviral Vaccination

Participants: F. Cazals, S. Marillet.

In collaboration with S. Magadan, L. Jouneau, S. Marillet, P. Boudinot (INRA, Virologie et Immunologie Moléculaires, Université Paris-Saclay, Jouy-en-Josas, France); M. Puelma Touzel, T. Mora, A. Walczak (Laboratoire de Physique Théorique, CNRS, Sorbonne Université, and Ecole Normale Supérieure (PSL), Paris, France); W. Chaara, A. Six (Sorbonne Université, INSERM, UMR S 959, Immunology-Immunopathology - Immunotherapy (I3), Paris, France); E. Quillet (INRA, Génétique Animale et Biologie Intégrative, Université Paris-Saclay, Jouy-en-Josas, France); O. Sunyer (Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, United States); S. Fillatreau (INEM, INSERM U1151/CNRS UMR8253, Institut Necker-Enfants Malades, Faculté de Médecine Paris Descartes, Paris, France; Faculté de Médecine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France; Assistance Publique des Hopitaux de Paris (AP-HP), Hopital Necker Enfants Malades, Paris, France).

Vaccination induces *public* antibody clonotypes common to all individuals of a species, that may mediate universal protection against pathogens. Only few studies tried to trace back the origin of these public B-cell clones. Here [16] we used Illumina sequencing and computational modeling to unveil the mechanisms shaping the structure of the fish memory antibody response against an attenuated Viral Hemorrhagic Septicemia rhabdovirus. After vaccination, a persistent memory response with a public VH5JH5 IgM component was composed of dominant antibodies shared among all individuals. The rearrangement model showed that these public junctions occurred with high probability indicating that they were already favored before vaccination due to the recombination process, as shown in mammals. In addition, these clonotypes were in the naive repertoire associated with larger similarity classes, composed of junctions differing only at one or two positions by amino acids with comparable properties. The model showed that this property was due to selective processes exerted between the recombination and the naive repertoire. Finally, our results showed that public clonotypes greatly expanded after vaccination displayed several VDJ junctions differing only by one or two amino acids with similar properties, highlighting a convergent response. The fish public memory antibody response to a virus is therefore shaped at three levels: by recombination biases, by selection acting on the formation of the pre-vaccination repertoire, and by convergent selection of functionally similar clonotypes during the response. We also show that naive repertoires of IgM and IgT have different structures and sharing between individuals, due to selection biases. In sum, our comparative approach identifies three conserved features of the antibody repertoire associated with public memory responses. These features were already present in the last common ancestors of fish and mammals, while other characteristics may represent species-specific solutions.

5.2. Modeling macro-molecular assemblies

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

5.2.1. Complexity Dichotomies for the Minimum F-Overlay Problem – Application for low resolution models of macro-molecular assemblies

Participant: D. Mazauric.

In collaboration with N. Cohen (CNRS, Laboratoire de Recherche en Informatique) and F. Havet (CNRS, Inria/I3S project-team Coati) and I. Sau (CNRS, Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier) and R. Watrigant (University Lyon I, Laboratoire de l'Informatique du Parallélisme).

In this article [14], we analyze a generalization of the minimum connectivity inference problem (MCI). MCI models the computation of low-resolution structures of macro-molecular assemblies, based on data obtained by native mass spectrometry. The generalization studied in this article, allows us to consider more refined constraints for the characterization of low resolution models of large assemblies. We model this problem by using hypergraphs: for a (possibly infinite) fixed family of graphs F, we say that a graph $GoverlaysF$ on a hypergraph H if $V(H)$ is equal to $V(G)$ and the subgraph of G induced by every hyperedge of H contains some member of F as a spanning subgraph. While it is easy to see that the complete graph on $|V(H)|$ overlays F on a hypergraph H whenever the problem admits a solution, the Minimum F-Overlay problem asks for such a graph with at most k edges, for some given $k \in \mathbb{N}$. This problem allows to generalize some natural problems which may arise in practice. For instance, if the family F contains all connected graphs, then Minimum F-Overlay corresponds to the MCI problem. Our main contribution is a strong dichotomy result regarding the polynomial vs. NP-complete status with respect to the considered family F. Roughly speaking, we show that the easy cases one can think of (e.g. when edgeless graphs of the right sizes are in F, or if F contains only cliques) are the only families giving rise to a polynomial problem: all others are NP-complete. We then investigate the parameterized complexity of the problem and give similar sufficient conditions on F that give rise to $W[1]$-hard, $W[2]$-hard or FPT problems when the parameter is the size of the solution.

5.3. Modeling the flexibility of macro-molecules

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

5.3.1. Characterizing molecular flexibility by combining lRMSD measures

Participants: F. Cazals, R. Tetley.

The root mean square deviation (RMSD) and the least RMSD are two widely used similarity measures in structural bioinformatics. Yet, they stem from global comparisons, possibly obliterating locally conserved motifs. We correct these limitations with the so-called combined RMSD [26], which mixes independent lRMSD measures, each computed with its own rigid motion. The combined RMSD is relevant in two main scenarios, namely to compare (quaternary) structures based on motifs defined from the sequence (domains, SSE), and to compare structures based on structural motifs yielded by local structural alignment methods. We illustrate the benefits of combined RMSD over the usual RMSD on three problems, namely (i) the assignment of quaternary structures for hemoglobin (scenario #1), (ii) the calculation of structural phylogenies (case study: class II fusion proteins; scenario #1), and (iii) the analysis of conformational changes based on combined RMSD of rigid structural motifs (case study: one class II fusion protein; scenario #2). Using these, we argue that the combined RMSD is a tool a choice to perform positive and negative discrimination of degree of freedom, with applications to the design of move sets and collective coordinates. Combined RMSD are available within the Structural Bioinformatics Library (http: //sbl.inria.fr).

5.3.2. Multiscale analysis of structurally conserved motifs

Participants: F. Cazals, R. Tetley.

This work [25] develops a generic framework to perform a multiscale structural analysis of two structures (homologous proteins, conformations) undergoing conformational changes. Practically, given a seed structural alignment, we identify structural motifs with a hierarchical structure, characterized by three unique properties. First, the hierarchical structure sheds light on the trade-off between size and flexibility. Second, motifs can be combined to perform an overall comparison of the input structures in terms of combined RMSD, an improvement over the classical least RMSD. Third, motifs can be used to seed iterative aligners, and to design hybrid sequence-structure profile HMM characterizing protein families. From the methods standpoint, our framework is reminiscent from the bootstrap and combines concepts from rigidity analysis (distance difference matrices), graph theory, computational geometry (space filling diagrams), and topology (topological

persistence). On challenging cases (class II fusion proteins, flexible molecules) we illustrate the ability of our tools to localize conformational changes, shedding light of commonalities of structures which would otherwise appear as radically different. Our tools are available within the Structural Bioinformatics Library (http://sbl.inria.fr). We anticipate that they will be of interest to perform structural comparisons at large, and for remote homology detection.

5.3.3. Hybrid sequence-structure based HMM models leverage the identification of homologous proteins: the example of class II fusion proteins

Participants: F. Cazals, R. Tetley.

In collaboration with P. Guardado-Calvo, J. Fedry, and F. Rey (Inst. Pasteur Paris, France).

In [27], we present a sequence-structure based method characterizing a set of functionally related proteins exhibiting low sequence identity and loose structural conservation. Given a (small) set of structures, our method consists of three main steps. First, pairwise structural alignments are combined with multi-scale geometric analysis to produce structural motifs i.e. regions structurally more conserved than the whole structures. Second, the sub-sequences of the motifs are used to build profile hidden Markov models (HMM) biased towards the structurally conserved regions. Third, these HMM are used to retrieve from UniProtKB proteins harboring signatures compatible with the function studied, in a bootstrap fashion. We apply these hybrid HMM to investigate two questions related to class II fusion proteins, an especially challenging class since known structures exhibit low sequence identity (less than 15%) and loose structural similarity (of the order of 15A in lRMSD). In a first step, we compare the performances of our hybrid HMM against those of sequence based HMM. Using various learning sets, we show that both classes of HMM retrieve unique species. The number of unique species reported by both classes of methods are comparable, stressing the novelty brought by our hybrid models. In a second step, we use our models to identify 17 plausible HAP2-GSC1 candidate sequences in 10 different drosophila melanogaster species. These models are not identified by the PFAM family HAP2-GCS1 (PF10699), stressing the ability of our structural motifs to capture signals more subtle than whole Pfam domains. In a more general setting, our method should be of interest for all cases functional families with low sequence identity and loose structural conservation. Our software tools are available from the FunChaT package of the Structural Bioinormatics Library (http://sbl.inria.fr).

5.3.4. Hamiltonian Monte Carlo with boundary reflections, and application to polytope volume calculations

Participants: F. Cazals, A. Chevallier.

In collaboration with S. Pion (Auctus, Inria Bordeaux).

This paper [23] studies HMC with reflections on the boundary of a domain, providing an enhanced alternative to Hit-and-run (HAR) to sample a target distribution in a bounded domain. We make three contributions. First, we provide a convergence bound, paving the way to more precise mixing time analysis. Second, we present a robust implementation based on multi-precision arithmetic – a mandatory ingredient to guarantee exact predicates and robust constructions. Third, we use our HMC random walk to perform polytope volume calculations, using it as an alternative to HAR within the volume algorithm by Cousins and Vempala. The tests, conducted up to dimension 50, show that the HMC RW outperforms HAR.

5.3.5. Wang-Landau Algorithm: an adapted random walk to boost convergence

Participants: F. Cazals, A. Chevallier.

The Wang-Landau (WL) algorithm is a recently developed stochastic algorithm computing densities of states of a physical system. Since its inception, it has been used on a variety of (bio-)physical systems, and in selected cases, its convergence has been proved. The convergence speed of the algorithm is tightly tied to the connectivity properties of the underlying random walk. As such, we propose in [22] an efficient random walk that uses geometrical information to circumvent the following inherent difficulties: avoiding overstepping strata, toning down concentration phenomena in high-dimensional spaces, and accommodating multidimensional distribution. Experiments on various models stress the importance of these improvements to

make WL effective in challenging cases. Altogether, these improvements make it possible to compute density of states for regions of the phase space of small biomolecules.

5.4. Algorithmic foundations

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

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

5.4.1. A Sequential non-parametric multivariate two-sample test

Participant: F. Cazals.

In collaboration with A. Lhéritier (Amadeus, France).

Given samples from two distributions, a nonparametric two-sample test aims at determining whether the two distributions are equal or not, based on a test statistic. Classically, this statistic is computed on the whole dataset, or is computed on a subset of the dataset by a function trained on its complement. We consider methods in a third tier [15], so as to deal with large (possibly infinite) datasets, and to automatically determine the most relevant scales to work at, making two contributions. First, we develop a generic sequential nonparametric testing framework, in which the sample size need not be fixed in advance. This makes our test a truly sequential nonparametric multivariate two-sample test. Under information theoretic conditions qualifying the difference between the tested distributions, consistency of the two-sample test is established. Second, we instantiate our framework using nearest neighbor regressors, and show how the power of the resulting two-sample test can be improved using Bayesian mixtures and switch distributions. This combination of techniques yields automatic scale selection, and experiments performed on challenging datasets show that our sequential tests exhibit comparable performances to those of state-of-the-art nonsequential tests.

5.4.2. Comparing two clusterings using matchings between clusters of clusters

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

In collaboration with R. Watrigant (University Lyon I, Laboratoire de l'Informatique du Parallélisme, France).

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. We go beyond these by providing a novel class of methods computing meta-clusters within each clustering–a meta-cluster is a group of clusters, together with a matching between these. Let the intersection graph of two clusterings be the edge-weighted bipartite graph in which the nodes represent the clusters, the edges represent the non empty intersection between two clusters, and the weight of an edge is the number of common items. We introduce the so-called D-family-matching problem on intersection graphs, with D the upper-bound on the diameter of the graph induced by the clusters of any meta-cluster. First we prove NP-completeness and APX-hardness results, and unbounded approximation ratio of simple strategies. Second, we design exact polynomial time dynamic programming algorithms for some classes of graphs (in particular trees). Then, we prove spanning-tree based efficient algorithms for general graphs. Our experiments illustrate the role of D as a scale parameter providing information on the relationship between clusters within a clustering and in-between two clusterings. They also show the advantages of our built-in mapping over classical cluster comparison measures such as the variation of information (VI).

5.4.3. How long does it take for all users in a social network to choose their communities?

Participant: D. Mazauric.

In collaboration with J.-C. Bermond (Inria/I3S project-team Coati) and A. Chaintreau (Columbia University in the city of New York) and G. Ducoffe (National Institute for Research and Development in Informatics and Research Institute of the University of Bucharest).

We consider a community formation problem in social networks, where the users are either friends or enemies. The users are partitioned into conflict-free groups (*i.e.*, independent sets in the *conflict graph*$G^- = (V, E)$ that represents the enmities between users). The dynamics goes on as long as there exists any set of at most k users, k being any fixed parameter, that can change their current groups in the partition *simultaneously*, in such a way that they all strictly increase their utilities (number of friends *i.e.*, the cardinality of their respective groups minus one). Previously, the best-known upper-bounds on the maximum time of convergence were $\mathcal{O}(|V|\alpha(G^-))$ for $k \leq 2$ and $\mathcal{O}(|V|^3)$ for $k = 3$, with $\alpha(G^-)$ being the independence number of G^-. Our first contribution in this paper consists in reinterpreting the initial problem as the study of a dominance ordering over the vectors of integer partitions. With this approach, we obtain for $k \leq 2$ the tight upper-bound $\mathcal{O}(|V| \min\{\alpha(G^-), \sqrt{|V|}\})$ and, when G^- is the empty graph, the exact value of order $\frac{(2|V|)^{3/2}}{3}$. The time of convergence, for any fixed $k \geq 4$, was conjectured to be polynomial [35], [41]. In this paper we disprove this. Specifically, we prove that for any $k \geq 4$, the maximum time of convergence is an $\Omega(|V|^{\Theta(\log |V|)})$.

See [19] for details.

5.4.4. Sequential metric dimension

Participant: D. Mazauric.

In collaboration with J. Bensmail (I3S, Inria/I3S project-team Coati) and F. Mc Inerney (Inria/I3S project-team Coati) and N. Nisse (Inria, Inria/I3S project-team Coati) and S. Pérennes (CNRS, Inria/I3S project-team Coati).

In the localization game, introduced by Seager in 2013, an invisible and immobile target is hidden at some vertex of a graph G. At every step, one vertex v of G can be probed which results in the knowledge of the distance between v and the secret location of the target. The objective of the game is to minimize the number of steps needed to locate the target whatever be its location.

We address the generalization of this game where $k \geq 1$ vertices can be probed at every step. Our game also generalizes the notion of the *metric dimension* of a graph. Precisely, given a graph G and two integers $k, \ell \geq 1$, the *localization* problem asks whether there exists a strategy to locate a target hidden in G in at most ℓ steps and probing at most k vertices per step. We first show that, in general, this problem is NP-complete for every fixed $k \geq 1$ (resp., $\ell \geq 1$). We then focus on the class of trees. On the negative side, we prove that the localization problem is NP-complete in trees when k and ℓ are part of the input. On the positive side, we design a $(+1)$-approximation for the problem in n-node trees, *i.e.*, an algorithm that computes in time $O(n \log n)$ (independent of k) a strategy to locate the target in at most one more step than an optimal strategy. This algorithm can be used to solve the localization problem in trees in polynomial time if k is fixed.

We also consider some of these questions in the context where, upon probing the vertices, the relative distances to the target are retrieved. This variant of the problem generalizes the notion of the *centroidal dimension* of a graph.

See [17], [18], [21] for details.

BEAGLE Project-Team

7. New Results

7.1. Dopamine interacts with endocannabinoids to regulate spike timing dependent plasticity

participants: H. Berry, I. Prokin

Dopamine modulates striatal synaptic plasticity, a key substrate for action selection and procedural learning. Thus, characterizing the repertoire of activity-dependent plasticity in striatum and its dependence on dopamine is of crucial importance. In collaboration with L. Venance Lab (CIRB, Collège de France) we recently unraveled a striatal spike-timing-dependent long-term potentiation (tLTP) mediated by endocannabinoids (eCBs) and induced with few spikes (5-15). Whether this eCB-tLTP interacts with the dopaminergic system remains to be investigated. We found that eCB-tLTP is impaired in a rodent model of Parkinson's disease and rescued by L-DOPA. Dopamine controls eCB-tLTP via dopamine type-2 receptors (D2R) located presynaptically in cortical terminals. Dopamine-endocannabinoid interactions via D2R are required for the emergence of tLTP in response to few coincident pre-and post-synaptic spikes and control eCB-plasticity by modulating the long-term potentiation (LTP)/depression (LTD) thresholds. While usually considered as a depressing synaptic function, our results show that eCBs in the presence of dopamine constitute a versatile system underlying bidirectional plasticity implicated in basal ganglia pathophysiology. These results have been published in Nature Communications [23]

7.2. Estimating the robustness of spike timing dependent plasticity to timing jitter

participants: H. Berry, I. Prokin

In Hebbian plasticity, neural circuits adjust their synaptic weights depending on patterned firing. Spike-timing-dependent plasticity (STDP), a synaptic Hebbian learning rule, relies on the order and timing of the paired activities in pre- and postsynaptic neurons. Classically, in ex vivo experiments, STDP is assessed with deterministic (constant) spike timings and time intervals between successive pairings, thus exhibiting a regularity that differs from biological variability. Hence, STDP emergence from noisy inputs as occurring in in vivo-like firing remains unresolved. In collaboration with the laboratories of L. Venance (CIRB, Collège de France) and A. De Kerchove d'Exaerde (Univ. Libre Bruxelles), we used noisy STDP pairings where the spike timing and/or interval between pairings were jittered. We explored with electrophysiology and mathematical modeling, the impact of jitter on three forms of STDP at corticostriatal synapses: NMDAR-LTP, endocannabinoid-LTD and endocannabinoid-LTP. We found that NMDAR-LTP was highly fragile to jitter, whereas endocannabinoid-plasticity appeared more resistant. When the frequency or number of pairings was increased, NMDAR-LTP became more robust and could be expressed despite strong jittering. Our results identify endocannabinoid-plasticity as a robust form of STDP, whereas the sensitivity to jitter of NMDAR-LTP varies with activity frequency. This provides new insights into the mechanisms at play during the different phases of learning and memory and the emergence of Hebbian plasticity in in vivo-like activity. These results have been published in Scientific Reports [14]

7.3. A new method to monitor gap junctional comunication in astrocytes

participants: H. Berry

Intercellular communication through gap junction channels plays a key role in cellular homeostasis and in synchronizing physiological functions, a feature that is modified in number of pathological situations. In the brain, astrocytes are the cell population that expresses the highest amount of gap junction proteins, named connexins. Several techniques have been used to assess the level of gap junctional communication in astrocytes, but so far they remain very difficult to apply in adult brain tissue. Using specific loading of astrocytes with sulforhodamine 101, we adapted in collaboration with C. Giaume's laboratory (CIRB, Collège de France) the gap-FRAP (Fluorescence Recovery After Photobleaching) to acute hippocampal slices from 9 month-old adult mice. We show that gap junctional communication monitored in astrocytes with this technique was inhibited either by pharmacological treatment with a gap junctional blocker or in mice lacking the two main astroglial connexins, while a partial inhibition was measured when only one connexin was knocked-out. We validate this approach using a mathematical model of sulforhodamine 101 diffusion in an elementary astroglial network and a quantitative analysis of the exponential fits to the fluorescence recovery curves. Consequently, we consider that the adaptation of the gap-FRAP technique to acute brain slices from adult mice provides an easy going and valuable approach that allows overpassing this age-dependent obstacle and will facilitate the investigation of gap junctional communication in adult healthy or pathological brain. These results have been published in J. Neuroscience Methods [24].

7.4. Kir4.1 upregulation in astrocytes of the lateral habenula is involved in depression

participants: H. Berry, A. Foncelle

Enhanced bursting activity of neurons in the lateral habenula (LHb) is essential in driving depression-like behaviours, but the cause of this increase has been unknown. In collaboration with H. Hu's laboratory (Zhejiang University, China), using a high-throughput quantitative proteomic screen, we show that an astroglial potassium channel (Kir4.1) is upregulated in the LHb in rat models of depression. Kir4.1 in the LHb shows a distinct pattern of expression on astrocytic membrane processes that wrap tightly around the neuronal soma. Electrophysiology and modelling data show that the level of Kir4.1 on astrocytes tightly regulates the degree of membrane hyperpolarization and the amount of bursting activity of LHb neurons. Astrocyte-specific gain and loss of Kir4.1 in the LHb bidirectionally regulates neuronal bursting and depression-like symptoms. Together, these results show that a glia–neuron interaction at the perisomatic space of LHb is involved in setting the neuronal firing mode in models of a major psychiatric disease. Kir4.1 in the LHb might have potential as a target for treating clinical depression. These results have been published in Nature [15] and were commented in the "News and views" section of the journal: Howe WM and Kenny PJ (2018). Burst firing sets the stage for depression.

7.5. The evolutionary complexity ratchet

participants: G Beslon, V Liard, D Parsons, Jonathan Rouzaud-Cornabas

Using the *in silico* experimental evolution platform Aevol, we evolved populations of digital organisms in conditions where a simple functional structure is best.

Strikingly, we observed that in a large fraction of the simulations, organisms evolved a complex functional structure and that their complexity increased during evolution despite being a lot less fit than simple organisms in other populations. However, when submitted to a harsh mutational pressure, we observed that a significant proportion of complex individuals ended up with a simple functional structure.

Our results suggest the existence of a complexity ratchet that is powered by epistasis and that cannot be beaten by selection. They also show that this ratchet can be overthrown by robustness because of the strong constraints it imposes on the coding capacity of the genome.

This result has been published in the International conference ALife in Tokyo (July 2018) where it received the best paper award [28]

7.6. Weight-based search to find clusters around medians in subspaces

participants: C Rigotti, G Beslon

There exist several clustering paradigms, leading to different techniques that are complementary in the analyst toolbox, each having its own merits and interests. Among these techniques, the K-medians approach is recognized as being robust to noise and outliers, and is an important optimization task with many different applications (e.g., facility location). In the context of subspace clustering, several paradigms have been investigated (e.g., centroid-based, cell-based), while the median-based approach has received less attention. Moreover, using standard subspace clustering outputs (e.g., centroids, medoids) there is no straightforward procedure to compute the cluster membership that optimizes the dispersion around medians. We advocated for the use of median-based subspace clustering as a complementary tool. Indeed, we showed that such an approach exhibits satisfactory quality clusters when compared to well-established paradigms, while medians have still their own interests depending on the user application (robustness to noise/outliers and location optimality). We showed that a weight-based hill climbing algorithm using a stochastic local exploration step can be sufficient to produce the clusters.

This research has been published in the proceedings of the ACM-SAC conference (Pau, March 2018) where it received the best paper award [26].

7.7. The surprising creativity of digital evolution

participants: C Knibbe, G Beslon

Natural evolution is a creative fount of complex adaptations that often surprise the scientists who discover them. However, the creativity of evolution is not limited to the natural world; artificial organisms evolving in computational environments are also able to elicit a similar degree of surprise and wonder from the researchers studying them. The process of evolution has proven to be an algorithmic process that transcends the substrate to which it is applied. Indeed, most digital evolution researchers can relate anecdotes highlighting how common it is for their algorithms to creatively subvert their expectations or intentions, expose unrecognized bugs in their code, produce unexpectedly potent adaptations, or engage in behaviors and outcomes uncannily convergent with ones found in nature. Such stories routinely reveal surprise and creativity by evolution in these digital worlds, but they rarely fit into the standard scientific narrative and are thus often treated as obstacles to be overcome rather than results that warrant publication in their own right. Bugs are fixed, experiments are refocused, one-off surprises are collapsed 1into a single data point. The stories themselves are traded among researchers through oral tradition, but that mode of information transmission is lossy, inefficient and error-prone. Moreover, the very fact that these stories tend to be confined to practitioners means that many natural scientists do not recognize how lifelike digital organisms are and how natural their evolution can be. We actively participated to a crowd-sourced research in which evolutionary computation researchers providing first-hand reports of such cases, and thus functions as a written, fact-checked collection of entertaining and important stories.

7.8. HPC support for Aevol

participants: Jonathan Rouzaud-Cornabas, David Parsons, Guillaume Beslon

During the year, we had three internships that focus around HPC. The three of them were founded through the Federation Informatique de Lyon (FIL FR2000) and were common between the Inria Beagle team (LIRIS) and the Inria Avalon team (LIP).

The first one (Lukas Schmidt - M2) was working on component-based software engineering and HPC with Aevol as use-case. The goal was to see if and how the COMET [1] task-based parallel component model (and its implementation Halley) can fit the parallelization requirement of Aevol. An extension of the model was proposed to support hierarchical data structure and a prototype implementation has been done. In the future, we will work on the formalization of the extension and an efficient implementation on it. The goal is to ease the development and replacement of core components of the Aevol software (e.g. be able to easily replace the 2-base DNA code by a 4-base one).

The second internship (Valentin Huguet - M2) was evolving around Aevol and how to ease the distribution of the computation. To do so, an extension of the DIET software [2] was proposed and a fully functional webboard was implemented. We have a first prototype that support the execution of a large set of distributed computing resources and the control of its execution through a webboard. Moreover, basic visualization of the simulation results can be done through the same webboard. A following internship (starting Feb. 2019) will continue the work. The goal is to support workflow composed of multiple execution of Aevol and its pre/post treatments to automate the execution of large campaign that are done manually at the moment.

The goal of the third internship (Nathan Payre - L3) was to propose a prototype of a bitset for Aevol and its efficient implementation on modern hardware (Intel Skylake and Intel Xeon Phi). Indeed, the current implementation of Aevol DNA (2 base) uses a char type (8bit) to store a bit value (0 or 1). Accordingly, working at the bitset level could save up to 8 time memory space and speed up the computation (as Aevol is memory bound, reducing the memory transfer by 8 could dramatically speed up the global execution). Moreover, modern processors have vectorization extension that are perfectly fitting our requirements (we could process 512bit per cycle with AVX512 extension). During the internship, the bitset and the different operation we use in Aevol model (e.g. Hamming distance) were formalized and implemented. The preliminary results show a speed up of 140x on these operations. A full evaluation on the impact of the performance of Aevol and how different modern processor react to such implementation will be done in the future.

Last, a part of the Beagle team (Guillaume Beslon, David Parsons, Jonathan Rouzaud-Cornabas) were selected and participate to the EuroHack 2018 GPU Programming Hackathon in Lugano (Switzerland) organized by CSCS (Swiss National Supercomputing Centre) and NVidia. The goal was to port Aevol to modern GPU and thus to the CUDA programming language. In order to be able to do so in a week, we propose a mini-application (mini-Aevol) of Aevol [3] that is representative of the computation and memory pattern of the full Aevol. This prototype will be reuse in our collaboration with team focusing on HPC research. At the end of the week, we had a full implementation of mini-Aevol on GPU. New core algorithms of Aevol have been proposed to support massively parallel processors such as GPU. The prototype will be transfer to the full Aevol code in the future to be able to support GPU. It is worth noting that this mini-apps is also used in teaching context (INSA Lyon - Computer Science M2) to learn how to parallelize and optimize code with OpenMP and CUDA.

[1] Olivier Aumage, Julien Bigot, Hélène Coullon, Christian Pérez, Jérôme Richard. Combining Both a Component Model and a Task-based Model for HPC Applications: a Feasibility Study on GYSELA. 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)., May 2017, Madrid, Spain.

[2] https://graal.ens-lyon.fr/diet/

[3] https://gitlab.inria.fr/rouzaudc/mini-aevol

[4] https://github.com/fomics/EuroHack18/blob/master/final/beagle_aevol.pdf

7.9. Exploring the evolution of chromatin-associated proteins

participants: A Crombach

Eukaryotic gene regulation depends strongly on chromatin state. High-throughput studies in the fruit fly Drosophila melanogaster have shown that instead of the canonical two types of chromatin, hetero- and euchromatin, one can subdivide chromatin into five states. These states are each characterized by a unique combination of chromatin-associated proteins (CAPs). We were interested in the evolution of CAPs and studied them by means of phylogenomic methods. We found three evolutionary trends. One type of heterochomatin, called GREEN, is specific to centromeres and some of its proteins are found to be under a Red Queen type evolution, where they rapidly accumulate amino acid changes. The second type of heterochromatin, BLUE, is tightly linked to Polycomb Group proteins. These proteins are important regulators in developmental processes and our findings confirm their origin in multicellular organisms. Finally, the two euchromatic types, YELLOW and RED, have strong lineage-specific characteristics. Their origins seem to date back to the start of eukaryotic life.

7.10. Evolutionary interplay of genome content and 3D spatial structure

participants: A Crombach
Genomes are hierarchically folded, which involves transposable elements (TEs). The most prominently observed folding domains are conserved between cell types and across species, yet their building blocks, TEs, are powerful mutagens. This paradox raises the question why we observe evolutionary stable folding domains. Using in silico evolution of polymer genomes, the aim is to elucidate the interplay between mutations and folding structure. We have built the software and are in the process of generating data. First results indicate that due to accessibility in 3D (some parts of the genome are more tightly compacted than others), a positive feedback is created between (1) where mutations happen, (2) how genome content is changed, and (3) how genomes fold in 3D.

7.11. Network inference for mammalian cortex development

participants: A Crombach
The mammalian cortex divides into two major regions, neocortex (NCx) and the structurally simpler allocortex. Whereas NCx is well-characterized, the allocortex is much less studied. Its best known region is the olfactory (piriform, PCx) cortex. The regions have a laminar structure, with distinct neuronal cell types in each of the layers: NCx has 6 layers, PCx has 3. The differentiation of precursor cells into various neuronal cell types determines to which layer these cells will migrate. This process is mostly studied in NCx and depends on the activity of 10–20 developmental genes. In PCx the same genes are used, yet they appear in other combinations and may indicate diverse target layers, sometimes violating rules-of-thumb derived from NCx. Current understanding is rather incomplete with respect to how cortical neurons are specified. We propose that, despite apparent contradictions, a single gene network can explain the development of distinct cortical regions.

In collaboration with Dr. A. Fleischmann at Brown University (USA), we are measuring the expression of genes involved in neurodevelopment at cellular resolution using light-sheet microscopy. These data will form the basis for the inference of a regulatory network describing neuronal differentiation in NCx and PCx. Inference is done by fitting mathematical models of gene regulation to the data using global optimization methods. Currently, we are processing the image data. Moreover, single cell RNA sequencing will allow the study of the temporal dynamics of the expression of these genes and many others. We are completing an in-depth statistical analysis of the resulting genome-wide expression data.

7.12. Gene transfers can date the tree of life

participants: E Tannier
Biodiversity has always been predominantly microbial, and the scarcity of fossils from bacteria, archaea and microbial eukaryotes has prevented a comprehensive dating of the tree of life. Here, we show that patterns of lateral gene transfer deduced from an analysis of modern genomes encode a novel and abundant source of information about the temporal coexistence of lineages throughout the history of life. We use state-of-the-art species tree-aware phylogenetic methods to reconstruct the history of thousands of gene families and demonstrate that dates implied by gene transfers are consistent with estimates from relaxed molecular clocks in Bacteria, Archaea and Eukarya. We present the order of speciations according to lateral gene transfer data calibrated to geological time for three datasets comprising 40 genomes for Cyanobacteria, 60 genomes for Archaea and 60 genomes for Fungi. An inspection of discrepancies between transfers and clocks and a comparison with mammalian fossils show that gene transfer in microbes is potentially as informative for dating the tree of life as the geological record in macroorganisms. [16]

7.13. The devil in the details of evolvability

Participants: E Tannier, P Biller, V Liard, G Beslon

The theory of Evolvability consists in studying the evolution of living organisms as a computational learning process. It defines the possibilities of a population under Darwinian selection, to evolve in a certain direction, in a reasonable amount of time. While its robustness to certain parameters has been theoretically assessed, this theory has not been experimentally tested. We use a standard *in silico* experimental evolution tool to compare some predictions of the theory and the behavior of digital populations designed to resemble biological organisms. We obtain that the evolvability of monotone conjunctions under the uniform distribution of environmental conditions, presented as a major result of the theory, is not reproduced by the experiments. We show that this is due to different mutation algorithms, by a proof of exponential expectation time to target under theoretical conditions closer to the experiments. We examine into detail the choices of mutation algorithms. In the Evolvability theory it is any Turing machine, while much more restricted in the experimental design. This definition allows a wider range of conditions and in a certain way is conform to biological reality, where mutators evolve and can be selected. However it also allows, if it is misused, for the inclusion of oracles that are incompatible with the principles of a Darwinian evolution. Unfortunately these oracles are extensively used in the current evolvability proofs.

BIGS Project-Team

6. New Results

6.1. Stochastic modelling

Participants: A. Gégout-Petit, S. Mézières, Y. Petot, P. Vallois

In the framework of the esca-illness of vines, we developed different spatial models and spatio-temporal models for different purposes: (1) study the distribution and the dynamics of esca vines in order to tackle the aggregation and the potential spread of the illness (2) propose a spatio-temporal model in order to capture the dynamics of cases and measure the effects of environmental covariates. For purpose (2), we developed an autologistic model (centered in a new way), estimators of the parameters, and showed their good properties, and proposed a way to choose between several neighborhood models. It is the object of preprint [41].

In the framework of chalara of ashand, through a collaboration with INRA researchers, we have proposed a mechanistic model of propagation whose parameters are estimated by bayesian estimation. It is the object of the communication [26].

In a collaboration with physicists from Nancy CHRU, we have worked about the interest to use the whole distribution of telomeres lengths until the mean that is usually used to characterise ageing of a cell. We have shown that the shape of the distribution can be seen as a individuals's signature. It is the object of the accepted paper [10].

We analyse the probabilistic features of the Choquet integral with respect to a capacity over a finite set where the entries are random variables. Despite the amount of studies, the question of uncertainty remains under-considered. Such a question is of first importance in many applications and uses.

In the multifactorial context of modelling for gliomas, we focused our attention on the acquisition of the tumor diameter from clinical-collected data [1]. 3-D reconstruction via an equivalent sphere from multiple contouring of the tumor leads us to characterize its infiltrating phenotype (infiltration rate, direction of infiltration, evolution of morphology over time), current work. Our aim is to incorporate this new factor in the modeling already started (to appear in JBHI, beginning 2019).
A brain cartography obtained by sensorial simulations during awake surgery with the aid of clustering analysis is in revision.

6.2. Estimation and control for Markov Processes

Participants: R. Azaïs, F. Bouguet, A. Gégout-Petit, F. Greciet, B. Scherrer

Piecewise-deterministic Markov processes form a class of stochastic models with a sizeable scope of applications. Such processes are defined by a deterministic motion punctuated by random jumps at random times, and offer simple yet challenging models to study. The issue of statistical estimation of the parameters ruling the jump mechanism is far from trivial. Responding to new developments in the field as well as to current research interests and needs, the book "Statistical Inference for Piecewise-deterministic Markov Processes" edited by Romain Azaïs and Florian Bouguet [33] gather 7 chapters by different authors on the topic. The idea for this book stemmed from a workshop organized in Nancy in the 2016-17 winter. Two chapters [48][31] have been co-authored by one or more BIGS members.

Multiple-step lookahead policies have demonstrated high empirical competence in Reinforcement Learning, via the use of Monte Carlo Tree Search or Model Predictive Control. In [13], multiple-step greedy policies and their use in vanilla Policy Iteration algorithms were proposed and analyzed. In [14], [12], we study multiple-step greedy algorithms in more practical setups: we describe and analyze a stochastic approximation variation and general sensitivity analyses to approximations. In [15], we describe a short study on an Anderson acceleration of the fixed point computation involved in Reinforcement Learning. These contributions resulted in one publication in ICML, in NeurIPS, and two in EWRL (the European Workshop on Reinforcement Learning).

6.3. Algorithms and Estimation for graph data

Participants: A. Gégout-Petit, A. Gueudin, C. Karmann

In the purpose to deal with inference for network of zero-inflated variables, we have developped a new regression model. We consider the problem of variable selection when the response is ordinal, that is an ordered categorical variable. In particular, we are interested in selecting quantitative explanatory variables linked with the ordinal response variable and we want to determine which predictors are relevant. In this framework, we choose to use the polytomous ordinal logistic regression model using cumulative logits which generalizes the logistic regression. We then introduce the Lasso estimation of the regression coefficients using the Frank-Wolfe algorithm. To deal with the choice of the penalty parameter, we use the stability selection method and we develop a new method based on the knockoffs idea. This knockoffs method is general and suitable to any regression and besides, gives an order of importance of the covariates. Finally, we provide some experimental results to corroborate our method and we present an application of this regression method for zero-inflated network inference. This work is the object of a presentation in a conference [28] and a preprint submitted in a journal [40].

6.4. Regression and machine learning

Participants: E. Albuisson, R. Azaïs (Inria, Lyon), T. Bastogne, L. Batista, K. Duarte, S. Ferrigno, A. Gégout-Petit, P. Guyot, J.-M. Monnez, N. Sahki, S. Mézières

In the purpose to detect change of health state for lung-transplanted patient, we have begun to work on breakdowns in multivariate physiological signals. Based on the CUSUM statistics, we have used dynamical thresholds of detection [27]. A more general talk about statistical learning and connected patient was given in a workshop "Evaluation des objets en santé connectée" [35].

We consider the analysis of cardiomyocyte signals (cardiac cells) for the cardiotoxicity assessment of new pharmaceutical compounds in preclinical assays. The experimental data are either impedance signals measuring the contractility of cardiomyocytes [39], [4], field potential signals measuring their functionality or fluorescence signals measuring the activity of some ion channels such as calcium pumps (Ca2+). At this preclinical level, our main contribution is the estimation of important characteristics such the field potential duration [17] or the identification of cardiotoxic events such as the early-afterdepolarization.We have also developed new methods for the analysis of electrocardiograms at patient level and more precisely the estimation of parameters such as the RR and QT intervals in long and noisy signals provided by wearable sensors [24], [30], [23], [25], [29]. We also study the efficacy of a new biomarker in radiotherapy. The objective is to compute a score able to predict risk of radiosensitivity for patients in radiotherapy [19], [20].We are also developing a new method to characterize the potential interactions between nanoparticles and biological compounds of complex media such as blood. This new method aims at predicting risks on the biodistribution and toxicity of the nanoparticles [16], [36].

In [7], we present a methodology for constructing a short-term event risk score from an ensemble predictor using bootstrap samples, two different classification rules, logistic regression and linear discriminant analysis for mixed data, continuous or categorical, and random selections of variables into the construction of predictors. We establish a property of linear discriminant analysis for mixed data and define an event risk measure by an odds-ratio. This methodology is applied to heart failure patients on whom biological, clinical and medical history variables were measured and the results obtained from our data are detailed.

The study [8] addresses the problem of sequential least square multidimensional linear regression, particularly in the case of a data stream, using a stochastic approximation process. To avoid the phenomenon of numerical explosion which can be encountered and to reduce the computing time in order to take into account a maximum of arriving data, we propose using a process with online standardized data instead of raw data and the use of several observations per step or all observations until the current step. Herein, we define and study the almost sure convergence of three processes with online standardized data: a classical process with a variable step-size and use of a varying number of observations per step, an averaged process with a constant step-size and use of a varying number of observations per step, and a process with a variable or constant step-size and use of all observations until the current step. Their convergence is obtained under more general assumptions than classical ones. These processes are compared to classical processes on 11 datasets for a fixed total number of observations used and thereafter for a fixed processing time. Analyses indicate that the third-defined process typically yields the best results.

Many articles were devoted to the problem of estimating recursively the eigenvectors and eigenvalues in decreasing order of the expectation of a random matrix using an i.i.d. sample of it. In [43], we make the following contributions. The convergence of a normed process is proved under more general assumptions: the random matrices are not supposed i.i.d. and a new data mini-batch or all data until the current step are taken into account at each step without storing them; three types of processes are studied; this is applied to online principal component analysis of a data stream, assuming that data are realizations of a random vector Z whose expectation is unknown and must be estimated online, as well as possibly the metrics used when it depends on unknown characteristics of Z.

Let $Y = m(X) + \sigma(X)\varepsilon$ be a regression model, where $m(\cdot)$ is the regression function, $\sigma^2(\cdot)$ the variance function and ε the random error term. Methods to assess how well a model fits a set of observations fall under the banner of goodness-of-fit tests. Many tests have been developed to assess the different assumptions for this kind of model. Most of them are “directional” in that they detect departures from mainly a given assumption of the model. Other tests are “global” in that they assess whether a model fits a data set on all its assumptions. We focus on the task of choosing the structural part $m(\cdot)$. It gets most attention because it contains easily interpretable information about the relationship between X and Y. To valid the form of the regression function, we consider three nonparametric tests based on a generalization of the Cramér-von Mises statistic. The first two are directional tests, while the third is a global test. To perform these goodness-of-fit tests based on a generalization of the Cramér-von Mises statistic, we have used Wild bootstrap methods and we also proposed a method to choose the bandwidth parameter used in nonparametric estimations. Then, we have developed cvmgof R package (being submitted), an easy-to-use tool for many users. The use of the package is illustrated using simulations to compare the three implemented tests [37].

In epidemiology, we are working with clinicians to study fetal development in the last two trimesters of pregnancy. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy) and from the EDEN cohort (INSERM). We propose to use non parametric methods of estimation to obtain reference curves of fetus and child growth. In addition, we want to develop a test, based on Z-scores, to detect any slope breaks in the fetal development curves (work in progress).

BONSAI Project-Team

6. New Results

6.1. Exploration of transcriptomes

In 2016 we produced a method called CG-Alcode able to compare transcripts repertoires of a given pair of orthologous genes. We applied our method to compare human and mouse transcriptomes. This year, in collaboration with C.Belleannée (DYLISS, Inria Rennes) we explored the comparison of multiple species. We inspected human, mouse and dog transcriptomes. We thus were able to predict a large number of putative transcripts in both human, mouse and dog based on known transcripts. Those results allow to investigate which functional sites are conserved and which genes have the same set of transcripts (known or putative).

6.2. Modeling of alternative transcripts with long reads

In the context of transcriptomic analyses based on third generation sequencing data (ONT), we started to explore the following problem : given a transcriptomic experiment, a gene of interest, select reads related to the given gene and find exon junctions. As we have done in the CG-Alcode project, we aim to model the gene as an alphabet of exonic blocks, each transcript being a word over this alphabet. This work takes place in the context of ANR ASTER for which we deal with mouse transcriptomic data in brain and liver. Built models will allow to query human genes to discover putative transcripts.

6.3. Read against read comparison for Nanopore data

In the team, we developed two years ago seeds with errors, which allow to find all common approximate patterns with a limited number of errors. The idea behind these seeds, called 01^*0 seeds, is to divide the sequence in blocks so that the distribution of errors is no longer random. This year, we have used these seeds in the context of long reads analysis. With this data, reads against reads comparison suffers from a high loss of sensitivity, because the single *read error-rate* is already high. Our application case is the detection of adapter sequences in ONT sequencing. We have shown that the use of these seeds instead of exact k-mers allowed a more accurate reconstruction of the sequences of the adapters. The method takes two steps: first the identification of k-mers potentially composing the adapter using a counting approach that takes into account errors in the read, and then the reconstruction of the complete sequence of the adapter with a greedy algorithm. Our results show that the seeds with errors allow to obtain accurate consensus sequences for more 80% of the samples, compared to 40% with the usual k-mer approach. This work was done within the ANR ASTER during the first year of the thesis of Quentin Bonenfant and was presented at the national workshop Seqbio 2018.

6.4. Annotation of the OC43 coronavirus genome

OC43 coronavirus is recognized as frequent cause of respiratory infection. We have conducted a bioinformatics study of 8 coronavirus genomes collected from patients at Lille hospital : gene annotation, phylogenetic analysis and amino acids substitutions. Several genotypes (B, E, F and G) were identified and two clusters of patients were defined from chronological data and phylogenetic trees based on the genomic sequences,. Analyses of amino acids substitutions of the S protein sequences identify substitutions specific of genotype F strains circulating among French people. This work is a collaboration with Anne Goffard (CHRU Lille and CIIL).

6.5. Small RNAs catalog in oilseed rape

Polyploidy – and notably allopolyploidy that involves interspecific hybridization – has played a major role in the evolution of plants, partly because this process is often associated with genomic structure and expression changes. Homeologous exchanges (HE) – i.e. between the constituent subgenomes– have been demonstrated to be frequent in allopolyploids and could be involved in the origin and maintenance of polyploids. While their influence on gene content has poorly been studied until recently, little is known about their impact on gene expression. Together with K. Alix (Inra Moulon), we have analyzed the impact of HEs that have been characterized in resynthesized oilseed rapes, on the repertoire of micro RNAs. Our main objective was to assess the relations that could exist between structural variation and modifications of gene expression through changes in miRNA regulation. The analysis was based on the small RNA-seq catalog obtained with the bioinformatic tool miRkwood, developed in BONSAI. We have built a microRNA database for the diploid subgenomes AA from *Brassica rapa* and CC for *Brassica oleracea* that correspond to the progenitors of the resynthesized *Brassica napus* allotetraploids (AACC). Integrating miRNA prediction and genomic location of HEs allowed us to infer relationships between microRNA restructuring and non-additivity of gene expression in polyploid hybrids.

6.6. Identifying systematic sequencing errors

Discovering over-represented approximate motifs in DNA sequences is an essential part of bioinformatics, which has been studied extensively. However, it remains a difficult challenge, especially with the huge quantity of data generated by high throughput sequencing technologies. We have developed an exact discriminative method for IUPAC motifs discovery in large sets of DNA sequences. The approach uses mutual information (MI) as an objective function to search for over-represented degenerate motifs in a lattice [7].

The algorithm was applied to the problem of *Sequence-Specific Errors*. Next Generation Sequencing, and further Single-Molecule Sequencing technologies are known to produce a highly variable error rate. A common method to overcome these sequencing errors is to increase the *coverage*. However, Sequence-Specific Errors are recurrent errors that depend on the upstream nucleotidic context, and can thus be confused with true genomic variations when the read coverage increases. Our algorithm was able to find motifs associated to sequencing errors and therefore to improve variant calling. This method has also tested on ChIP-seq datasets, and compared with five state-of-the art methods, where it was experimentally shown to perform as well as the best one, while be resistant to down-sampling.

This work was done during the thesis of Chadi Saad, and as a collaboration with Martin Figeac (Univ. Lille - Plateau de génomique fonctionnelle et structurale), Julie Leclerc and Marie-Pierre Buisine (CHRU de Lille - JPARC), and Hugues Richard (Sorbonne Université - Laboratory Computational and Quantitative Biology).

6.7. Indexing labelled sequences

We designed a compressed full-text index structure able to index a whole text with labels attached to every letter in the text [6]. This work will be applied to DNA sequences and more precisely V(D)J recombinations which are complex genomic rearrangements occurring in lymphocytes. The index will be used to index labelled V(D)J recombinations, which are labelled with their V, D and J gene. As the index we conceived is scalable, we will index V(D)J recombinations from thousands of samples and give access to this data through the Vidjil platform.

6.8. Tree representations

We found an intriguing duality between two well-known representations of trees [12]. This work concerns data structures and succinct tree representations. The Balanced Parenthesis representation of trees consists of encoding the structure of any tree using a series of opening and closing parentheses. The DFUDS representation is similar, but differs in how each node is encoded (also using parentheses). By relating both BP and DFUDS representations, we obtained improvements for a basic fundamental problem: the Minimum Length Interval Query problem. We also reported unnoted commonalities in recent solution to the Range Minimum Query problem.

6.9. Co-linear chaining on graphs

We reported the first algorithm that perform co-linear chaining between a sequence and a directed acyclic graph (DAG) [9]. This work concerns dynamic programming algorithms and sequencing alignment. The problem of co-linear chaining is a classical bioinformatics problem, which has immediate application to sequence alignment, as it is used as a filter remove spurious alignment seeds. Co-linear chaining is typically solved using a simple dynamic programming algorithm. Yet, representations of genomes using graphs instead of sequences have recently became an active research topic. As a results, the problem of aligning a sequence to a sequence graph merits consideration. This work provides the first step towards tackling practical sequence-to-graph alignment instances, by first considering the case when the graph is a DAG. We designed a $O(k|E|log|V|)$ algorithm to solve co-linear chaining on DAGs, which matches the optimal solution for the classical sequence variant, i.e. when the graph is a path.

6.10. Representations of de Bruijn graphs

We designed the first practical data structure for representing large de Bruijn graphs, which supports insertions and deletions of nodes [3]. This work concerns *de novo* assembly and several other k-mer-related bioinformatics problems. The representation of de Bruijn graphs is a transversal bioinformatics question that has enjoyed recent applications in genome, metagenome and transcriptome assembly and quantification. To this date, efficient data structures were essentially static. In this work we provided an implementation of a dynamic data structure that combines perfect hashing, Karp-Rabin hashing, and forests. Practical tests show that this structure is highly competitive with the state of the art.

6.11. Readability of overlap graphs

We report further progress on the study of a theoretical parameter of graph named *readability* [8]. This work concerns graph theory mainly. The readability parameter measures the minimal length of strings that would be needed in order to label a graph such that it is an overlap graph over a set of strings of that length. So far, recent works on readability have not elucidated many aspects related to this parameter: the complexity of computing it is open, and it is not even known whether the corresponding decision problem is in NP. The only upper bound known for this parameter is exponential. This work focuses on certain graph families: bipartite chain graphs, grids, induced subgraphs of grids, and provides a characterization of bipartite graphs of readability 2.

6.12. Nonribosomal peptides

Norine is a comprehensive public database for non-ribosomal peptides developed by the team for more than 10 years. The Norine database quality has been enhanced through a semi-automatic curation process of data. Particularly, more than 500 SMILES annotations have been added or updated. This allowed us to check and correct the monomeric graphs, i.e. a 2D representation of the monomeric composition of the NRPs, thanks to dedicated tools like Smiles2Monomers. This update was done in collaboration with members of the Proteome Informatics Group from SIB (Swiss Institute of Bioinformatics). New annotations on monoisotopic mass and molecular formulas have also been added. The Norine interface was improved and new features are available, such as the possibility to access the complete change history of each entry. To encourage new submissions of NRPs, authors of new NRPs are now visible as contributors on Norine home page. Finally, we published this year, in the field of biocontrol (a contraction of “biological control”), a paper on bioinformatic tools for the discovery of new lipopeptides [5], essentially based on the Norine platform.

CAPSID Project-Team

7. New Results

7.1. Drug Targeting and Adverse Drug Side Effects

Identifying new molecular targets using comparative genomics and knowledge of disease mechanisms is a rational first step in the search for new preventative or therapeutic drug treatments [63]. We are mostly concerned with three global health problems, namely fungal and bacterial infections and hypertension. Through on-going collaborations with several Brazilian laboratories (at University of Mato Grosso State, University of Maringá, Embrapa, and University of Brasilia), we previously identified several novel small-molecule drug leads against *Trypanosoma cruzi*, a parasite responsible for Chagas disease [91]. With the University of Maringá, we subsequently found several active molecules against the flavoenzyme TRR1 in *Candida albicans*, and two manuscripts are in preparation. We also proposed several small-molecule inhibitors against *Fusarium graminearum*, a fungal threat to global wheat production [63], [43]. Two further manuscripts on this topic are currently in preparation. Concerning hypertension, we continued our collaboration with Prof. Catherine Llorens-Cortes at Collège de France to study the interaction between the apelin receptor (a transmembrane protein important for blood pressure regulation) and the aminopetidase A enzyme [47].

It is well known that many therapeutic drug molecules can have adverse side effects. However, when patients take several combinations of drugs it can be difficult to determine which drug is responsible for which side effect. In collaboration with Adrien Coulet (Orpailleur team co-supervisor of Gabin Personeni) and Prof. Michel Dumontier (Biomedical Informatics Research Laboratory, Stanford), we developed an approach which combines multiple ontologies such as the Anatomical Therapeutical Classification of Drugs, the ICD-9 classification of diseases, and the SNOMED-CT medical vocabulary together with the use of Pattern Structures (an extension of Formal Concept Analysis) in order to extract association rules to analyse the co-occurrence of adverse drug effects in patient records [74], [73]. A paper describing this work has been published in the Journal of Biomedical Semantics [70].

7.2. Docking Symmetrical Protein Structures

Many proteins form symmetrical complexes in which each structure contains two or more identical copies of the same sub-unit. We recently developed a novel polar Fourier docking algorithm called “Sam” for automatically assembling symmetrical protein complexes. A journal article describing the Sam algorithm has been published [8]. An article describing the results obtained when using Sam to dock several symmetrical protein complexes from the “CASP/CAPRI” docking experiment has also been published [53]. This study showed that many of the models of protein structures built by members of the “CASP” fold prediction community are “dockable” in the sense that Sam is able to find acceptable docking solutions from amongst the CASP models.

More recently, we are working to extend the polar Fourier correlation algorithm to use very high angular resolution spherical Bessel basis functions. As part of this work, we have developed a very fast recursive algorithm for calculating high order Clebsch-Gordan coupling coefficients [30]. A manuscript describing this work has been submitted to a quantum mechanics journal.

7.3. Multiple Flexible Protein Structure Alignments

Comparing two or more proteins by optimally aligning and superposing their backbone structures provides a way to detect evolutionary relationships between proteins that cannot be detected by comparing only their primary amino-acid sequences. The latest version of our “Kpax” protein structure alignment algorithm can flexibly align pairs of structures that cannot be completely superposed by a single rigid-body transformation, and can calculate multiple alignments of several similar structures flexibly [9]. In collaboration with Alain

Hein of the INRA lab "Agronomie et Environnement", we used Kpax to help study the structures of various "Cyp450" enzymes in plants [81]. In collaboration with Emmanuel Levy of the Weizmann Institute, we used Kpax to superpose and compare all of the symmetrical protein complexes in the Protein Databank in order to verify or remediate their quaternary structure annotations. A manuscript describing this work has been published in Nature Methods [15].

7.4. Large-Scale Annotation of Protein Domains and Sequences

Many protein chains in the Protein Data Bank (PDB) are cross-referenced with Pfam domains and Gene Ontology (GO) terms. However, these annotations do not explicitly indicate any relation between EC numbers and Pfam domains, and many others lack GO annotations. In order to address this limitation, as part of the PhD thesis project of Seyed Alborzi, we developed the CODAC approach for mining multiple protein data sources (i.e. SwissProt, TremBL, and SIFTS) in order to associate GO molecular function terms with Pfam domains, for example. We named the software implementation "GO-DomainMiner". This work was first presented at IWBBIO 2017 [36]. A full paper has recently been accepted for a special issue of *BMC Bioinformatics* [13].

In collaboration with Maria Martin's team at the European Bioinformatics Institute (EBI), we combined the CODAC approach with a novel combinatorial association rule based approach called "CARDM" for annotating protein sequences. When applied to the large UniProt/TrEMBL sequence database of 63 million protein entries, CARDM predicted over 24 million Enzyme Commission (EC) numbers and 188 million GO terms for those entries. A journal paper in collaboration with the EBI on comparing the quality of these predicted annotations with other state of the art annotation methods is in preparation, and a poster was presented at ISMB-ECCB-2017 [35]. As part of the PhD thesis of Bishnu Sarker, we also developed GrAPFI, a graph-based protein function annotation approach. GrAPFI applies a label propagation algorithm to a complex network representation of protein sequence data. A full paper on this work has recently been accepted by the International Conference on Complex Networks and their Applications [24].

7.5. Distributed Protein Graph Processing

The huge number of protein sequences in protein databases such as UniProtKB calls for rapid procedures to annotate them automatically. We are using existing protein annotations to predict the annotations of new or non-reviewed proteins. In this context, we developed the "DistNBLP" method for annotating protein sequences using a graph representation and a distributed label propagation algorithm. DistNBLP uses the BLADYG framework [38] to process protein graphs on multiple compute nodes by applying a neighbourhood-based label propagation algorithm in a distributed way. We applied DistNBLP in the recent "CAFA 3" (critical Assessment of Protein Function Annotation) community experiment to annotate new protein sequences automatically. This work was presented as a poster at ISMB/ECCB-2017 [34]. We are also interested in feature selection for subgraph patterns. In collaboration with the LIMOS laboratory at Université Clermont Auvergne we also developed a scalable approach using MapReduce for identifying sub-graphs having similar labels in very large graphs [51].

7.6. Flexible Docking of Protein-GAG Complexes

Modeling how flexible polymers bind to proteins presents enormous computational challenges due to the large conformational search space that arises from the many internal rotational degrees of freedom in polymer structures. In collaboration with Sergey Samsonov (Gdansk University, Poland), we extended our fragment-based flexible docking approach [83], [42] to model how flexible Glycosaminoglycans (GAGs) might bind to the surface of a known protein structure. A paper has been submitted to the Journal of Computational Chemistry.

In collaboration with Sjoerd de Vries (Univ Paris Diderot), we have created a new protein-glycan interaction force-field and integrated it in the ATTRACT docking engine [83]. We also participated in a comparative study of the main current protein-GAG docking methods.

7.7. Stochastic Decision Trees for Similarity Computation

We have designed a method to compute similarities on unlabeled data using stochastic decision trees [20]. The main idea of Unsupervised Extremely Randomized Trees (UET) is to randomly and iteratively split the data until a stopping criterion is met. Pairwise similarity values are computed based on the co-occurrence of samples in the leaves of each generated tree. We evaluate our method on synthetic and real-world datasets by comparing the mean similarities between samples with the same label and the mean similarities between samples with distinct labels. Empirical studies show that the method effectively gives distinct similarity values between samples belonging to distinct clusters, and gives indiscernible values when there is no cluster structure. We also assessed some interesting properties such as invariance under monotone transformations of variables and robustness to correlated variables and noise. Our experiments show that the algorithm outperforms existing methods in some cases, and can reduce the amount of preprocessing needed with many real-world datasets. We plan to study the application of this "global" pairwise similarity computation to quantify protein structural similarities. Two interesting problems will concern the representation of the protein structure and how to tackle extra constraints such as invariance under rotational and translational transformations.

DYLISS Project-Team

7. New Results

7.1. Scalable methods to query data heterogenity

Participants: Guillaume Alviset, Olivier Dameron, Xavier Garnier, Vijay Ingalalli, Marine Louarn, Yann Rivault, Anne Siegel, Denis Tagu.

Ontology design and integration [*O. Dameron, Y. Rivault*] We have contribued to several technics improving data integration in onthologies

- The ATOL ontology [link to ontology] supports the annotation of phenotype traits in livestock. It was extended with health-related traits. For each organism, livestock diseases are organized according to their type (infectious, genetic, metabolic,...), their transmission and their symptoms. [32]
- queryMed is an R package [url] that provides both high-level and low-level functions for facilitating the integration of reference ontologies and datasets represented in RDF as Linked Data. It currently focuses on drugs indications, interactions and contra-indications by integrating the Drug Indication Database (DID) and the Drug Interaction Knowledge Base (DIKB). Typical applications concern public health and pharmaco-epidemiology. [27], [26]

Using AskOmics to integrate heterogeneous data [*O. Dameron, A. Siegel*]

- We contributed to the conversion of an Alzheimer's disease map into a heavyweight ontology, the Alzheimer's Disease Map Ontology (ADMO, [url]), an ontological upper model based on systems biology terms. It provides the ontological formalization for the existing disease map AlzPathway that gives a detailed and broad account of Alzheimer's Disease pathophysiology [25], [20].
- We also contributed to decipher the role of small non-coding RNAs in the regulation of animal reproduction, especially the role of miR-202 in female fecundity by regulating medaka oncogenesis [16].

Graph compression and analysis [*L. Bourneuf*]. We introduced a general approach combining procedural and logical languages to specify graph objects. This is a generalization of previous work [37], using the reconstruction of Formal Concept Analysis framework example to target the AI community [23].

7.2. Metabolism: from enzyme sequences to systems ecology

Participants: Meziane Aite, Arnaud Belcour, Marie Chevallier, Mael Conan, François Coste, Olivier Dameron, Clémence Frioux, Jeanne Got, Jacques Nicolas, Anne Siegel, Hugo Talibart.

Efficient identification of substitutable context-free grammars by reduction [*F. Coste, J. Nicolas*] To study more formally the approach by reduction initiated by ReGLiS [40], we introduced a formal characterization of the grammars in reduced normal form (RNF) which can be learned by this approach. Modifying the core of ReGLiS to ensure polynomial running time, we show that local substitutable languages represented by RNF context-free grammars are identifiable in polynomial time and thick data (IPTtD) from positive examples by reduction [19].

Learning grammars capturing 3D structural features of proteins [*F. Coste, H. Talibart*] With the team of Witold Dyrka in Polland, we investigated the problem of learning context-free grammars modeling well protein sequences with respect to their 3D structures.

- A preliminary step is to be able to quantify the relevance of a grammar with respect to a structure. In [21], we introduced and assessed quantitative measures for comparing the topology of the parse tree of a protein sequence analyzed by a context-free grammar with the topology of the protein structure.
- In [24], we established a new framework for learning probabilistic context-free grammars for protein sequences using predicted or experimentally assessed amino acid 3D contacts. We relied on maximum-likelihood and contrastive estimators of parameters in this setting and an implementation for simple yet practical grammars. Tested on samples of protein motifs, grammars developed within the framework showed improved precision in recognition and higher fidelity to protein structures.

Metabolic pathway inference from non genomic data [*A. Belcour, M. Aite, J. Nicolas, A. Siegel, N. Théret, V. Dellannée, M. Conan*] We designed methods for the identification of metabolic pathways for which enzyme information is not precise enough.

- Heterocyclic Aromatic Amines (HAAs) are environmental and food contaminants classified as probable carcinogens. Our approach based on a refinement of molecular predictions with enzyme activity scores allowed us to accurately predict HAAs biotransformation and their potential DNA reactive compounds [13].
- We designed a prototype (*Pathmodel*) implementing inference methods to reconstruct biochemical reactions and metabolite structures to cope with metabolic pathway drift mechanisms. Using known metabolic pathways and metabolomics data, the tool infers alternative pathways compatible with the species known metabolites [29].

Large-scale eukaryotic metabolic network reconstruction [*A. Siegel, M. Chevallier, C. Frioux, M. Aite, J. Cambefort*] Metabolic network reconstruction has attained high standards but is still challenging for complex organisms such as eukaryotes.

- In this direction, we developed AuReMe for a flexible and reproducible reconstruction of these models. Together with a convenient mean for exploration through a local wiki, AuReMe is well-suited for the study of non-model organisms [12].
- In addition, a new gap-filling method satisfying the two main semantics of activation in metabolism is available. It enables to refine the models by pinpointing reactions such that metabolic objectives are met [15].

Systems ecology: design of microbial consortia [*C. Frioux, A. Siegel*]. Finding key elements among hundreds or thousands in microbiota to explain metabolic behaviours or prepare biological experimentations is a highly combinatorial problem. We introduced a two-step approach, MiSCoTo to screen the metabolic capabilities of microbiotas and exhaustively select members of interest by solving optimization problems with logic programming. We applied these methods to data from the Human Microbiome Project and a system composed of the Human metabolic network and 773 models for gut bacteria [14], [11].

7.3. Regulation and signaling: detecting complex and discriminant signatures of phenotypes

Participants: Catherine Belleannée, Samuel Blanquart, Célia Biane-Fourati, Nicolas Guillaudeux, Marine Louarn, Maxime Folschette, François Moreews, Anne Siegel, Nathalie Théret, Pierre Vignet, Méline Wery.

Comparative-genomics based prediction of non-model transcriptomes [*C. Belleannée, S. Blanquart, N. Guillaudeux*] In order to annotate the transcriptome of a non-model species, *Canis lupus familiaris*, we developed a method to predict whether or not a transcript known in a given species/gene could be expressed in an other species/gene. Exploiting knowledge in human, mouse and dog , we predicted a total of 7201 unknown yet transcripts and interpreted the evolutionary dynamics of gene's isoform sets. [30]

Signaling network identification [*M. Folschette, A. Siegel*] [22], [17]

- We introduced a new method to learn an interaction graph from the knowledge of its state space, without assumption on the semantics that was used to produce it. Proofs and characterizations are given for the synchronous, asynchronous and generalized semantics.
- We also used the caspo time-series software to integrate large-scale time series phosphoproteomic data (HPN-DREAM Breast Cancer challenge) into protein signaling networks and infer a family of Boolean Networks. The method highlights commonalities and discrepancies between the four cell lines.

Static analysis of ruled-based models [*P. Vignet, N. Théret*] We used a model of TGF-β to illustrate the main features of Kasa, a static analyzer for Kappa models. Kappa is a rule based language that describes systems of mechanistic interactions between proteins by the means of site-graph rewriting rules. The cornerstone of KaSa is a fix-point engine which detects some patterns that may never occur whatever the evolution of the system may be [18].

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2018.

We tried to organise these along the four axes as presented above. Clearly, in some cases, a result obtained overlaps more than one axis. In such case, we chose the one that could be seen as the main one concerned by such results.

We did not indicate here the results on more theoretical aspects of computer science if it did not seem for now that they could be relevant in contexts related to computational biology. Actually, those on string [32], [33], [36], [11] and graph algorithms in general [2], [35], [38], [37], [39], [41], [54], [42], [44], [43], [40], [47], [5], [45], [48], [49], [53], [23], [24], or on more general algorithmic problems notably related to data structures are already relevant for life sciences (biology or ecology) or in the future could become more specifically so. We do in particular believe that dynamic graph approaches could be of great interest in the future for some of the enumeration problems we constantly meet in biology.

A few other results of 2018 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised [52], or the work represented a survey *e.g.* [21]). Likewise, also for space reasons, we do not detail the results presented in some biological papers of the team when these did not require a mathematical or algorithmical input or are surveys [1], [4], [7], [8], [12], [19], [20], [22], [25], [31].

On the other hand, we do mention a couple of works that were submitted towards the end of 2018.

6.2. Axis 1: Genomics

Genome hybrid assembly
Long read sequencing technologies are considered to be the solution for handling genome repeats, allowing near reference-level reconstructions of large genomes. However, long read de novo assembly pipelines are computationally intense and require a considerable amount of coverage, thereby hindering their broad application to the assembly of large genomes. Alternatively, hybrid assembly methods that combine short and long read sequencing technologies can reduce the time and cost required to produce de novo assemblies of large genomes. In [10], we proposed a new method, called FAST-SG, that uses a new ultrafast alignment-free algorithm specifically designed for constructing a scaffolding graph using lightweight data structures. FAST-SG can construct the graph from either short or long reads. This allows the reuse of efficient algorithms designed for short read data and permits the definition of novel modular hybrid assembly pipelines. Using comprehensive standard datasets and benchmarks, we showed how FAST-SG outperforms the state-of-the-art short read aligners when building the scaffolding graph and can be used to extract linking information from either raw or error-corrected long reads. We also showed how a hybrid assembly approach using FAST-SG with shallow long-read coverage (5X) and moderate computational resources can produce long-range and accurate reconstructions of the genomes of *Arabidopsis thaliana* (Ler-0) and human (NA12878). We are currently working on the assembly process itself, using the scaffolding graphs obtained with FAST-SG. The results obtained so far are extremely promising and a paper is currently in preparation. This is part of the work done by Alex di Genova, postdoc in ERABLE.

Variant annotation

Genome-wide analyses estimate that more than 90% of multi exonic human genes produce at least two transcripts through a genomic variant called alternative splicing (AS). Various bioinformatics methods are available to analyse AS from RNAseq data. Most methods start by mapping the reads to an annotated reference genome, but some start by a *de novo* assembly of the reads. In [3], we presented a systematic comparison of a mapping-first approach (FARLINE) and an assembly-first approach (scKisSplice). We applied these methods to two independent RNAseq datasets and found that the predictions of the two pipelines overlapped (70% of exon skipping events were common), but with noticeable differences. The assembly-first approach allowed to find more novel variants, including novel unannotated exons and splice sites. It also predicted AS in recently duplicated genes. The mapping-first approach allowed to find more lowly expressed splicing variants, and splice variants overlapping repeats. This work demonstrated that annotating AS with a single approach leads to missing out a large number of candidates, many of which are differentially regulated across conditions and can be validated experimentally. We therefore advocate for the combined use of both mapping-first and assembly-first approaches for the annotation and differential analysis of AS from RNAseq datasets. This was part of the work of Clara Benoît-Pilven, postdoc at Inserm and in ERABLE, to which also participated other current or ex-members of ERABLE, namely Camille Marchet (during her stay as ADT engineer with ERABLE), Emilie Chautard (when she was postdoc Inserm and in ERABLE), Gustavo Sacomoto (when he was PhD and then for one year postdoc in ERABLE), and Leandro Lima (current PhD student of ERABLE).

Another type of variant, namely SNPs was also considered in [51]. In this paper, mutations are detected by eBWT (extended Burrows-Wheeler Transform). Indeed, we notices that eBWT of a collection of DNA fragments tend to cluster together the copies of nucleotides sequenced from a genome. We showed that it is thus possible to accurately predict how many copies of any nucleotide are expected inside each such cluster, and that a precise LCP array based procedure can locate these clusters in the eBWT. These theoretical insights were validated in practice with SNPs being clustered in the eBWT of a reads collection. We developed a tool for finding SNPs with a simple scan of the eBWT and LCP arrays. Preliminary results show that our method requires much less coverage than the state-of-the-art tools while drastically improving precision and sensitivity.

Both types of variants correspond to special types of st-paths in graphs, a topic that was also explored from a more purely theoretical point of view in two papers, one already accepted [46] and on that is about to be submitted and extends the results obtained in 2017 on bubble (as st-paths are also called in bioinformatics) generators in directed graphs.

Full-length *de novo* viral quasispecies assembly through variation graph construction
Viruses populate their hosts as a viral quasispecies: a collection of genetically related mutant strains. Viral quasispecies assembly refers to reconstructing the strain-specific haplotypes from read data, and predicting their relative abundances within the mix of strains, an important step for various treatment-related reasons. Reference-genome-independent ("de novo") approaches have yielded benefits over reference-guided approaches, because reference-induced biases can become overwhelming when dealing with divergent strains. While being very accurate, extant *de novo* methods only yield rather short contigs. It remains to reconstruct full-length haplotypes together with their abundances from such contigs. In [34], we first constructed a variation graph, a recently popular, suitable structure for arranging and integrating several related genomes, from the short input contigs, without making use of a reference genome. To obtain paths through the variation graph that reflect the original haplotypes, we solved a minimisation problem that yields a selection of maximal-length paths that is optimal in terms of being compatible with the read coverages computed for the nodes of the variation graph. We output the resulting selection of maximal length paths as the haplotypes, together with their abundances. Benchmarking experiments on challenging simulated data sets showed significant improvements in assembly contiguity compared to the input contigs, while preserving low error rates. As a consequence, our method outperforms all state-of-the-art viral quasispecies assemblers that aim at the construction of full-length haplotypes, in terms of various relevant assembly measures. The tool, called VIRUS-VG, is availabe at https://bitbucket.org/jbaaijens/virus-vg.

A member of ERABLE was also involved in the Second Annual Meeting of the European Virus Bioinformatics Center (EVBC), held in Utrecht, Netherlands, and whose focus was on computational approaches in virology, with topics including (but not limited to) virus discovery, diagnostics, (meta-)genomics, modeling, epidemiology, molecular structure, evolution, and viral ecology. Approximately 120 researchers from around the world attended the meeting this year. An overview of new developments and novel research findings that emerged during the meeting was published in the journal *Viruses* [16].

Bacterial genome-wide association studies (GWAS)
Genome-wide association study (GWAS) methods applied to bacterial genomes have shown promising results for genetic marker discovery or detailed assessment of marker effect. Recently, alignment-free methods based on k-mer composition have proven their ability to explore the accessory genome. However, they lead to redundant descriptions and results which are sometimes hard to interpret. In [17], we introduced DBGWAS, an extended k-mer-based GWAS method producing interpretable genetic variants associated with distinct phenotypes. Relying on compacted de Bruijn graphs (cDBG), our method gathers cDBG nodes, identified by the association model, into subgraphs defined from their neighbourhood in the initial cDBG. DBGWAS is alignment-free and only requires a set of contigs and phenotypes. In particular, it does not require prior annotation or reference genomes. It produces subgraphs representing phenotype-associated genetic variants such as local polymorphisms and mobile genetic elements (MGE). It offers a graphical framework which helps interpret GWAS results. Importantly, it is also computationally efficient (the experiments took one hour and a half on average). We validated our method using antibiotic resistance phenotypes for three bacterial species. DBGWAS recovered known resistance determinants such as mutations in core genes in *Mycobacterium tuberculosis*, and genes acquired by horizontal transfer in *Staphylococcus aureus* and *Pseudomonas aeruginosa* along with their MGE context. It also enabled us to formulate new hypotheses involving genetic variants not yet described in the antibiotic resistance literature. This is part of the work of Magali Jaillard, PhD student of Laurent Jacob who is an external collaborator of ERABLE, and of Leandro I. S. de Lima, PhD student co-supervised by three members of ERABLE.

6.3. Axis 2: Metabolism and post-transcriptional regulation

Multi-objective metabolic mixed integer optimisation: with an application to yeast strain engineering
In a paper submitted and already available in bioRxiv (https://www.biorxiv.org/content/early/2018/11/22/476689), we explored the concept of multi-objective optimisation in the field of metabolic engineering when both continuous and integer decision variables are involved in the model. In particular, we proposed a multi-objective model which may be used to suggest reaction deletions that maximise and/or minimise several functions simultaneously. The applications may include, among others, the concurrent maximisation of a bioproduct and of biomass, or maximisation of a bioproduct while minimising the formation of a given by-product, two common requirements in microbial metabolic engineering. Production of ethanol by the widely used cell factory *Saccharomyces cerevisiae* was adopted as a case study to demonstrate the usefulness of the proposed approach in identifying genetic manipulations that improve productivity and yield of this economically highly relevant bioproduct. We did an *in vivo* validation and we could show that some of the predicted deletions exhibit increased ethanol levels in comparison with the wild-type strain. The multi-objective programming framework we developed, called MOMO, is open-source and uses POLYSCIP as underlying multi-objective solver. This is part of the work of Ricardo de Andrade, postdoc at Unversity of São Paulo with Roberto Marcondes, and in ERABLE. It is joint work with Susana Vinga, external collaborator of ERABLE and partner of the Inria Associated Team Compasso.

Metabolic shifts

With the increasing availability of so-called 'omics data – transcriptomics, proteomics, and metabolics – there has been growing interest in various ways of integrating them with the metabolic network. When the network is represented by a graph, 'omics data can guide the extraction of subnetworks of interest to find metabolic pathways or sets of related genes. Within the framework of constraint-based modelling, 'omics data can be used to improve the prediction of metabolic behaviour and to build context-specific metabolic models. One interesting application of metabolic reconstructions in conjunction with 'omics data is to use the two to understand metabolic shifts. When an organism encounters a change in environmental conditions, often a re-organisation of metabolism follows. Comparative measurements of gene expression and metabolite concentrations can be used to gain insight into these changes but this data is "structureless", meaning it lacks the information about how the metabolic components relate to each other. A metabolic network on the other hand contains this information, and can thus greatly benefit such an analysis. We developed a new method, called MOOMIN, that combines the results of a differential expression analysis comparing the gene expression levels in two different conditions with a metabolic network to produce a hypothesis of a metabolic shift. The idea is to use the network structure to define feasible global changes in metabolism. These changes are then scored based on the gene expression data with the goal of finding the change that best agrees with the observations. Finding the best-scoring change is formulated into an optimisation problem that can be solved using Mixed-Integer Linear Programming. This is part of the work of Henri Taneli Pusa, co-supervised by 3 members of ERABLE, whose manuscript was submitted to the reviewers and who should be defending his PhD in early February 2019. The paper on MOOMIN will be submitted soon, and the software then made available. Participated also in this work Mariana G. Ferrarini, postdoc at Insa and in ERABLE, and Ricardo Andrade, postdoc at University of São Paulo with Roberto Marcondes and in ERABLE.

Metabolic games
The PhD of Taneli also investigated game theory in the context of metabolism. Game theory is a branch of applied mathematics that deals with interacting rational agents with conflicting goals. When rationality is replaced with natural selection, *evolutionary* game theory can be used to explain the "decisions" taken by even microscopic organisms. The PhD manuscript presents the idea of a *metabolic game*, a game theoretical model for the prediction of metabolic behaviour. In contrast to Flux Balance Analysis, where the metabolic state is predicted using simple optimisation, a metabolic game takes into account the fact that optimality is influenced by the surrounding members of a microbial community. By changing the availability of nutrients, or secreting beneficial or harmful molecules, microbes essentially create their own environment and make optimal behaviour context-specific. A paper is submitted that reviews the literature that has applied game theory to the study of microbes, with a focus on metabolism and especially games derived using metabolic networks and constraint-based modelling. In the PhD manuscript, Taneli further explains the idea behind a metabolic game and discusses different aspects of defining such games: the choice of players, actions, and payoffs.

6.4. Axis 3: (Co)Evolution

Exploring the robustness of the parsimonious reconciliation method in host-symbiont cophylogeny
Following our previous work on reconciliation methods for cophylogeny, in [29], we explored the robustness of the parsimonious host-symbiont tree reconciliation method under editing or small perturbations of the input. The editing involved making different choices of unique symbiont mapping to a host in the case where multiple associations exist. This is made necessary by the fact that the tree reconciliation model is currently unable to handle such associations. The analysis performed could however also address the problem of errors. The perturbations were re-rootings of the symbiont tree to deal with a possibly wrong placement of the root specially in the case of fast-evolving species. In order to do this robustness analysis, we introduced a simulation scheme specifically designed for the host-symbiont cophylogeny context, as well as a measure to compare sets of tree reconciliations, both of which are of interest by themselves. This work was also part of the PhD of a previous student of ERABLE, Laura Urbini.

Geometric medians in reconciliation spaces

Recently, there has been much interest in studying spaces of tree reconciliations (as used in cophylogenetic studies), which arise by defining some metric d on the set $\mathcal{R}(P, H, \phi)$ of all possible reconciliations between two trees P and H where ϕ represents the map between the leaf-sets of P and H (corresponding to present-day associations). In [14], we studied the following question: how do we compute a *geometric median* for a given subset Ψ of $\mathcal{R}(P, H, \phi)$ relative to d, *i.e.* an element $\psi_{med} \in \mathcal{R}(P, H, \phi)$ such that

$$\sum_{\psi' \in \Psi} d(\psi_{med}, \psi') \leq \sum_{\psi' \in \Psi} d(\psi, \psi')$$

holds for all $\psi \in \mathcal{R}(P, H, \phi)$? For a model where so-called host-switches or transfers are not allowed, and for a commonly used metric d called the *edit-distance*, we showed that although the cardinality of $\mathcal{R}(P, H, \phi)$ can be super-exponential, it is still possible to compute a geometric median for a set Ψ in $\mathcal{R}(P, H, \phi)$ in polynomial time. We expect that this result could be useful for computing a summary or consensus for a set of reconciliations (*e.g.* for a set of suboptimal reconciliations). The collaboration with Katharina Huber and Vincent Moulton from the School of Computing Sciences at the University of New Anglia was made possible by a Royal Society Grant obtained by the two partners (UNA and ERABLE).

Exploring and Visualising Spaces of Tree Reconciliations
A common approach to tree reconciliation involves specifying a model that assigns costs to certain events, such as cospeciation, and then tries to find a mapping between two specified phylogenetic trees which minimises the total cost of the implied events. For such models, it has been shown, including by the ERABLE members in previous papers, that there may be a huge number of optimal solutions, or at least solutions that are close to optimal. It is therefore of interest to be able to systematically compare and visualise whole collections of reconciliations between a specified pair of trees. In [13] , we considered various metrics on the set of all possible reconciliations between a pair of trees, some that have been defined before but also new metrics that we proposed. We showed that the diameter for the resulting spaces of reconciliations can in some cases be determined theoretically, information that we used to normalise and compare properties of the metrics. We also implemented the metrics and compared their behaviour on several host parasite datasets, including the shapes of their distributions. In addition, we showed that in combination with multidimensional scaling, the metrics can be useful for visualising large collections of reconciliations, much in the same way as phylogenetic tree metrics can be used to explore collections of phylogenetic trees. Implementations of the metrics can be downloaded from https://team.inria.fr/erable/en/team-members/blerina-sinaimeri/reconciliation-distances/. This work was also funded by a Royal Society Grant obtained by the two partners (at University of New Anglia and ERABLE).

Variants of phylogenetic network problems
Although not falling within the general topic of coevolution, phylogenetic networks are of great interest as another way of representing the evolution of a set of species. In the context of such representations, unrooted and root-uncertain variants of several well-known phylogenetic network problems were explored. The hybridisation number problem requires to embed a set of binary rooted phylogenetic trees into a binary rooted phylogenetic network such that the number of nodes with indegree two is minimised. However, from a biological point of view accurately inferring the root location in a phylogenetic tree is notoriously difficult and poor root placement can artificially inflate the hybridisation number. To this end, we studied in [30] a number of relaxed variants of this problem. We started by showing that the fundamental problem of determining whether an unrooted phylogenetic network displays (*i.e.* embeds) an unrooted phylogenetic tree, is NP-hard. On the positive side, we show that this problem is FPT in the reticulation number. In the rooted case, the corresponding FPT result is trivial, but here we required more subtle argumentation. Next we showed that the hybridisation number problem for unrooted networks (when given two unrooted trees) is equivalent to the problem of computing the tree bisection and reconnect distance of the two unrooted trees. In the third part of the paper, we considered the "root uncertain" variant of hybridisation number. Here we were free to choose the root location in each of a set of unrooted input trees such that the hybridisation number of the resulting rooted trees is minimised. On the negative side, we showed that this problem is APX-hard. On the positive side, we showed that the problem is FPT in the hybridisation number, via kernelisation, for any number of input trees.

6.5. Axis 4: Human, animal and plant health

Hydrogen peroxide production and myo-inositol metabolism as important traits for virulence of Mycoplasma hyopneumoniae

Mycoplasma hyopneumoniae is the causative agent of enzootic pneumonia. In a previous work, we had reconstructed the metabolic models of this species along with two other mycoplasmas from the respiratory tract of swine: *Mycoplasma hyorhinis*, considered less pathogenic but which nonetheless causes disease and *Mycoplasma flocculare*, a commensal bacterium. We had identified metabolic differences that partially explained their different levels of pathogenicity. One important trait was the production of hydrogen peroxide from the glycerol metabolism only in the pathogenic species. Another important feature was a pathway for the metabolism of myo-inositol in *M. hyopneumoniae*. In the paper accepted this year [9], we tested these traits to understand their relation to the different levels of pathogenicity, comparing not only the species but also pathogenic and attenuated strains of *M. hyopneumoniae*. Regarding the myo-inositol metabolism, we showed that only *M. hyopneumoniae* assimilated this carbohydrate and remained viable when myo-inositol was the primary energy source. Strikingly, only the two pathogenic strains of *M. hyopneumoniae* produced hydrogen peroxide in complex medium. We also showed that this production was dependent on the presence of glycerol. Although further functional tests are needed, this work enabled to identify two interesting metabolic traits of *M. hyopneumoniae* that might be directly related to its enhanced virulence. This is part of the work of Mariana G. Ferrarini, currently postdoc at Insa and in ERABLE, and of Scheila G. Mucha whose PhD (defended in Sept. 2018) was co-supervised by Arnaldo Zaha and by a member of ERABLE.

Cancer

A member of ERABLE continues deeply involved with the Centre Léon Bérard in Lyon, and in that context, a number of works are running, all related to cancer genomics. In the first [28], an integrated genomic study was performed of 25 tumour tissues from radical prostatectomy of aggressive (defined by the International Society of Urological Pathology) prostate cancer patients (10 African Caribbean and 15 French Caucasian) using single nucleotide polymorphism arrays, whole-genome sequencing, and RNA sequencing. The results showed that African Caribbean tumours are characterised by a more frequent deletion at 1q41-43 encompassing the DNA repair gene PARP1, and a higher proportion of intra-chromosomal rearrangements including duplications associated with CDK12 truncating mutations. Transcriptome analyses showed an over-expression of genes related to androgen receptor activity in African Caribbean tumours, and of PVT1, a long non-coding RNA located at 8q24 that confirms the strong involvement of this region in prostate tumours from men of African ancestry. In a second study [15], gene-expression profiling data was used to build and validate a predictive model of outcome for patients with follicular lymphoma. A robust 23-gene expression-based predictor of progression-free survival that is applicable to routinely available formalin-fixed, paraffin-embedded tumour biopsies from such patients was thus developed and validated. Applying this score could allow individualised therapy for patients according to their risk category. In a third study, an integrated analysis highlighted APC11 protein expression as a likely new independent predictive marker for colorectal cancer [6].

In a parallel work by another member of ERABLE [27], it was proposed that cancer Is not (only) a senescence problem. Age is indeed one of the strongest predictors of cancer and risk of death from cancer. Cancer is therefore generally viewed as a senescence-related malady. However, cancer also exists at subclinical levels in humans and other animals, but its earlier effects on the body are poorly known by comparison. What was argued in [27] is that cancer is a significant but ignored burden on the body and is likely to be a strong selective force from early during the lifetime of an organism. It was thus proposed that time has come to adopt this novel view of malignant pathologies to improve our understanding of the ways in which oncogenic phenomena influence the ecology and evolution of animals long before their negative impacts become evident and fatal.

Xylella fastidiosa epidemiological model

Xylella fastidiosa is a notorious plant pathogenic bacterium that represents a threat to crops worldwide. Its subspecies, *Xylella fastidiosa* subsp. *fastidiosa* is the causal agent of Pierce's disease of grapevines. Pierce's disease has presented a serious challenge for the grapevine industry in the United States and turned into an epidemic in Southern California due to the invasion of the insect vector *Homalodisca vitripennis*. In an attempt to minimize the effects of *Xylella fastidiosa* subsp. *fastidiosa* in vineyards, various studies have been developing and testing strategies to prevent the occurrence of Pierce's disease, *i.e.*, prophylactic strategies. Research has also been undertaken to investigate therapeutic strategies to cure vines infected by *Xylella fastidiosa* subsp. *fastidiosa*. In [18], we explicitly review all the strategies published to date and specifies their current status. Furthermore, an epidemiological model of *Xylella fastidiosa* subsp. *fastidiosa* is proposed and key parameters for the spread of Pierce's disease deciphered in a sensitivity analysis of all model parameters. Based on these results, it is concluded that future studies should prioritise therapeutic strategies, while investments should only be made in prophylactic strategies that have demonstrated promising results in vineyards. This is part of the PhD of Henri Taneli Pusa in the context of the H2020 ITN MicroWine, together with another PhD student of the ITN, Ifigeneia Kyrkou. Ifigeneia was the first author of the paper [18] but the mathematical model is the work of Taneli.

GENSCALE Project-Team

7. New Results

7.1. Data Structure

7.1.1. Quasi-dictionary data structure

Participants: Camille Marchet, Lolita Lecompte, Pierre Peterlongo.

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

The quasi-dictionary, is freely available at https://github.com/pierrepeterlongo/quasi_dictionary. The associated paper has been published in Discrete Applied Mathematics [17].

7.2. Algorithms & Methods

7.2.1. Genome assembly of targeted organisms in metagenomic data

Participants: Wesley Delage, Cervin Guyomar, Fabrice Legeai, Claire Lemaitre.

In this work, we propose a two-step reference-guided assembly method tailored for metagenomic data. First, a subset of the reads belonging to the species of interest are recruited by mapping and assembled *de novo* into backbone contigs using a classical assembler. Then an all-versus-all contig gap-filling is performed using a modified version of MindTheGap with the whole metagenomic dataset. The originality and success of the approach lie in this second step, that enables to assemble the missing regions between the backbone contigs, which may be regions absent or too divergent from the reference genome. The result of the method is a genome assembly graph in gfa format, accounting for the potential structural variations identified within the sample. We showed that this method is able to assemble the *Buchnera aphidicola* genome in a single contig in pea aphid metagenomic samples, even when using a divergent reference genome, it runs at least 5 times faster than classical *de novo* metagenomics assemblers and it is able to recover large structural variations co-existing in a sample. The modified version of MindTheGap is freely available at http://github.com/GATB/MindTheGap (version $>$ 2.1.0) [31].

7.2.2. De Novo Clustering of Long Reads by Gene from Transcriptomics Data

Participants: Camille Marchet, Lolita Lecompte, Jacques Nicolas, Pierre Peterlongo.

Long-read sequencing currently provides sequences of a few thousand base pairs. It is therefore possible to obtain complete transcripts, offering an unprecedented vision of the cellular transcriptome. However the literature lacks tools for *de novo* clustering of such data, in particular for Oxford Nanopore Technologies reads, because of the inherent high error rate compared to short reads. Our goal is to process reads from whole transcriptome sequencing data accurately and without a reference genome in order to reliably group reads coming from the same gene. This *de novo* approach is therefore particularly suitable for non-model species, but can also serve as a useful pre-processing step to improve read mapping. Our contribution both proposes a new algorithm adapted to clustering of reads by gene and a practical and free access tool that allows to scale the complete processing of eukaryotic transcriptomes. We sequenced a mouse RNA sample using the MinION device. This dataset is used to compare our solution to other algorithms used in the context of biological clustering. We demonstrate that it is the best approach for transcriptomics long reads. When a reference is available to enable mapping, we show that it stands as an alternative method that predicts complementary clusters.

The tool, called CARNAC-LR, is freely available at https://github.com/kamimrcht/CARNAC-LR. This work has been published in Nucleic Acids Research journal [16] and presented in several conferences [33], [28].

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

Participant: Camille Marchet.

In this work we compared an assembly-first and a mapping-first approach to analyze RNA-seq data and find alternative splicing (AS) events. Assembly-first approach enables to identify novel AS events and to detect events in paralog genes that are hard to find using mapping because of multiple equivalent matches. On the other hand, the mapping-first approach is more sensitive and detects AS events in lowly expressed genes, and is also able to find AS events with exons containing transposable elements. In addition we support these results with experimental validation. We showed that in order to extensively study the alternative splicing via RNA-seq data and retrieve the most candidates, both approaches should be led. We provide a pipeline consituted of parallel local *de novo* assembly executed by KisSplice and mapping using a novel mapping workflow called FaRLine [11].

7.2.4. Short read correction

Participant: Pierre Peterlongo.

We proposed a new method to correct short reads using de Bruijn graphs, and we implemented it as a tool called Bcool. As a first step, Bcool constructs a corrected compacted de Bruijn graph from the reads. This graph is then used as a reference and the reads are corrected according to their mapping on the graph. We showed that this approach yields a better correction than kmer-spectrum techniques, while being scalable, making it possible to apply it to human-size genomic datasets and beyond [27].

7.2.5. Long read splitting of heterozygous genomes

Participants: Dominique Lavenier, Maxime Bridoux.

This study aims to directly split long reads of highly heterozygous genomes to help assembly. Long read technologies provide very noisy sequences with many short indel errors. Standard assembly software do not really make difference between heterozygosity and sequencing errors. For highly heterozygous genomes this confusion may lead to misassembly. To separate long reads accordingly to their haplotype, we developed a new k-mer based method. After an alignment step to group similar reads, we build slices of 1 kbp along the multiple alignment containing a representative number of reads. The splitting is done by focusing on k-mers that are absent in one group and not in another one. This in an ongoing work started by the internship of M. Bridoux [29] in the framework of the France Genomique ALPAGA project.

7.3. Optimisation

7.3.1. Distance-Constrained Elementary Path Problem

Participants: Sebastien François, Rumen Andonov.

Given a directed graph $G = (V, E, l)$ with weights $l_e \geq 0$ associated with arcs $e \in E$ and a set of vertex pairs with distances between them (called *distance constraints*), the problem is to find an elementary path in G that satisfies a maximum number of distance constraints. We call it *Distance-Constrained Elementary Path (DCEP)* problem. This problem is motivated by applications in genome assembly. We describe three Mixed Integer Programming (MIP) formulations for this problem and discuss their advantages [25].

7.3.2. Complete Assembly of Circular Genomes Based on Global Optimization

Participants: Sebastien François, Rumen Andonov, Dominique Lavenier.

The goal here is to develop a new methodology and tools based on strong mathematical foundations and novel optimization techniques for solving the genome assembly problem. During the current year we focused on the last two stages of genome assembly, namely scaffolding and gap-filling, and showed that they can be solved as part of a single optimization problem. We obtained this by modeling genome assembly as a problem of finding a simple path in a specific graph that satisfies as many as possible of the distance constraints encoding the insert-size information. We formulated it as a mixed-integer linear programming problem and applied an optimization solver to find the exact solution on a benchmark of chloroplasts. Our tool is called GAT (Genscale Assembly Tool) and we tested it on a set of 33 chloroplast genome data. Comparisons with some of the most popular recent assemblers show that our tool produces assemblies of significantly higher quality than these heuristics [26]. These results fully justify the efforts for designing exact approaches for genome assembly.

7.4. Parallelism

7.4.1. Variant detection using processing-in-memory technology

Participants: Dominique Lavenier, Mohamed Moselhy.

The concept of Processing-In-Memory aims to dispatch the computer power near the data. Together with the UPMEM company (http://www.upmem.com/), which is currently developing a DRAM memory enhanced with computing units, we parallelized the detection of small mutations on the human genome. Traditionally, this process is split into 2 steps: a mapping step and a variant calling step. Here, thanks to the high processing power of this new type of memory, the mapping step can nearly be done at the disk transfer rate. In 2018, we define an ad-hoc data structure allowing the variant calling step to be performed simultaneously on the host processor. Basically, the two steps are overlapped in such a way that reads are mapped by packet. When a packet is mapped, the mapping results of the previous one dynamically feed the variant calling data structure. Performance evaluation on the FPGA UPMEM memory prototype indicates a very high speed-up (two orders of magnitude) compared with state-of-the-art software (specifically GATK).

7.5. Benchmarks and Reviews

7.5.1. Evaluation of error correction tools for long Reads

Participants: Lolita Lecompte, Camille Marchet, Pierre Peterlongo.

Long read technologies, Pacific Biosciences and Oxford Nanopore, have high error rates (from 9% to 30%). Hence, numerous error correction methods have been recently proposed, each based on different approaches and, thus, providing different results. As this is important to assess the correction stage for downstream analyses, we designed the ELECTOR software, providing evaluation of long read correction methods. This software generates additional quality metrics compared to previous existing tools. It also scales to very long reads and large datasets and is compatible with a wide range of state-of-the-art error correction tools.

ELECTOR is freely available at https://github.com/kamimrcht/ELECTOR. It as been presented during the Jobim2018 conference [32]

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

Participant: Pierre Peterlongo.

We took part to the redaction of the review paper that proposes a state of the art of the pan-genomics current status, methods and future orientations. This paper has been published in *Briefings in Bioinformatics* [18].

7.6. Bioinformatics Analysis

7.6.1. Metagenomic analysis of pea aphid symbiotic communities

Participants: Cervin Guyomar, Fabrice Legeai, Claire Lemaitre.

We worked on a methodological framework adapted to the study of genomic diversity and evolutionary dynamics of the pea aphid symbiotic community from an extensive set of metagenomics datasets. The framework is based on mapping to reference genomes and whole genome SNP-calling. We explored the genotypic diversity associated to the different symbionts of the pea aphid at several scales : across host biotypes, amongst individuals of the same biotype, and within individual aphids. Thorough phylogenomic analyses highlighted that the evolutionary dynamics of symbiotic associations strongly varied depending on the symbiont, reflecting different evolutionary histories and possible constraints [14].

7.6.2. Analysis of pea aphid genomic polymorphism

Participants: Fabrice Legeai, Claire Lemaitre.

We participated in the analyses of a large re-sequencing dataset of pea aphid individuals and populations. We performed the data cleaning, mapping to the reference genome and variant calling steps. The resulting polymorphism data shed light on two novel findings regarding the pea aphid genome evolution.

First, we showed that relaxed selection is likely to be the greatest contributor to the faster evolution of the X chromosome compared to autosomes [15]. Secondly, we looked for genomic bases of adaptation to novel environments, and identified 392 genomic hotspot regions of differentiation spanning 47.3 Mb and 2,484 genes. Interestingly, these hotspots were significantly enriched for candidate gene categories that are related to host–plant selection and use. These genes represent promising candidates for the genetic basis of host–plant specialization and ecological isolation in the pea aphid complex [21].

7.6.3. A de novo approach to disentangle partner identity and function in holobiont systems

Participants: Camille Marchet, Pierre Peterlongo.

Study of meta-transcriptomic datasets involving non-model organisms represents bioinformatic challenges that affect the study of holobiont meta-transcriptomes. Hence, we proposed an innovative bioinformatic approach and tested it on marine models as a proof of concept.

We considered three holobiont models, of which two transcriptomes were previously published and a yet unpublished transcriptome, to analyze and sort their raw reads using Short Read Connector (see section 7.1.1). Before assembly, we thus defined four distinct categories for each holobiont meta-transcriptome: host reads, symbiont reads, shared reads, and unassigned reads. Afterwards, we observed that independent *de novo* assemblies for each category led to a diminution of the number of chimeras compared to classical assembly methods. Moreover, the separation of each partner's transcriptome offered the independent and comparative exploration of their functional diversity in the holobiont. Finally, our strategy allowed to propose new functional annotations for two well-studied holobionts (a Cnidaria-Dinophyta, a Porifera-Bacteria) and a first meta-transcriptome from a planktonic Radiolaria-Dinophyta system forming widespread symbiotic association for which our knowledge is considerably limited [19].

7.6.4. Whole genome detection of micro-satellites

Participant: Dominique Lavenier.

This study has been done in cooperation with the federal university of de São João del-Rei, Brazil. The objective was to locate tens of thousands of micro-satellite loci for an endangered piracema (i.e. migratory) South American fish, *Brycon orbignyanus*. Together with the Brazil group we designed a specific pipeline that first assembles short paired-end reads into contigs and then performs micro-satellite oriented scaffolding processing [23].

7.6.5. Analysis of the genes and genomes involved in plant and insects interactions

Participant: Fabrice Legeai.

These study has been done in cooperation with various laboratories. In particular, we characterized the effectors (secreted proteins suppressing plant defense) of the pea aphid fed on different plants, by firstly identifying these genes in the pea aphid genome, then studying and comparing their expression between different conditions, and then finally by observing their evolution among a broad set of phytophagous insects [10]. We also identified microRNAs from smallRNA datasets from *Spodoptera frugiperda* strains fed on different host-plants [20]. Finally, we predicted the transposable elements in the genome of *Cephus cinctus*, an important insect pest [22].

7.6.6. Analysis of the expression and identification of the targets of mir202 during the medaka oogenesis

Participant: Fabrice Legeai.

This study has been done in cooperation with the INRA LPGP laboratory (Rennes). Its goal was to identify the role of small non-coding RNAs in the regulation of the reproduction of the fish model *Oryzias latipes* (medaka). We predicted the putative targets of the microRNA miR202, already observed as being specifically expressed in gonads. In the second part of the work, we identified important genes and functions targeted by miR202 and differentially expressed in the gonads when the microRNA was artificially repressed [13].

IBIS Project-Team

6. New Results

6.1. 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). Over the years, many useful resources have appeared, such as libraries of reporter strains for model organisms and computer tools for designing reporter plasmids. Moreover, the widespread adoption of thermostated microplate readers in experimental laboratories has made it possible to automate and multiplex reporter gene assays on the population level. This has resulted in large time-series data sets, typically comprising $10^5 - 10^6$ measurements of absorbance, fluorescence, and luminescence for 10^3 wells on the microplate. In order to fully exploit these data sets, we need sound mathematical methods to infer biologically relevant quantities from the primary data and computer tools to apply the methods in an efficient and user-friendly manner.

In the past few years we 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 to be solved, notably the inference of growth rate, promoter activity, and protein concentration profiles. The linear inversion methods, published in *Bioinformatics* in 2015 [12], have been implemented in the Python package WELLFARE and integrated in the web application WELLINVERTER. Funded by a grant from the Institut Français de Bioinformatique (IFB), we improved WellInverter by developing a parallel computational architecture with a load balancer to distribute the analysis queries over several back-end servers, a new graphical user interface, and a plug-in system for defining high-level routines for parsing data files produced by microplate readers from different manufacturers. This has resulted in a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. This year the web service has been redeployed on the new IFB cloud and on an Inria server, accompanied by extensive user documentation, online help, and a tutorial. Moreover, we submitted a journal paper on WELLINVERTER illustrating the use of the tool by analyzing data of the expression of a fluorescent reporter gene controlled by a phage promoter in growing *Escherichia coli* populations. We notably show that the expression pattern in different growth media, supporting different growth rates, corresponds to the pattern expected for a constitutive gene.

Compared to most reporter gene assays based on fluorescence proteins, luciferase reporters have a superior signal-to-noise ratio, since they do not suffer from the high autofluorescence background of the bacterial cell. At the same time, however, luciferase reporters have the drawback of constant light emission, which leads to undesired cross-talk between neighbouring wells on a microplate. To overcome this limitation, Marco Mauri in collaboration with colleagues from the Philipps-Universität Marburg developed a computational method to correct for luminescence bleed-through and to estimate the "true" luminescence activity for each well of a microplate. As the sole input our algorithm uses the signals measured from a calibration plate, in which the light emitted from a single luminescent well serves as an estimate for the "light-spread function". We show that this light-spread function can be used to deconvolve any other measurement obtained under the same technical conditions. Our analysis demonstrates that the correction preserves low-level signals close to the background and shows that it is universally applicable to different kinds of microplate readers and plate types. A journal article on this work was submitted this year.

6.2. Microdomain formation of bacterial membrane proteins

Fluorescent reporters can be used not only to quantify gene expression, but also to localize proteins in different compartments of the cell, In particular, in bacteria proteins within the cytoplasmic membrane display distinct localization patterns and arrangements. While multiple models exist describing the dynamics of membrane proteins, to date there have been few systematic studies, particularly in bacteria, to evaluate how protein size, number of transmembrane domains, and temperature affect their diffusion, and if conserved localization patterns exist.

Marco Mauri in collaboration with colleagues from the Philipps-Universität in Marburg has used fluorescence microscopy, single-molecule tracking (SMT), and computer-aided visualization methods to obtain a better understanding of the three-dimensional organization of bacterial membrane proteins, using the model bacterium *Bacillus subtilis*. First, we carried out a systematic study of the localization of over 200 *B. subtilis* membrane proteins, tagged with monomeric mVenus-YFP at their original gene locus. Their subcellular localization could be discriminated in polar, septal, patchy, and punctate patterns. Almost 20% of membrane proteins specifically localized to the cell poles, and a vast majority of all proteins localized in distinct structures, which we term microdomains. Dynamics were analyzed for selected membrane proteins, using SMT. Diffusion coefficients of the analyzed transmembrane proteins did not correlate with protein molecular weight, but correlated inversely with the number of transmembrane helices, *i.e.*, transmembrane radius. We observed that temperature can strongly influence diffusion on the membrane, in that upon growth temperature upshift, diffusion coefficients of membrane proteins increased and still correlated inversely to the number of transmembrane domains, following the Saffman–Delbrück relation.

The vast majority of membrane proteins were observed to localize to distinct multimeric assemblies. Diffusion of membrane proteins can be suitably described by discriminating diffusion coefficients into two protein populations, one mobile and one immobile, the latter likely constituting microdomains. Moreover, in this study published in *BMC Biology* [18], we provided a method to correct the diffusion coefficient for the membrane curvature. Our results show there is high heterogeneity and yet structural order in the cell membrane, and provide a roadmap for our understanding of membrane organization in prokaryotes. Given the exceptionally richness of the data obtained, both further analysis on membrane lateral movements and a more detailed theory on the effect of membrane curvature is possible.

6.3. 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 rather than the deterministic models underlying the linear inversion methods discussed in Section 6.6 . 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 parametric intrinsic noise models, on the one hand, and the nonparametric inference of gene expression statistics, on the other hand, from population snapshot data. Along with modelling and inference, identifiability analysis is dedicated special attention. Other problems related with single-cell modelling, extracellular variability and inheritance of traits at cell division are considered also through external collaborations, as discussed below (Section 6.4).

Concerning identification of intrinsic noise dynamics in single cells, previous results on the contribution of stochasticity to parameter identifiability and on reconstruction of unknown gene regulatory networks have been taken further. For the case of population snapshot meaurements, where the dynamics of the population statistics are observed by simple time-lapse experiments, our earlier results showing that variance measurements may provide tremendous improvement in network reconstruction relative to sole mean measurements have been developed into a full-blown method for first-order gene network reconstruction. Additionally, parameter identifiability methods and results initially developed for gene expression models have been generalized to the whole class of first-order stochastic reaction networks. These developments have been presented and demonstrated by simulation in a paper published in the journal *Processes* [15].

Reconstruction of promoter activity statistics from reporter gene population snapshot data has been further investigated, leading to a full-blown spectral analysis and reconstruction method for reporter gene systems. In

the context of the ANR project MEMIP (Section 7.2), we have characterized reporter systems as noisy linear systems operating on a stochastic input (promoter activity), and developed an inversion method for estimation of promoter activation statistics from reporter population snapshots that is nonparametric, in the sense that it does not assume any parametric model for the unknown promoter dynamics. This analysis rests on a more general, original generalization of moment equations that we have developed for stochastic reaction networks with state-affine rates subject to random input processes. The theoretical results, together with a demonstration of the reporter gene inversion method on simulated data, have been accepted for publication in *Automatica* this year [16]. In addition to utilization of the method on real gene-expression data, the results lend themselves to several additional applications, among which the study of extrinsic noise and the optimal design of reporter systems.

6.4. Modelling and analysis of cellular trait dynamics over lineage trees

The investigation of cellular populations at a single-cell level has already led to the discovery of important phenomena, such as the occurrence of different phenotypes in an isogenic population. Nowadays, several experimental techniques, such as microscopy combined with the use of microfluidic devices, enable one to take investigation further by providing time-profiles of the dynamics of individual cells over entire lineage trees. The difficulty, and at the same time the opportunity, in exploiting these data and inferring mathematical models from them is the fact that the behavior of different cells is correlated because of inheritance.

From the modelling point of view, lineage trees are well described by structured branching population models where the life cycle of each cell depends on individual characteristics, such as size and internal protein dynamics, which play a key role in the mechanisms of cell division. One important aspect in the analysis of these population models consists in the investigation of biases arising from the sampling of a finite set of observed individuals. In order to characterize bias, we studied the dynamics of a structured branching population where the trait of each individual evolves in accordance with a Markov process. We assumed that the rate of division of each individual is a function of its trait and when a branching event occurs, the trait of the descendants at birth depends on their number and on the trait of the mother. We explicitly described the Markov process, named auxiliary process, corresponding to the dynamics of the trait of a "typical" individual by deriving its associated infinitesimal generator. In particular, we proved that this process characterizes exactly the process of the trait of a uniformly sampled individual in a large population approximation. This work, carried out by Aline Marguet, has been accepted for publication in the journal *Bernoulli* [19].

We also investigated the long-time behavior of the population and proved that a typical individual in the population asymptotically behaves like the auxiliary process previously introduced. These results have been submitted for publication [22]. Structured branching processes and their analysis also provide the basis for identification tools for lineage-tree data. In particular, in the context of a bifurcating Markov chain, where each individual is characterized by a trait evolving in accordance with a scalar diffusion, we proved that the maximum-likelihood estimator of the division rate is asymptotically efficient and demonstrate the method on simulated data. This work, in collaboration with M. Hoffmann at Univ Paris-Dauphine, has also been submitted for publication [21].

Along the same lines, modelling and identification of gene expression models with mother-daughter inheritance are being investigated in the context of the ANR project MEMIP. Starting from an earlier work of the group [7], with reference to an application on osmotic shock response by yeast, the key question is to what extent leveraging an inheritance model improves inference of individual cell dynamics as well as of inheritance dynamics themselves, relative to state-of-art approaches where inheritance is not accounted for at a modelling stage.

6.5. 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.

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

The same focus on the dynamics of physiological processes has shaped a project on the post-transcriptional control of carbon central metabolism in *E. coli*. In the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique INRA-Inria, the collaboration of Delphine Ropers with Muriel Cocaign-Bousquet and Brice Enjalbert at INRA/INSA Toulouse has demonstrated the key role played by the post-transcriptional regulatory system CSR in growth transitions in a series of publications in the past few years (*e.g.*, [9]). 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. This work is further supported by the ANR project ECORIB accepted this year and an IXXI grant in collaboration with the ERABLE project-team (Section 7.2).

6.6. Modelling bacterial growth

Various mathematical approaches have been used in the literature to describe the networks of biochemical reactions involved in microbial growth. With various levels of detail, the resulting models provide an integrated view of these reaction networks, including the transport of nutrients from the environment and the metabolism and gene expression allowing the conversion of these nutrients into biomass. The models hence bridge the scale between individual reactions to the growth of cell populations. Analysing the dynamics of some of these models mentioned above becomes quickly intractable, when mathematical functions are for instance given by complex algebraic expressions resulting from the mass balance of biochemical reactions. In a paper published in the *Bulletin of Mathematical Biology* [13], Edith Grac, former post-doc in IBIS, Delphine Ropers, and Stefano Casagranda and Jean-Luc Gouzé from the BIOCORE project-team, have studied how monotone system theory and time-scale arguments can be used to reduce high-dimension models based on the mass-action law. Applying the approach to an important positive feedback loop regulating the expression of RNA polymerase in *E. coli*, made it possible to study the stability of the system steady states and relate the dynamical behaviour of the system to observations on the physiology of the bacterium *E. coli*.

In another paper published in *BMC Systems Biology* [14], Delphine Ropers and BIOCORE members Stefano Casagranda, Jean-Luc Gouzé, and Suzanne Touzeau, have developed a new approach to deal with model complexity. The approach, named Principle Process Analysis, allows to identify processes playing a key role in the model dynamics and to reduce the complex dynamics to these core processes, omitting processes that are

inactive. In particular, it has allowed the reduction of a well-known model of circadian rhythms in mammals into a succession of simpler submodels. Their analysis has resulted in the identification of the source of circadian oscillations, the main oscillator being the negative feedback loop involving proteins PER, CRY, CLOCK-BMAL1, in agreement with previous modelling and experimental studies.

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. In recent years, in the framework of the PhD thesis of Nils Giordano, we extended the study of microbial growth strategies to dynamical environments, using a self-replicator model. In collaboration with the BIOCORE project-team, we formulated dynamical growth maximization as an optimal control problem that can be solved using Pontryagin's Maximum Principle and we compared the theoretical results thus obtained with different possible implementations of growth control in bacterial cells [5]. The extension and experimental validation of some of these results are currently being carried out by Antrea Pavlou in the framework of her PhD project, funded by the ANR project Maximic (Section 7.2).

6.7. Growth control in bacteria and biotechnological applications

The ability to experimentally control the growth rate is crucial for studying bacterial physiology. It is also of central importance for applications in biotechnology, where often the goal is to limit or even arrest growth. Growth-arrested cells with a functional metabolism open the possibility to channel resources into the production of a desired metabolite, instead of wasting nutrients on biomass production. In recent years we obtained a foundation result for growth control in bacteria [6], in that we engineered an *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.

The experimental work underlying the growth switch has been continued in several directions in the context of the Maximic project by Célia Boyat. Moreover, in collaboration with colleagues from the BIOCORE project-team, we have formulated the maximization of metabolite production by means of the growth switch as a resource reallocation problem that can be analyzed by means of the self-replicator models of bacterial growth mentioned in Section 6.6 in combination with methods from optimal control theory. In a publication accepted for the *Journal of Mathematical Biology* this year [20], we study various optimal control problems by means of a combination of analytical and computational techniques. We show that the optimal solutions for biomass maximization and product maximization are very similar in the case of unlimited nutrient supply, but diverge when nutrients are limited. Moreover, external growth control overrides natural feedback growth control and leads to an optimal scheme consisting of a first phase of growth maximization followed by a second phase of product maximization. This two-phase scheme agrees with strategies that have been proposed in metabolic engineering. More generally, this work shows the potential of optimal control theory for better understanding and improving biotechnological production processes. Extensions concerning the effect on growth and bioproduction of the (biological or technological) costs associated with discontinuous control strategies, and of the time allotted to optimal substrate utilization, are described in a contribution to a control theory conference submitted this year.

LIFEWARE Project-Team

7. New Results

7.1. Graphical Requirements for Multistationarity in CRNs and their Verification in BioModels

Participants: Adrien Baudier, François Fages, Sylvain Soliman.

Thomas's necessary conditions for the existence of multiple steady states in gene networks have been proved by Soulé with high generality for dynamical systems defined by differential equations. When applied to (protein) reaction networks however, those conditions do not provide information since they are trivially satisfied as soon as there is a bimolecular or a reversible reaction. Refined graphical requirements have been proposed to deal with such cases. In [1], we present for the first time a graph rewriting algorithm for checking the refined conditions given by Soliman, and evaluate its practical performance by applying it systematically to the curated branch of the BioModels repository. This algorithm analyzes all reaction networks (of size up to 430 species) in less than 0.05 second per network, and permits to conclude to the absence of multistationarity in 160 networks over 506. The short computation times obtained in this graphical approach are in sharp contrast to the Jacobian-based symbolic computation approach. We also discuss the case of one extra graphical condition by arc rewiring that allows us to conclude on 20 more networks of this benchmark but with a high computational cost. Finally, we study with some details the case of phosphorylation cycles and MAPK signalling models which show the importance of modelling the intermediate complexations with the enzymes in order to correctly analyze the multistationarity capabilities of such biochemical reaction networks.

7.2. Influence Networks compared with CRNs: Semantics, Expressivity and Attractors

Participants: François Fages, Thierry Martinez [former member], David Rosenblueth [former member], Sylvain Soliman, Denis Thieffry.

Biochemical reaction networks are one of the most widely used formalism in systems biology to describe the molecular mechanisms of high-level cell processes. However modellers also reason with influence diagrams to represent the positive and negative influences between molecular species and may find an influence network useful in the process of building a reaction network. In [4], we introduce a formalism of influence networks with forces, and equip it with a hierarchy of Boolean, Petri net, stochastic and differential semantics, similarly to reaction networks with rates. We show that the expressive power of influence networks is the same as that of reaction networks under the differential semantics, but weaker under the discrete semantics. Furthermore, the hierarchy of semantics leads us to consider a (positive) Boolean semantics without test for absence, that we compare with the (negative) Boolean semantics with test for absence of gene regulatory networks à la Thomas. We study the monotonicity properties of the positive semantics and derive from them an algorithm to compute attractors in both the positive and negative Boolean semantics. We illustrate our results on models of the literature about the p53/Mdm2 DNA damage repair system, the circadian clock, and the influence of MAPK signaling on cell-fate decision in urinary bladder cancer.

As an application, in [11] methods are shown to add dynamics to large molecular influence maps.

7.3. Reducing CRNs by Tropicalization

Participants: Eléonore Bellot, François Fages, Aymeric Quesne, Sylvain Soliman, Eliott Suits.

We have shown in the past that model reduction relationships between CRNs can be detected on a large scale by the graph matching notion of subgraph epimorphism [0], furthermore quite efficiently with constraint programming or SAT solving techniques. However this approach does not allow us to actually reduce models. In the framework of the ANR-DFG SYMBIONT project [10] we are investigating model reduction methods based on tropicalization and constraint programming techniques [0] together with correctness conditions based on Tikhonov theorem.

7.4. Compiling mathematical functions and programs in CRNs

Participants: Auriane Cozic, Elisabeth Degrand, François Fages, Mathieu Hemery, Wei-Chih Huang, Lena Le Quellec, Sylvain Soliman.

In a previous paper, we have proven that any computable function over the reals in the sense of computable analysis (i.e. computable with finite yet arbitrary precision by a Turing machine) is computable by a continuous CRN over a finite set of molecular species. In this approach, the real-valued molecular concentrations are the information carriers and computation can be purely analog. We have derived from the proof of this result a compiler of real functions (of either time or input concentrations) specified by polynomial initial value problems (PIVP) in elementary CRNs. This compiler makes it possible to automate the design of abstract CRNs for implementing arbitrary computable functions over the reals presented by PIVPs, in particular arithmetic, trigonometric, sigmoid and logical functions. The compilation of sequentiality, program control flows and mixed analog-digital imperative programs lead us however to consider more efficient implementations of Heavyside functions with simple CRNs that have no simple mathematical expression as input/output functions. Our goal is to develop a compiler of high-level mixed analog-digital programs in efficient abstract CRNs amenable to practical implementation with real enzymes in DNA-free vesicles, as illustrated in Section 7.8 .

7.5. Evolving CRNs from data time series

Participants: Elisabeth Degrand, François Fages, Jérémy Grignard [former&future Member], Mathieu Hemery, Sylvain Soliman.

Another approach to CRN design is by evolutionary algorithms. Given a function given by its graph with a finite set of points, and using the same framework based on PIVPs as above, we have designed a genetic algorithm which interleaves the evolution of a population of PIVPs with parameter optimization using CMA-ES for fitting the input curve. On the cosine function, this algorithm recovers PIVPs equivalent to the standard PIVP for cosine, while on Heavyside functions, the algorithm finds (mathematically mysterious) CRNs that are much simpler than Hill functions of high order for instance.

7.6. Learning CRNs from data time series

Participants: François Fages, Jérémy Grignard [former&future Member], Nicolas Levy, Julien Martinelli, Sylvain Soliman.

The problem of learning a mechanistic model from observed data is more difficult than learning a blackbox model fitting the data, due to the difference between causal relationships and correlations. In a biological context, learning a mechanistic model from experimental data would help understanding the underlying biological processes. To that end, considering multiple time series data generated by a hidden CRN from different initial states (by either stochastic or differential simulation), we develop a clustering-based algorithm for the inference of biological reaction networks. The output is a set of reactions which can be used to generate new traces. A model selection method is derived from these newly generated traces. We evaluate the performance of this algorithm on a range of models from Biomodels.

[0] Steven Gay, François Fages, Thierry Martinez, Sylvain Soliman, Christine Solnon. On the subgraph Epimorphism Problem. Discrete Applied Mathematics, 162:214–228, 2014.

[0] Sylvain Soliman, François Fages, Ovidiu Radulescu. A constraint solving approach to model reduction by tropical equilibration. Algorithms for Molecular Biology, 9(24), 2014.

7.7. Optimizing CRN robustness

Participants: François Fages, Lucia Nasti, Sylvain Soliman.

In [7] we present two complementary notions of robustness of a system with respect to a property of its behaviour expressed in temporal logic: first the statistical notion of model robustness to parameter perturbations, defined as its mean functionality; and second, a metric notion of formula satisfaction robustness, defined as the penetration depth in the validity domain of the temporal logic constraints. We show how the formula robustness can be used in BIOCHAM-4 with no extra cost as an objective function in the parameter optimization procedure, to actually improve CRN robustness. We illustrate these unique features with a classical example of the hybrid systems community and provide some performance figures on a model of MAPK signalling with 37 parameters.

7.8. Robust biochemical programming of synthetic microreactors

Participants: Auriane Cozic, François Fages, Wei-Chih Huang, Lena Le Quellec, Lucia Nasti, Sylvain Soliman.

Biological systems have evolved efficient sensing and decision-making mechanisms to maximize fitness in changing molecular environments. Synthetic biologists have exploited these capabilities to engineer control on information and energy processing in living cells. While engineered organisms pose important technological and ethical challenges, de novo assembly of non-living biomolecular devices could offer promising avenues towards various real-world applications. However, assembling biochemical parts into functional information processing systems has remained challenging due to extensive multidimensional parameter spaces that must be sampled comprehensively in order to identify robust, specification compliant molecular implementations. In [3], we introduce a systematic methodology based on automated computational design and microfluidics enabling the programming of synthetic cell-like microreactors embedding biochemical logic circuits, or protosensors, to perform accurate biosensing and biocomputing operations in vitro according to temporal logic specifications. We show that proof-of-concept protosensors integrating diagnostic algorithms detect specific patterns of biomarkers in human clinical samples. Protosensors may enable novel approaches to medicine and represent a step towards autonomous micromachines capable of precise interfacing of human physiology or other complex biological environments, ecosystems or industrial bioprocesses.

7.9. Identification of individual cells from z-stacks of bright-field microscopy images

Participants: Grégory Batt, Chiara Fracassi [former Member], Jean-Baptiste Lugagne [former Member].

Obtaining single cell data from time-lapse microscopy images is critical for quantitative biology, but bottlenecks in cell identification and segmentation must be overcome. In [5], we propose a novel, versatile method that uses machine learning classifiers to identify cell morphologies from z-stack bright-field microscopy images. We show that axial information is enough to successfully classify the pixels of an image, without the need to consider in focus morphological features. This fast, robust method can be used to identify different cell morphologies, including the features of *E. coli, S. cerevisiae* and epithelial cells, even in mixed cultures. Our method demonstrates the potential of acquiring and processing Z-stacks for single-layer, single-cell imaging and segmentation.

7.10. Applying ecological resistance and resilience to dissect bacterial antibiotic responses

Participants: Virgile Andreani, Grégory Batt.

An essential property of microbial communities is the ability to survive a disturbance. Survival can be achieved through resistance, the ability to absorb effects of a disturbance without a notable change, or resilience, the ability to recover after being perturbed by a disturbance. These concepts have long been applied to the analysis of ecological systems, although their interpretations are often subject to debate. In [6], we show that this framework readily lends itself to the dissection of the bacterial response to antibiotic treatment, where both terms can be unambiguously defined. The ability to tolerate the antibiotic treatment in the short term corresponds to resistance, which primarily depends on traits associated with individual cells. In contrast, the ability to recover after being perturbed by an antibiotic corresponds to resilience, which primarily depends on traits associated with the population. This framework effectively reveals the phenotypic signatures of bacterial pathogens expressing extended-spectrum β-lactamases when treated by a β-lactamase antibiotic. Our analysis has implications for optimizing treatment of these pathogens using a combination of a β-lactamase and a β-lactamase inhibitor. In particular, our results underscore the need to dynamically optimize combination treatments based on the quantitative features of the bacterial response to the antibiotic or the Bla inhibitor.

MORPHEME Project-Team

6. New Results

6.1. Exact biconvex reformulation of the $\ell_2 - \ell_0$ minimization problem

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

We focus on the problem of minimizing the least-squares loss function under the constraint that the reconstructed signal is at maximum k-sparse. This is called the ℓ_2-ℓ_0 constrained problem. The ℓ_0 pseudo-norm counts the number of non-zero elements in a vector. The minimization problem is of interest in signal processing, with a wide range of applications such as compressed sensing, source separation, and super-resolution imaging.

Based on the results of [20], we reformulate the ℓ_0 pseudo-norm exactly as a convex minimization problem by introducing an auxiliary variable. We then propose an exact biconvex reformulation of the $\ell_2 - \ell_0$ constrained and penalized problems. We give correspondence results between minimizer of the initial function and the reformulated ones. The reformulation is biconvex. This property is used to derive a minimization algorithm.

We apply the algorithm to the problem of Single-Molecule Localization Microscopy and compare the results with the well-known IHT algorithm [13]. Both visually and numerically the biconvex reformulations perform better. This work has been presented at the iTWIST 2018 workshop [5].

Furthermore, the algorithm has been compared to the IRL1-CEL0 [14] and Deep-STORM [15] (see figure 1). The IRL1-CEL0 minimizes an exact relaxation [19] of the $\ell_2 - \ell_0$ penalized form and Deep-STORM is an algorithm that uses deep-learning and convolutional network to localize the molecules. This work has been accepted to the ISBI 2019 conference.

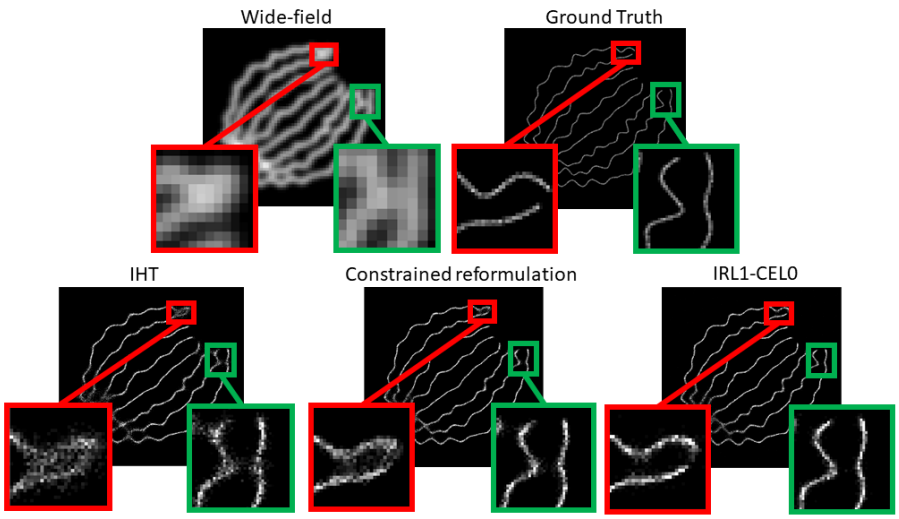


Figure 1. Reconstruction by the different algorithms. Data set from ISBI 2013 challenge [18].

6.2. Reconstruction of mosaic of microscopic images

Participants: Kevin Giulietti, Eric Debreuve, Grégoire Malandain.

This work takes place within the ANR PhaseQuant.

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

Such an imaging protocol is available on many microscopy software. Basically, displacements of the table on which lies the material to be imaged are programmed, and used to reconstruct the mosaic. However, it appears (at the overlapping areas) that a residual offset is still present. Analysis of acquisitions of both real and controlled experiments demonstrate that the table motions are not exactly reproducible (see figure 2), and that the cause of the offset is twofold: first a mis-alignement of the micrometer table axis with respect to the microscope axis, and second errors in the displacement computed by the micrometer table. Thanks to an image-based calculation of the axis mis-alignement, it has been shown that the first type error can easily be corrected.

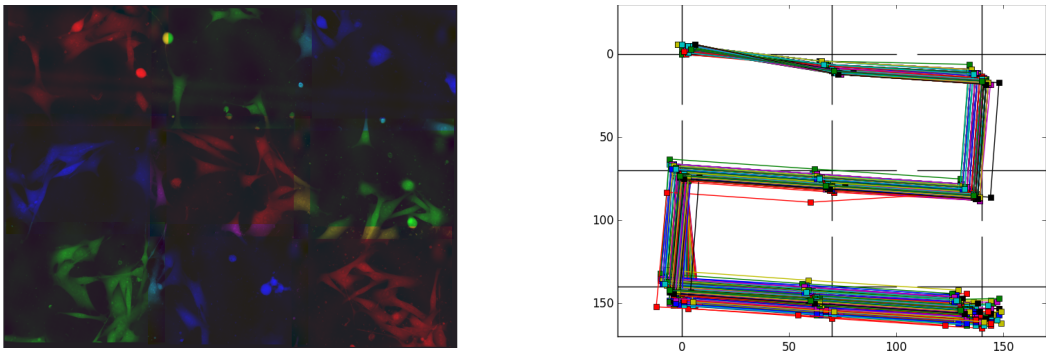


Figure 2. Example of a mosaic reconstructed for one acquisition timepoint. Estimation of the relative image position through time for the whole sequence (displacements with respect to the expected position have been magnified for visualization purpose).

6.3. Cytoplasm segmentation from cells confocal microscopy images

Participants: Somia Rahmoun, Fabienne de Graeve, Eric Debreuve, Xavier Descombes.

This work takes place within the ANR RNAGRIMP.

As part of the ANR project RNAGRIMP, two series of images have been acquired using fluorescence microscopy: one where the cell cytoplasm has been stained with GFP (Green Fluorescent Protein), the second where the nuclei have been stained with DAPI (4',6-diamidino-2-phenylindole). The first steps are detecting the nuclei on the DAPI images and learning a classification procedure into living cell or dead cell based on morphological and radiometric nuclei properties (average intensity, area, granularity, circularity ...).

A specific CellProfiler pipeline has been developed for this, and CellProfiler Analyst has been used to learn a decision tree for automatic nuclei (hence, cell) classification.

The next step is to segment the cell cytoplasms on the GFP images. Indeed, the target RNP-IMP granules appear in that compartment of the cell and are visible through their GFP response. This segmentation problem is particularly difficult due the heterogeneity of the cells intensity. This heterogeneity even appears within a given cell. Besides, cells sometimes form clusters in which there is no clear separation between adjacent cells.

In this context, we have considered a two steps algorithm to segment the cytoplasm. The first step consists of the image segmentation in small areas called superpixels that represent adjacent pixels with similar intensity. We have evaluated and compared different strategies (based on iterative clustering, minimum spanning tree, persistent edge selection ...) to achieve such a segmentation. Finally, we have selected an automatic algorithm based on the watershed transform. We are currently developing an algorithm to merge superpixels into the final segmentation.

Meanwhile, we have developed a supervised software to manually merge the superpixels (see Fig. 3). This tool can also be used by biologist to correct any segmentation error.

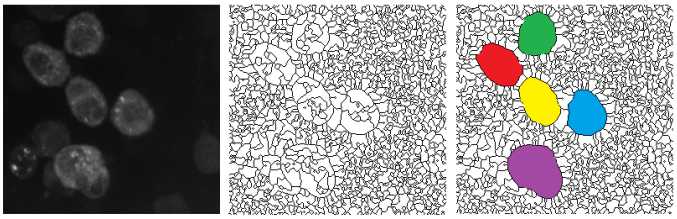


Figure 3. Superpixels merging: each color corresponds to a cell that is obtained by merging several superpixels.

6.4. Cytoneme detection and characterization

Participants: Eric Debreuve, Xavier Descombes.

This work is made in collaboration with Caterina Novelli, Tamas Matusek, Pascal Thérond (iBV).

This work is supported by the ANR project HMOVE. Cellular communication is one of the most important processes for controlling the morphogenesis of organs (i.e. the set of laws that determine the structure of tissues and organs during embryonic development). Understanding the communication both ways is an important issue in the field of developmental biology and it has recently been shown that the exchange of information between cells is controlled by long cellular extensions called "cytonemes". Last year, we had developed a pipeline for automatic cell membrane detection and cytoneme extraction from in vivo images obtained by confocal microscopy. When testing the proposed method on new images with varying acquisition conditions, we found it to be less reliable than expected. While retaining the same general philosophy (use of Frangi enhancement filter, skeletonization, and Dijkstra shortest path algorithm), we largely rethought the approach to make it more robust to acquisition conditions, more reliable in general, and faster (see Fig. 4). Some topological and geometrical features are then computed on the graph-based representation of the cytonemes in order to characterize in which respect wild-type and mutant conditions are different or similar. A journal paper is in preparation based on the analysis and interpretation of these results by our biologist colleagues.

6.5. Classification and Modeling of the Fibronectin Network in Extracellular Matrices

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

This work is done in collaboration with Ellen Van Obberghen-Schilling and Georgios Efthymiou (iBV).

We are interested in the numerical analysis and modeling of the Fibronectin (FN) network, a major extracellular matrix (ECM) molecule expressed in pathological states (fibrosis, cancer, etc). Our goal is to develop numerical quantitative biomarkers that describe the organization of the different FN networks from 2D confocal miscroscopy images (Figure 5).

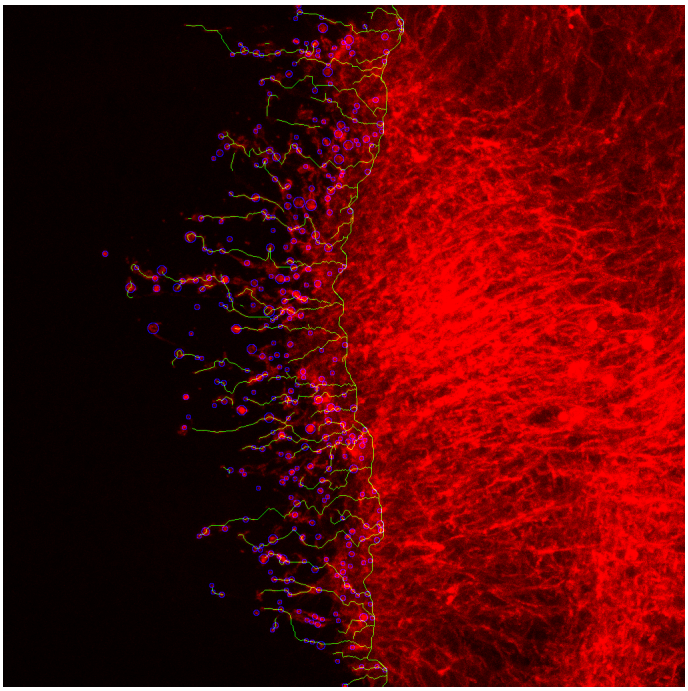


Figure 4. A result of membrane detection and cytoneme extraction.

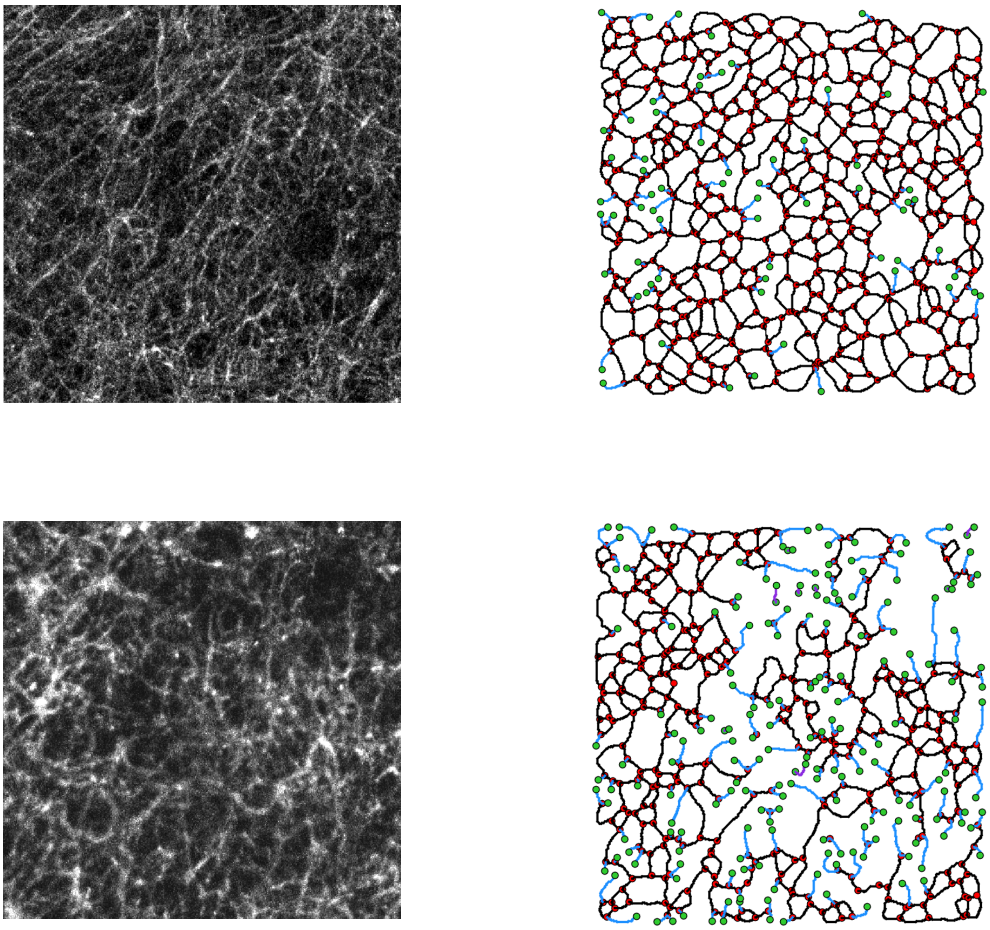


Figure 5. Different variants of FN and their associated graph networks. Top row: A+ fibronectin; bottom row: A-B- fibronectin. Left: confocal images; right: associated graphs.

In a previous work, we have derived a pipeline to classify a given tissue among the four FN variants (cell-derived matrices), based on a decomposition into discrete fast curvelet transform coefficients. We ensured the invariance to rotation of the coefficients and then fed them to a DAG-SVM multiclassifier, in order to prove their discriminative ability in the context of classification of the four FN variants. The results were published in [7].

The second step of our work consists in setting up the modeling of the FN networks starting from a graph-based representation, built on top of Gabor features (fiber scale, orientation, etc). More specifically, Gabor filters are used to enhance the fibrillar structures, followed by a morphological skeletonization of the maximum response of the Gabor filter set. We then derive the corresponding graph networks that generate relevant fiber geometry statistics (e.g fiber length, node degree, node density, etc).

Starting from the graph networks, we manage to reconnect the missing fibers in the skeleton, that are due to previous morphological operations. To do so, we use the Gabor maximum response as a guideline for reconnection, and connect the fibers within a predefined cone sector around the local fiber orientation. The graph parameters corresponding to the improved skeletonizations of the four FN variants, are then classified by a DAG-SVM. It is thus shown that graph features can discriminate among the FN variants.

The next step concerns the development of a metric between graph networks that takes into account their topology, to provide a meaningful distance between them. We currently investigate methods based on optimal transport, that are able to compare discrete probability distributions and respect the local geometry. The techniques that rely on graph structures to compute a geodesic distance (e.g. Gromov-Wasserstein) and/or barycenter of structured data (e.g. mesh structures) [16], serve as inspiration for our work. The distance is obtained following a minimization of the cost of transport of the mass from one distribution to the other. Despite the fact that they consider the intrinsic distance within each space (i.e graphs) in the cost formulation, these methods don't explicitly take into account the graph structure defined by the adjacency matrices. To counteract some of the shortcomings, we consider parallel graph-matching methods and redefine our problem in a many-to-many graph matching context, where the distance between the graphs is given by the optimal alignment of their structure determined by the mapping between the vertices [21].

Finally, we analyze the advantages and drawbacks of the two techniques both for small-size graphs as well as for FN graphs to derive an appropriate formulation of the distance among them, which will be useful to compare the FN fiber networks.

6.6. Detection of Brain Strokes Using Microwave Tomography

Participant: Laure Blanc-Féraud.

This work is done in collaboration with Vanna Lisa Coli and Juliette Leblond (EPI Factas, Inria Sophia), Pierre-Henri Tournier (Université Sorbonne, CNRS, LJLL, Inria, Paris) , Victorita Dolean (Université Côte d'Azur, CNRS, LJAD, Nice), Ibtissam El Kanfoud, Christian Pichot, Claire Migliaccio (Université Côte d'Azur, CNRS, LEAT, Sophia Antipolis).

Brain strokes are one of the leading causes of disability and mortality in adults in developed countries. The ischemic stroke (85% of total cases) and hemorrhagic stroke (15%) must be treated with opposite therapies, so that the determination of the stroke nature must be made quickly to apply the appropriate treatment. Recent works in biomedical imaging showed that strokes produce variations on brain tissues complex electric permittivity that can be detected by means of microwave tomography.

We present here some synthetic results obtained with an experimental microwave tomography-based portable system for the early detection and monitoring of brain strokes (Figure 6). The determination of electric permittivity requires the solution of a coupled direct-inverse problem, where massive parallel computation from domain decomposition method and regularization techniques for optimization methods are employed. Synthetic data are obtained with electromagnetic simulations and a noise model developed for the specific problem, which has been derived from measurements errors with the experimental imaging system.

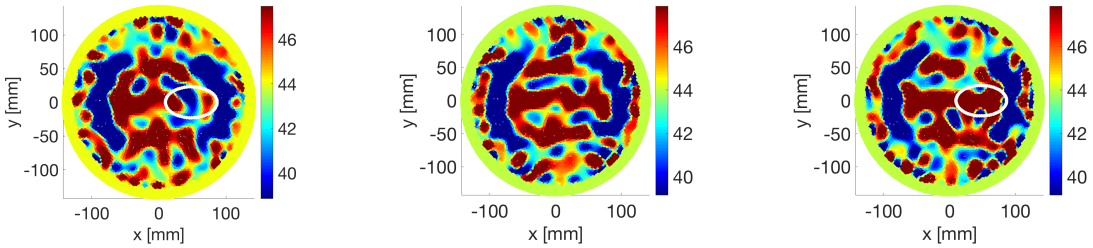


Figure 6. From left to right: brain with ischemic stroke, healthy brain and brain with hemorrhagic stroke (real part of permittivity ϵ_r).

6.7. Organoid growth tracking

Participants: Cédric Girard Riboulleau, Xavier Descombes.

This work is a collaboration with F.-R. Roustan, S. Torrino, S. Clavel and F. Bost from C3M. It was partially supported by the UCA Jedi Idex.

Organoid culture is a major challenge toward personalized medicine. It is now possible to partially reconstruct the structure of organs from a single biopsy. This new technology named organoid for sane cells or tumoroid for cancer cells allows the test of different molecules on cells withdrawn on a patient to retrieve the most efficient one for this patient. In this context, the main goal of this work is to develop a numerical scheme to automatically assess the effect of a given treatment on the organoid growth. We consider a time sequence of 2D confocal microscopy images of a population of organoids. We have considered different approaches to detect the organoids in the images. These approaches include edge detection using Canny filter, thresholding combined with mathematical morphology tools, texture analysis through Markov Random Fields, marked point processes. We have also modified several times the imaging protocol in order to simplify the object detection. To evaluate the growth of each organoid we have to match the objects detected in two consecutive frames. We have developed a matching algorithm based on a majority voting. We compute the vector between every pair of detected objects on both frames. Assuming that there is only a translation between the two frames we estimate it as the most represented vector. With this framework we have shown that the studied treatment has stopped the organoid growth (see figure 7).

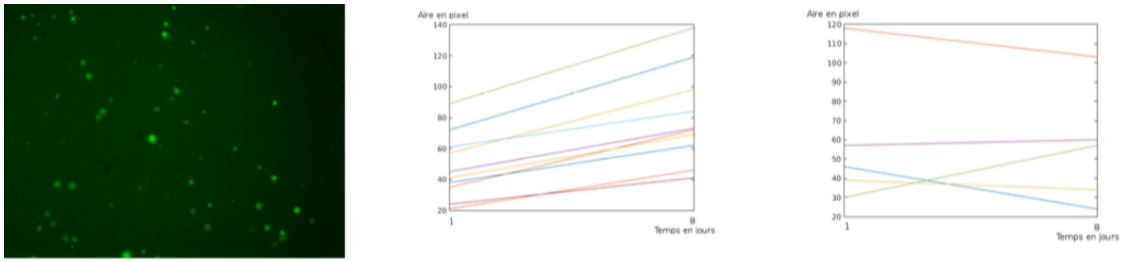


Figure 7. Image of organoids (left), Temporal evolution of some organoid size without treatment (middle) and with treatment (right).

6.8. Vesicles tracking

Participants: Raphael Pages, Xavier Descombes.

This work is a collaboration with P. Juan, M. Furthauer from iBV. It was partially supported by the UCA Jedi Idex.

We take advantage of the optical transparency of the zebrafish to study the formation, transport and function of extracellular vesicles in vivo. In the zebrafish Left/Right Organizer (LRO) a cilia-driven fluid flow promotes the directional transport of extracellular vesicles in the organ lumen. We have developed a software to analyze the vesicles trajectory. Assuming that the speed is slow enough, the detection is performed by considering the 2D+t time sequence of data as a 3D volume in which the vesicle trajectories are represented by tubular shapes. We first remove the background in each slice by subtracting a temporal mean computed on a sliding window. The trajectories are than enhanced by a Frangi filter followed by a top hat operator. Finally, the trajectories are obtained by a threshold and filtered to remove those corresponding to cilia movement. To compare different populations we then compute a mean shape of the LRO using an elastic shape metric. The trajectories detected on the different samples of a given population are then projected onto this common space. To have a dense representation of the vesicles speed inside the LRO we then extrapolate the detected trajectories on the whole population using a Markov random field regularization (see figure 8).

6.9. Comparison of tracking strategies

Participants: Sarah Laroui, Grégoire Malandain, Gaël Michelin.

This work takes place within the ANR PhaseQuant.

In video-microscopy, subject-based studies require the tracking of every individual to both quantify its dynamics (speed, etc) and detect special events (mitosis). In high throughput experiments, manual annotation or correction of sequences is not feasible, and computed-based strategies are definitevely prefered. In such a context, where cells have already been segmented in video-microscopy images (by a third party method), this work aims to assess different tracking strategies in presence of unavoidable segmentation errors (missing cells, over- or under-segmentations).

Two main strategies have been under examination. In the first one, all pairing hypothesis (based on a proximity criteria) have been generated. Further stages of both selection of plausible pairings or rejection of non-plausible ones have been tested to end up with tracking results. In the second one, pairings are built progressively based on their plausibility (one cell can be paired forward to 0, 1 or 2 cells; one cell can be paired backward to 0 or 1 cell). In both strategies, jumps are allowed to take into account possible segmentation errors.

It appears that the first strategie is more likely to end up with undecidable unplausible situations, that can not occur, by construction, with the second one.

6.10. 3D Coronary vessel tracking in x-ray projections

Participants: Emmanuelle Poulain, Grégoire Malandain.

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

Percutaneous Coronary Intervention (PCI) is a minimally procedure which is used to treat coronary artery narrowing. The physician intervenes on the patient under the guidance of an x-ray imaging system. This system is not able to display a visual assessment of the coronary wall, contrary to the pre-operative Computed Tomography Angiography (CTA). To help physician to exploit this information during the course of the procedure, registering these two modalities would be useful. To this aim, we first proposed in a previous work a method of 3D coronary tracking of the main vessel in x-ray projections [17]. Although, we faced a segmentation problem when we wanted to move from the tracking of one vessel to the entire set. For this reason, we have worked this year on the vessel centerline extraction in x-ray projection images.

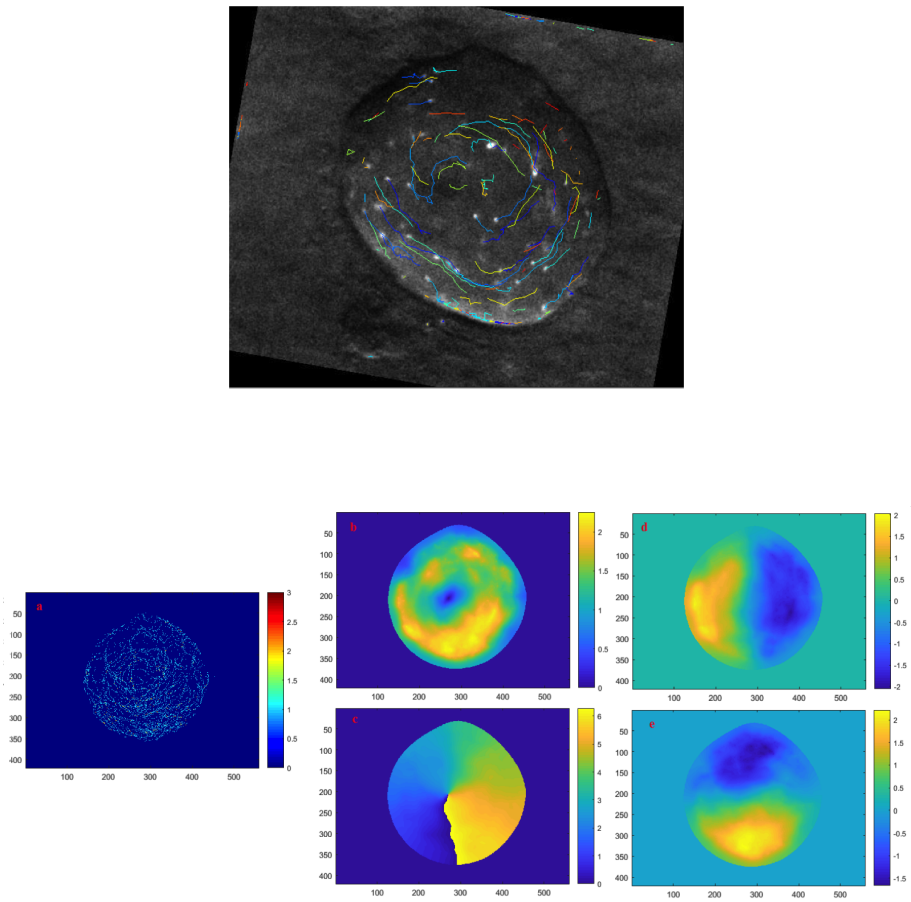


Figure 8. Trajectories detected on one LRO (top). Bottom: trajectories detected the whole population (right a), horizonal (b), vertical (c) radial (d) and angular (e) speeds.

2D Angiographic images are often first enhanced, before centerline extraction, by dedicated filters, e.g. Hessian based filters. Such filters exhibit critical defects, one of them being the non-uniform response for vessels of different sizes. This fact largely compromised the next step of centerline extraction. This last step requires a threshold step which is usually not clearly explained in other established methods. We worked on a model-based study of two widely used Hessian-based filter. It demonstrates that the non-uniform response for vessels of different sizes is due to the projective effect, and further enables to propose an X-ray projection dedicated method for centerline extraction which overpass this behavior. It is complemented by a component-based hysteresis thresholding. Last, the huge variability of coronary image aspect, due to imaging parameters, makes the threshold choice quite complex. We have shown that the thresholds can be determined automatically taking into account the kilovoltage peak (kVp), one key technical parameter of the X-ray acquisition. These thresholds are determined without any a priori on the image content. This technique not only allows to obtain an almost optimal segmentation, but also performs well for non-injected frames. Results of our proposed method and methods from state of the art are presented in Fig. 9 .

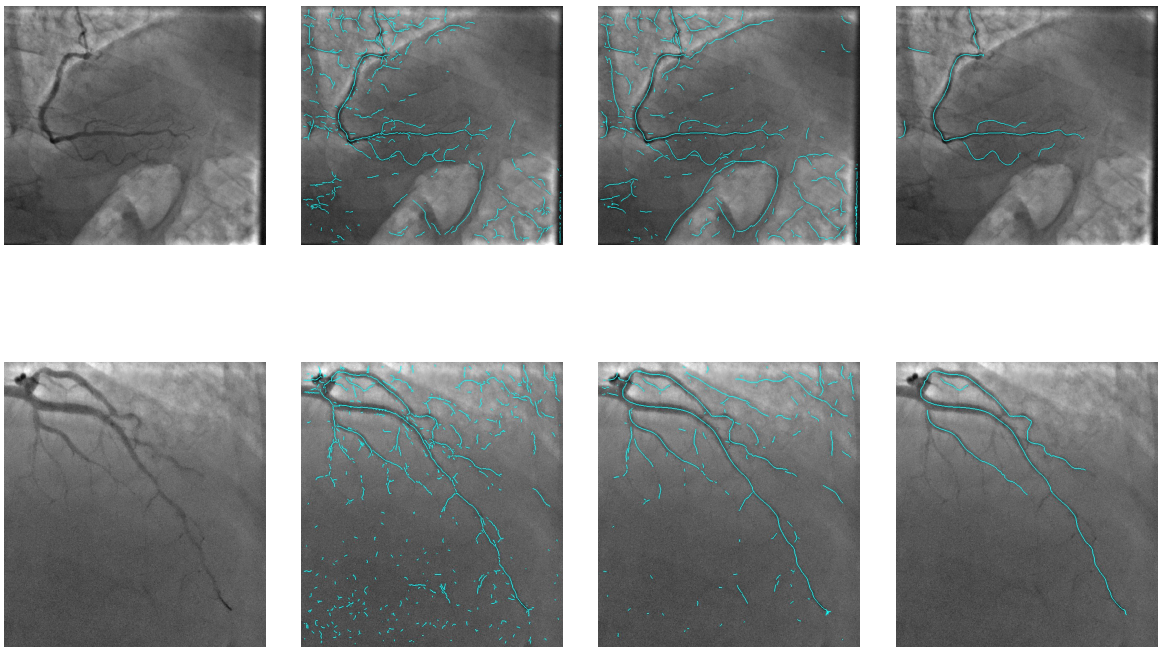


Figure 9. The obtained results of the different methods on a right coronary sample (first line), and left coronary sample (second line). From left to right: the original image, the result of Frangi (OPHT), the result of Krissian (OPHT), and the result of the proposed method (CCHT). The two first methods were tuned to obtain the same sensitivity than the third one.

6.11. Mitochondrial network detection and classification

Participants: Guillaume Lavisse, Xavier Descombes.

This work is a collaboration with C. Badot and M. Chami from IPMC and A. Charezac, S. Clavel and F. Bost from C3M. It was partially supported by the UCA Jedi Idex.

Last year we had developed a framework to classify mitochondrial networks. In this framework, mitchondrial networks are first binarized using our algorithm ATOLS. Some geometrical features are then computed for each connected component providing a clustering at the object level in the feature space. A signature of a

given image is then defined by the ratio of objects in the different classes. A second classification, performed by an SVM on this signature, provides a global class for the image. The different classes are defined as fragmented, tubular and filamentous. This year, we have validated this framework on two other databases, one consisting of cultured cells, the other being constituted of Alzheimer neuronal cells. The results were not satisfactory compared to those obtained on the first database last year. This is mainly due to the signal heterogeneity within an image. To compensate this heterogeneity we have applied a local normalization (see figure 10). We then have recovered classification performances comparable to those obtained by an expert. The next step consists in following in time the mitochondria of Alzheimer neuron. To this aim, we have developed a matching algorithm between two sets of mitochondria based on geometrical features and location of detected objects at two different instants.

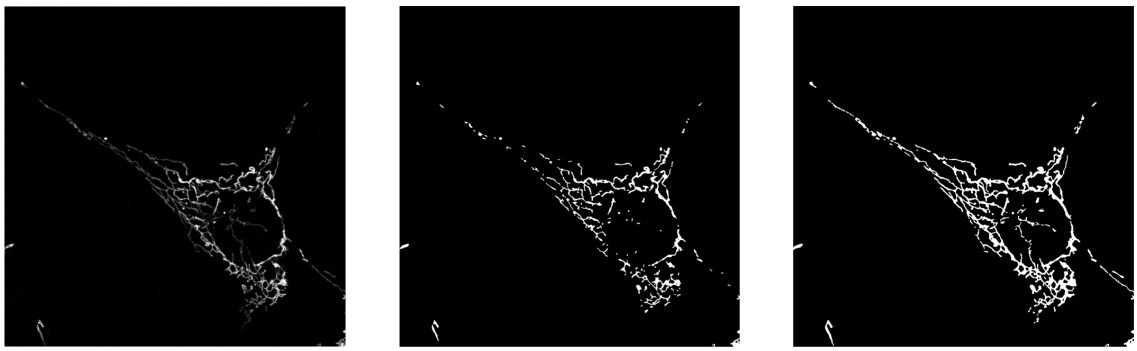


Figure 10. Image of Mitochondria (left), Binarization obtained without local normalization (middle) and with local normalization (right).

6.12. Botrytis cinerea phenotype recognition and classification: toward the establishment of links between phenotypes and antifungal molecules

Participants: Sarah Laroui, Eric Debreuve, Xavier Descombes.

This work is a collaboration with Aurelia Vernay (Bayer, Lyon, France).

Botrytis cinerea is a reference model of filamentous phytopathogen fungi. Some chemical treatments can lead to characteristic morphological changes, or phenotypic signatures, observable with transmitted light microscopy. These phenotypes could be associated with the treatment Mode of Action (figure 11). The goal of this work is the recognition of already known phenotypes but also the detection of new phenotypes. Because of the different dose-response effects, each given molecule is tested at ten concentrations.

In this context, we are developing a robust image analysis and classification framework relying on morphometric and topological characteristics to automatically recognize such phenotypes. Specifically, these characteristics are used in a supervised machine-learning framework to learn a Random Forest classifier.

After object detection, we calculate the skeleton of each object and we converted them into graphs, a more convenient data structure. Two types of parameters were extracted: those calculated globally on all the objects of an image like for example the number of objects and the skeleton length variance, and those computed on each object of an image like the number of nodes, the mean branch length and the object area.

6.13. Automatic zooplankton classification using hierarchical approaches

Participants: Eric Debreuve, Baptiste Pouthier.

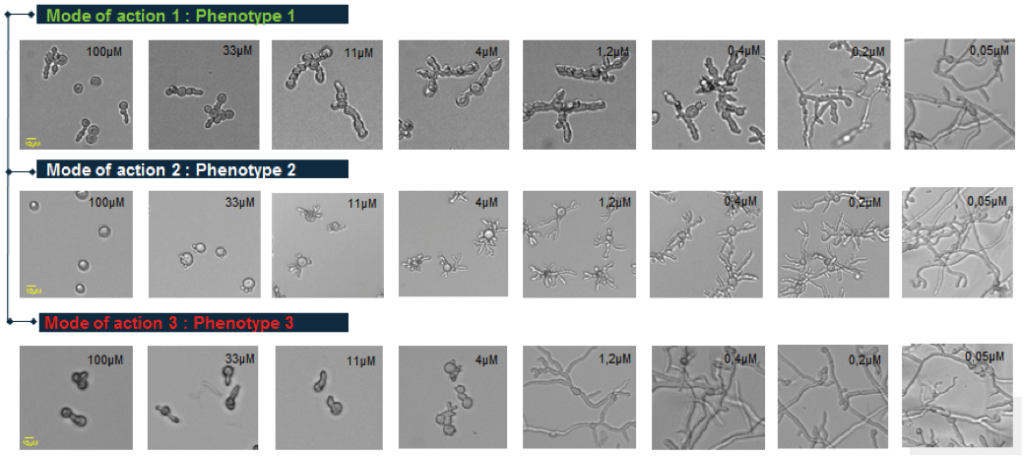


Figure 11. Each row depicts the observed phenotypic signatures associated with a given molecules. Columns correspond to different molecule concentrations.

This work is made in collaboration with Frédéric Precioso (I3S) and Jean-Olivier Irisson (Laboratoire d'Océanographie de Villefranche-sur-mer).

In marine ecology routine, zooplankton organisms are imaged using a single camera system. With the purpose of building an automatic classifier of plankton images, databases of annotated images are built. For each species, such databases contain a set of images of similar organisms, but seen under different point of views, i.e., having potentially very different appearances. Hence, learning an automatic classifier for zooplankton from such databases is difficult. In consequence, feeding the learning process with as much information as possible is essential. One piece of information we have access to is a taxonomic structure (hence hierarchical structure) of zooplankton species established by environmental biologists (see Fig. 12). Therefore, we have explored (and we continue to explore) different strategies to use such a hierarchy in the learning process, from the straightforward one consisting of learning a coarse-to-fine tree of independent convolutional neural networks (CNNs) to using neural network architectures explicitly accounting for hierarchical constraints.

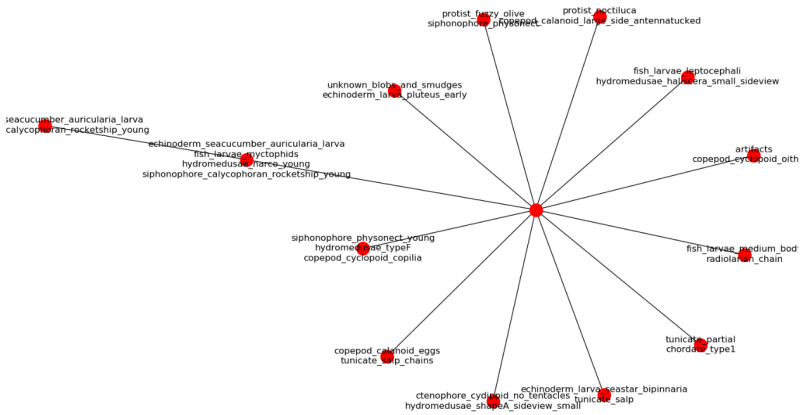


Figure 12. An Illustration of the zooplankton taxonomy.

MOSAIC Team

6. New Results

6.1. Dynamical characterization of morphogenesis at cellular scale

Participants: Guillaume Cerutti, Emmanuel Faure [External Collaborator], Christophe Godin, Bruno Leggio, Jonathan Legrand, Patrick Lemaire [External Collaborator], Grégoire Malandain [External Collaborator], Jan Traas [External Collaborator].

- Research Axes: **RA1** (*Representation of biological organisms and their forms in silico*) & **RA3** (*Plasticity & robustness of forms*)
- Key Modeling Challenges: **KMC3** (*Realistic integrated digital models*)

The modeling of morphogenesis requires to explore the interconnection of different spatial and temporal scales of developing organisms. Non-trivial questions such as whether the observed robustness of morphogenesis is rooted in some highly conserved properties at the cellular level or whether it emerges as a macroscopic phenomenon, necessitate precise, quantitative analyses of complex 3D dynamic structures. The study of dynamical properties at the cellular scale poses at the same time key technical challenges and fundamental theoretical questions. An example of the former category is how to characterize and follow the change of shape of cells within tissues and of tissues within organs, and how to couple this change with, for instance, gene expression dynamics; an illustration of the latter is how to define cell-scale variability of morphogenesis within and between species. Our team has produced this year several results in this context:

Cells spatio-temporal properties and patterns characterization. Over the past few years, we have achieved quantitative characterization of some of the cells physical properties, such as volumes or curvatures, in a developing tissue. Together with cell lineaging, it also enabled the quantification of temporal properties at cellular scale such as volumetric growth rate or strain patterns. To ease-up the analysis and to structure the previously described data, we have implemented a dedicated spatio-temporal graph structure, formalizing the cell network and its change in time.

To further characterize the tissue development, we developed clustering methods to identify cellular patterns based on a selection of quantified cell properties, including topology. Since such data are highly structured, both in time and space, we developed two complementary approaches:

1. spatial oriented: this approach use the cell neighborhood and a selection of cell descriptors to create pairwise distance maps latter clustered by a distance-based method, such as Ward's hierarchical clustering.
2. temporal oriented: this approach uses the lineage forest and a selection of cell descriptors to infer cell identities using Hidden Markov Tree (HMT) models.

Both approaches allow later characterization of the detected cluster or groups of cells based on their properties and should be published during the first half of 2019.

Atlases. One fundamental requirement to understand morphogenesis is the creation of atlases of different properties and different species. This year we have started creating two morphogenetic atlases: the atlas of gene expression patterns in the *Arabidopsis thaliana* flower development and the atlas of early embryonic development of the ascidian *Phallusia mammillata*.

Phallusia mammillata embryos develop with an invariant cellular lineage and with a relatively low number of cells (~ 700) up until the end of neurulation. This allows the creation of atlases with cellular resolution. Developing embryos from in-vitro fertilised dechorionated eggs have been injected with mRNA to fluorescently label their cell membranes and imaged by light-sheet microscopy for several hours of development. Automated image reconstruction through the segmentation pipeline ASTEC allowed to collect a large number of wild-type and mutated development with single-cell resolution and with a temporal resolution of two minutes. Based on this amount of data and on the invariant early ascidian lineage, we started curating an atlas of wild-type cellular, tissue and embryonic properties. Each cell, classified by its unique name, is identified in each wild-type embryo and analyzed through the dedicated computational pipelines. The result of this work provides a comprehensive view on the variability (in time, within and between embryos) of properties such as cell volume, cell surface, cell and tissue shape, cell topology, length of cell cycle, cell position within its tissue and globally within the embryo, orientation of cell's cleavage plane. This cellular networks have been coupled via cell names with genetic data coming from the the ascidian genetic database (ANISEED) and a specific tool, Morphonet, has been developed to explore these morphodynamic atlases seamlessly within a web-browser (paper in revision).

On the other hand, developing digital atlases of organism or organs development is a complex challenge for organisms presenting a strong variability in the cellular layout. Indeed contrary to *C. Elegans* or *P. mammillata*, for instance, that posses a very strict cell lineage, the development of most organisms or organs is under the influence of robust genetic patterns but without a unique cellular layout. In that respect, proposing a cell-based atlas of flower development for instance is not straightforward and specific methods have been developed to choose a representative examples of the developing *A. thaliana* flower. Using this representative flower we have generated an atlas in which we have introduced manually the expression patterns of 27 genes. The knowledge generated by the creation of this atlas makes it possible to have a first quantitative (correlative) view on the relation between gene activity and growth.

Both these works should result in publications in 2019.

Robustness of ascidian embryonic development. The image segmentation pipeline ASTEC developed by the team allows the 3D dynamic reconstruction of early ascidian embryogenesis at cellular resolution. Based on the high-quality wild-type data of our ascidian morphogenetic atlas and on ANISEED, we investigated the robustness of ascidian embryonic development and established a model to explain its origin. Thanks to the image-analysis pipelines we developed, we could extract relevant information from data and to perform cell-to-cell comparisons between different embryos of the same species (Phallusia mammillata). Since embryos developing from dechorionated eggs are left-right symmetric, we assessed the degree of cell-level variability between two embryos with different genomes (genetic variability) by comparing it to the intrinsic left-right variability in cellular properties within each embryo (stochastic variability). We showed that the same degree of variability is observed within and between embryos, demonstrating how ascidian embryonic development is highly canalised, and that the high reproducibility of shapes observed during embryogenesis is rooted in the robustness of cellular geometry and topology. Based on these observations, we studied the dynamics of embryonic patterning by developing a quantitative mathematical model for cellular fate-restriction events based on kinetic equations describing biochemical signalling. This model suggests that the robustness of cell topology and geometry is necessary for cell-cell biochemical interactions to give rise to the correct fate restriction events, a phenomenon which might represent a strong evolutive constraint to cell-scale variability in ascidians.

These results gave rise to a work which is currently under review and published as a preprint [16].

Digital reconstruction of developing *Arabidopsis* ovule. The ovule is a relatively simple organ, with limited developmental variability, which makes it an excellent case study for the computational modeling of organ development. In order to test various hypotheses of cellular growth, we reconstructed a first 4D digital tissue structure of a developing ovule as a triangulated cellular complex. It can be used as an input for FEM-based simulations, and will allow to compare quantitatively the results of growth models with actual ovule development.

This work was part of the *Imago* project.

6.2. Reconstruction of macroscopic forms from images and characterization of their variability

Participants: Guillaume Cerutti, Christophe Godin, Jonathan Legrand, Katia Mirande.

- Research Axes: **RA1** (*Representations of forms in silico*) & **RA3** (*Plasticity & robustness of forms*)
- Key Modeling Challenges: **KMC3** (*Realistic integrated digital models*)

To study the variability of macroscopic forms resulting from development, it is necessary to both develop digital reconstruction methods, typically based on image acquisitions, and statistical tools to define notions of distance or average between these forms. The automatic inference of computational representations of forms or organ traits from images of different types is therefore an essential step, for which the use of prior knowledge can be very beneficial. Realistic synthetic models of forms can guide the reconstruction algorithms and/or assess their performances. Computational representations of forms can then be used to analyze how forms vary at the scale of a population, of a species or between species, with potential applications in species identification and genetic or environmental robustness estimation.

Automatized characterization of 3D plant architecture. The digital reconstruction of branching and organ forms and the quantification of phenotypic traits (lengths of internodes, angles between organs, leaf shapes) is of great interest for the analysis of plant morphology at population scale. We develop an automated processing pipeline that involves the 3D reconstruction of plant architecture from RGB image acquisitions performed by a robot, and the segmentation of the reconstructed plant into organs. To provide validation data for the pipeline, we designed a generative model of *Arabidopsis thaliana* simulating the development of the plant architecture at organ scale. This model was used to develop the method for the measurement of angles of organs and test its accuracy. In a second phase, the model will be used to generate training data for machine learning techniques introduced in the reconstruction methods.

This work is part of the *ROMI* project.

Identification of plant species from morphological traits. The description of morphological traits of the various organs of the plants (leaves, bark, flowers and fruits) is essential for the characterization of a phenotype, and is highly relevant in the context of species or variety identification. In the context of tree species identification from RGB images of their organs, we study methods to represent the morphological charateristics of the plant organs, and the way to combine those different sources of information to enhance the classification performance. We demonstrated that botany-inspired descriptors of bark improves tree species classification based on leaves [13]. We also explore the possibility of using deep learning techniques to train a system to extract botanically relevant information from images [3].

This work is part of the *ReVeRIES* ANR project, in which the team is not directly involved.

This work has led to a publication in *Ecological Informatics* and to a participation at the *International Worskhop on Image Analysis Methods for the Plant Sciences* in Nottingham in January 2018.

6.3. Analysis of tree data

Participants: Romain Azaïs, Christophe Godin, Florian Ingels, Clément Legrand.

- Related Research Axes: **RW1** (*Representations of forms in silico*)
- Related Key Modeling Challenges: **KMC1** (*A new paradigm for modeling tree structures in biology*)

Tree-structured data naturally appear at different scales and in various fields of biology where plants and blood vessels may be described by trees. In the team, we aim to investigate *a new paradigm for modeling tree structures in biology* in particular to solve complex problems related to the *representation of biological organisms and their forms in silico*.

In 2018, we investigated the following questions linked to the analysis of tree data. (i) How to control the complexity of the algorithms used to solve queries on tree structures? For example, computing the edit distance matrix of a dataset of large trees is numerically expensive. (ii) How to estimate the parameters within a stochastic model of trees? And finally, (iii) how to develop statistical learning algorithms adapted to tree data? In general, trees do not admit a Euclidean representation, while most of classification algorithms are only adapted to Euclidean data. Consequently, we need to study methods that are specific to tree data.

Approximation of trees by self-nested trees. Complex queries on tree structures (*e.g.*, computation of edit distance, finding common substructures, compression) are required to handle tree objects. A critical question is to control the complexity of the algorithms implemented to solve these queries. One way to address this issue is to approximate the original trees by simplified structures that achieve good algorithmic properties. One can expect good algorithmic properties from structures that present a high level of redundancy in their substructures. Indeed, one can take account these repetitions to avoid redundant computations on the whole structure. In the team, we think that the class of self-nested trees, that are the most compressed trees by DAG compression scheme, is a good candidate to be such an approximation class.

In [7], we have proved the algorithmic efficiency of self-nested trees through different questions (compression, evaluation of recursive functions, evaluation of edit distance) and studied their combinatorics. In particular, we have established that self-nested trees are roughly exponentially less frequent than general trees. This combinatorics can be an asset in exhaustive search problems. Nevertheless, this result also says that one can not always take advantage of the remarkable algorithmic properties of self-nested trees when working with general trees. Consequently, our aim is to investigate how general trees can be approximated by simplified trees in the class of self-nested trees from both theoretical and numerical perspectives.

We conjecture that the problem of optimal approximation by a self-nested tree is NP-hard. Despite a substantial work in 2018 (internship of Clément Legrand), this remains an open question. Consequently, we have developed a suboptimal approximation algorithm based on the *height profile* of a tree that can be used to very rapidly predict the edit distance between two trees, which is a usual but costly operation for comparing tree data in computational biology [7]. Another algorithm based on the simulation of Gibbs measures on the space of trees is currently under development. This work should result in a publication next year.

Statistical inference. The main objective of statistical inference is to retrieve the unknown parameters of a stochastic model from observations. A Galton-Watson tree is the genealogical tree of a population starting from one initial ancestor in which each individual gives birth to a random number of children according to the same probability distribution, independently of each other. In a recent work [12], we have focused on Galton-Watson trees conditional on their number of nodes. Several main classes of random trees can be seen as conditioned Galton-Watson trees. For instance, an ordered tree picked uniformly at random in the set of all ordered trees of a given size is a conditioned Galton-Watson tree with offspring distribution the geometric law with parameter $1/2$. Statistical methods were developed for conditioned Galton-Watson trees in [19]. We have introduced new estimators and stated their consistency. Our techniques improve the existing results both theoretically and numerically. A simulation study shows the good behavior of our procedure on finite-sample sizes and from missing or noisy data.

In a very different context, a substantial work has been made on statistical inference for piecewise-deterministic processes [2], [9], [8].

Kernel methods for tree data. In statistical learning, one aims to build a decision rule of a qualitative variable Y as a function of a feature X (typically a vector of $\mathbb{R}^d$) from a training dataset $(X_i, Y_i)_{1 \leq i \leq n}$. We assume that X is a tree, ordered or not, with or without labels. This framework is quite original since the state space of X is not endowed with a canonical inner product. Kernel methods are particularly adapted to this setting since they enable to transform the raw data into a Hilbert space. In this context, the main issue is related to the construction of a *good kernel*. A kernel function adapted to trees is the subtree kernel introduced [24]. While the literature has never been focused on the weight function involved in the subtree kernel, we have shown that this function is crucial in prediction problems. We have proposed a new algorithm for computing the subtree kernel. It has been designed to allow learning the weight function directly from the data. On some difficult datasets, the prediction error is dramatically decreased from $> 50\%$ to 3%.

This work is part of the *ROMI* project, that aims to develop an open and lightweight robotics platform for microfarms. This project requires to investigate advanced analysis and modeling techniques for plant structures. A main issue that arises in this context is to predict a feature of the plant (species, health status, etc) from its topology.

Invited talk on tree structures and algorithms Christophe Godin gave a invited talk entitled *Can we manipulate tree forms like numbers?* that was prepared with Romain Azaïs at the workshop on *Mathematics for Developmental Biology* organized at the Banff International Research Station for Mathematical Innovation and Discovery, organized by P. Prusinkiewicz and E. Mjolsness (Banff, Canada, December 2017).

Abstract: Tree-forms are ubiquitous in nature and recent observation technologies make it increasingly easy to capture their details, as well as the dynamics of their development, in 3 dimensions with unprecedented accuracy. These massive and complex structural data raise new conceptual and computational issues related to their analysis and to the quantification of their variability. Mathematical and computational techniques that usually successfully apply to traditional scalar or vectorial datasets fail to apply to such structural objects: How to define the average form of a set of tree-forms? How to compare and classify tree-forms? Can we solve efficiently optimization problems in tree-form spaces? How to approximate tree-forms? Can their intrinsic exponential computational curse be circumvented? In this talk, we presented a recent work to approach these questions from a new perspective, in which tree-forms show properties similar to that of numbers or real functions: they can be decomposed, approximated, averaged, and transformed in dual spaces where specific computations can be carried out more efficiently. We will discuss how these first results can be applied to the analysis and simulation of tree-forms in developmental biology (https://www.birs.ca/events/2017/5-day-workshops/17w5164).

6.4. Mechanics of tissue morphogenesis

Participants: Olivier Ali, Arezki Boudaoud [External Collaborator], Guillaume Cerutti, Ibrahim Cheddadi [External Collaborator], Christophe Godin, Bruno Leggio, Jonathan Legrand, Hadrien Oliveri, Jan Traas [External Collaborator].

- Research Axes: **RA2** (*Data-driven models*) & **RA3** (*Plasticity & robustness of forms*)
- Key Modeling Challenges: **KMC2** (*Efficient computational mechanical models of growing tissues*) & **KMC3** (*Realistic integrated digital models*)

As deformations supporting morphogenesis require the production of mechanical work within tissues, the ability to simulate accurately the mechanical behavior of growing living tissues is a critical issue of the MOSAIC project. From a macroscopic perspective, tissues mechanics can be formalized through the framework of continuum mechanics. However, the fact that they are composed, at the microscopic level, by active building blocks out of equilibrium (namely cells) offers genuine modeling challenges and opportunities. This section describes the team's efforts on integrating cellular behaviors such as mechano-sensitivity, intercellular fluxes of materials and cell division into a macroscopic mechanical picture of morphogenesis.

Mechanical influence of inner tissues. Mechanical stress patterns within plant tissues emerge from the balance between inner-pressure-induced forces and the elastic response of the cell wall [0] over the entire tissue. Being able to derive, from a specific cellular architecture, the corresponding pattern of stresses within a tissue is crucial for the study of morphogenesis. It requires a precise description of the tissue as a network of connected cells and the ability to run numerical simulations of force balance on such heterogeneous structures.

To that end, we developed numerical methods to generate finite element meshes from: i) 3D microscopic images with sub-cellular resolution (referred to as *bio-inspired* structures) and ii) 3D cellularized geometrical volumes (referred to as *artificial* structures). Combined with a FEM-based simulation framework previously developed within the team [20], we generated quantitative maps of stress distributions in multilayered reconstructed tissues. The combined analysis of stress patterns on *bio-inspired* and *artificial* structures showed how mechanical stresses experienced by cells convey geometrical information to cells about the global shape of the tissue as well as the local shape of cells.

[0] A thick protective exoskeleton surrounding plant cells

This work was part of the *Morphogenetics* IPL and Jan Traas ERC grant *Morphodynamics*.

This work is currently under review in the *Bulletin of Mathematical Biology* and has been presented at the *19th International Conference of System Biology* held in Lyon at fall.

Shape regulation. Reproducible and robust morphogenesis requires growth coordination of thousands of cells. How such coordination can be "implemented" in living organism is a core question for *RA3*. One identified mechanism in plant to coordinate growth rely on cells mechano-sensitivity [0]. Combined with the geometrical dependency of mechanical stress (*c.f.* previous subsection), this suggests the existence of a feedback mechanism that regulates tissue shape changes. We have been investigating closely the consequences of such a mechanism.

To that end, we first modeled the bio-molecular pathway relating mechanical stress experienced by cells to actual modification of their mechanical properties (*e.g.* cell wall stiffness). This work enabled us to describe plant tissues as an active material featuring large-scale properties, such as stress stiffening [0], emerging from sub-cellular dynamics. This work has been published in *Journal of Mathematical Biology* [5].

In parallel, we modeled the influence of cell wall elasticity (value, orientation) on the growth dynamics of tissues. This was done in the context of plant organogenesis, in close collaboration with biologists investigating the effect of cell-wall-related mutations on plant organ initiation. Our modeling approach was based on our previously developed *strain-based growth* model [20]. This joint study has been published in *Development* earlier this year [1].

We then studied how initial spherical symmetry (e.g. dome-shaped primordia) can be potentially broken during development in such active tissues and lead to elongated or flat shapes. For this, we integrated the *stress feedback* model with the *strain-based growth* model to investigate how their interplay could influence the morphogenesis of 3D cellularized structures. In particular, we showed that a stress-based feedback mechanism can maintain the typical plant growth modes (*i.e.* axial elongation or 2D flat expansion) and amplify asymmetries. This computational approach to symmetry breaking in growing tissues has been developed in parallel to experimental investigations addressing the shape evolution of sepals [0].

This work was part of the *Mophogenetics* IPL and Jan Traas ERC grant *Morphodynamics*.

The whole story has been presented at the *9th International Plant Biomechanics Conference* in Montreal this summer. A journal article combining both our modeling approach and experimental work in the context of symmetry breaking during plant organogenesis is currently being written.

Influence of water fluxes on plant morphogenesis. Since pressure appears as the "engine" behind growth-related deformation in plants, its regulation by cells is a major control mechanism of morphogenesis. We developed 2D computational models to investigate the morphological consequences of the interplay between cell expansion, water fluxes between cells and tissue mechanics. This interdisciplinary work, combining experiments and modeling, addresses the influence of turgor pressure heterogeneities on relative growth rate between cells. We showed that the coupling between fluxes and mechanics allows us to predict observed morphological heterogeneities without any *ad hoc* assumption. It also reveals the existence of a putative inhibitory action of organ growth on growth in immediatemy neighboring regions, due to the hydraulic coupling between cells during growth.

This work was part of the Agropolis fundation project *MecaFruit3D* and Arezki Boudaoud's ERC *PhyMorph.*

Two papers report the results of this work (one currently under review in *Nature Physics* [21] and a second one that is about to be submitted. These results have also been presented last summer at the *9th International Plant Biomechanics Conference* in Montreal.

Influence of dividing cells on tissue mechanics during morphogenesis in ascidians. The control of cell division orientation is of prime importance for patterning and shape emergence, especially in animal embryos where the first developmental stages happen at constant volume. In recent years, the Hetwig's rule appeared

[0] the ability to probe mechanical stress around them and to modify accordingly their growth behavior
[0] the ability of the tissue to re-enforce itself in the directions of high mechanical solicitations
[0] leaf-like organs surrounding and protecting flowers

as a physical model accounting for orientation of cell division. Within animal tissues it has been shown that the coupling of externally induced strain and Hertwig's rule leads to the orientation of cell divisions with the main stress direction.

We investigated through modeling the consequences, in a multicellular context, of such stress-based regulation of cell division orientation. To that end, we developed a theoretical standpoint on the many-body energetic thermodynamics of cell divisions in the presence of external anisotropic stress. We showed that Hertwig's rule emerges as a limiting-case behavior and how anisotropic mechanical stresses can provide important cues to guide cell divisions. Our model accounts for the division pattern observed in the epidermis of the embryo of ascidian *Phallusia mammillata*, including those reproducible observed deviations from Hertwig's rule which have so far eluded explanation.

This work was part of the *Digem* project.

This work has been presented in two national conferences: the *IBC Scientific Days* and the *Cell Cycle Days* both held in Montpellier. A paper is currently being written.

Automatic quantification of adhesion defects in microscopy images. Direct measurements of mechanical stresses experienced by living tissues are not yet feasible. To circumvent this limitation, we developed an indirect method based on measurements of cracks in tissues: Our biologist colleagues developed cell-adhesion mutants in which strong connections between epithelial cells are impaired. As a consequence, mechanical stresses within the tissue produce cracks. Distribution and orientation of these cracks can be related to the main directions of the mechanical forces at play. We developed a 2D image analysis pipeline to detect and quantify these cracks in microscopy projections of epithelia, and deduce the magnitude and orientation of tensions in organs and tissues. This tool has been used to evidence new mechanical signaling mechanisms in *Arabidopsis*.

This analysis pipeline has been published in [6] and used by collaborators in the analysis performed in [26].

6.5. Signaling and transport for tissue patterning

Participants: Romain Azaïs, Guillaume Cerutti, Christophe Godin, Bruno Leggio, Jonathan Legrand, Teva Vernoux [External Collaborator].

- Research Axes: **RA1** (*Representations of forms in silico*) & **RA2** (*Data-driven models*)
- Key Modelling Challenges: **KMC3** (*Realistic integrated digital models*)

One central mechanism in the shaping of biological forms is the definition of regions with different genetic identities or physiological properties through bio-chemical processes operating at cellular level. Such patterning of the tissue is often controlled by the action of molecular signals for which active or passive transport mechanisms determine patterning spatial precision. The shoot apical meristem (SAM) of flowering plants is a remarkable example of such finely controlled system where the dynamic interplay between the hormone auxin and the polarization of efflux carriers PIN1 during growth governs the rhythmic patterning of organs, and the consequent emergence of phyllotaxis. Using *Arabidopsis thaliana* as a model system, we developed an integrated view of the meristem as a self-organizing dynamical form by reconstructing the dynamics of physiological processes from living tissues, and by proposing computational models integrating transport and signaling to study tissue patterning *in silico*.

Automatic quantification of auxin transport polarities. Time-lapse imaging of living SAM tissues marked with various fluorescent proteins allows monitoring the dynamics of cell-level molecular processes. Using a co-visualization of functional fluorescent auxin transporter (PIN1-GFP) with a dye staining of cell walls with propidium iodide (PI), we developed a method to quantify in 3D the polarization of auxin transport for every anticlinal wall of the first layer of cells. The digitally reconstructed network evidenced an overall stable convergence of PIN1 polarities towards the center of the meristem, with local front lines matching dynamic accumulations of auxin [15]. It also showed that the apparent crescent shape often thought to indicate polarities in cells might sometimes be misleading, and opens the way for a new view of how auxin transport is regulated.

Temporal auxin signaling in meristem organ patterning. Morphogenetic signals such as auxin define spatial distributions that are thought to control tissue patterning, but it has been proposed in animals that they also carry temporal information in their dynamics. A recent model developed by our group has postulated the existence of a stochastic mechanism to explain disturbed phyllotaxis patterns. This model assumes that organ initiation results from a temporal integration of a morphogenetic signal that buffers molecular noise [22]. Using a quantitative analysis of the dynamics of auxin distribution and response, we provide evidence that organ initiation in the SAM is indeed dependent on the temporal integration of the auxin signal [15]. The duration of cell exposition to auxin is used to differentiate temporally sites of organ initiation, and provide robustness to the rhythmic organ patterning.

Computational models of integrated transport and signaling. To interpret these new observations of auxin signaling and transport in the meristem, we investigate theoretical and computational models to study dynamic auxin distributions and the consequent organ patterning at the level of the meristem. Building on existing models of auxin transport [23], [25], we investigate different competing hypotheses on the auxin-PIN interplay, through numerical simulations based on rate equations for molecular transport and efflux carrier polarization. Quantitative comparisons with in vivo observations will provide cues on how the system responses are linked to memory effects and information exchanges between auxin and PINs.

These works were part of the *BioSensors* HFSP project and are carried out in the *Phyllo* ENS-Lyon project and gave rise to a journal article submitted for publication. These results have been presented at the *International Worskhop on Image Analysis Methods for the Plant Sciences* in Nottingham in January 2018 and in several invited talks given by Teva Vernoux and Christophe Godin.

6.6. Regulation of branching mechanisms in plants

Participants: Romain Azaïs, Frédéric Boudon [External Collaborator], Christophe Godin.

- Research Axes: **RA2** (*Data-driven models*) & **RA3** (*Plasticity & robustness of forms*)
- Key Modelling Challenges: **KMC3** (*Realistic integrated digital models*)

Branching in plants results from the development of apical meristems that recursively produce lateral meristems. These meristems may be more or less differentiated with respect to the apical meristem from which they originate, potentially leading to different types of lateral branches or organs. They also can undergo a more or less long period of inactivation, due to systemic regulation. The understanding of branching systems morphogenesis in plants thus relies on the analysis of the regulatory mechanisms that control both meristem differentiation and inactivation.

Analysis of the diversity of inflorescence architecture in different rice species. Rice is a major cereal for world food security and understanding the genetic and environmental determinants of its branching habits is a timely scientific challenge. The domestication, i.e., the empirical selection by humans, of rice began 10 000 years ago in Asia and 3 000 years ago in Africa. It thus provides a short-term model of the processes of evolution of plants.

Hélène Adam and Stéphane Jouannic from the group Evo-Devo de l'Inflorescence of UMR DIADE at IRD (Montpellier) have collected for years on the different continents an outstanding database of panicle-type inflorescence phenotypes in Asian and African, cultivated and wild, rice species. Classical statistical analysis based on the extraction of characteristic traits for each individual branching system were able to separate wild species from cultivated ones, but could not discriminate between wild species, suggesting that the entire branching structure should be used for classification methods to operate. For this, we are currently developing statistical methods on tree structures (see section 6.3) that should allow us to achieve better discrimination between panicles, based on their branching topology in addition to geometric traits. By coupling the quantitative study of the panicles to genomic analyses carried out by the IRD group, we should be able to highlight which regulation pathways have been selected or altered during the domestication process.

The role of sugars in apical dominance. The outgrowth of axillary buds is a key process in plant branching and which is often shown to be suppressed by the presence of auxin in nodal stems. However, local auxin levels are not always sufficient to explain bud outgrowth inhibition. Recent studies have also identified a contribution of sugar deprivation to this phenomenon. Whether sugars act independently of auxin or other hormones auxin regulates is unknown. Auxin has been shown to induce a decrease of cytokinin levels and to upregulate strigolactone biosynthesis in nodes. Based on rose and pea experiments, both in vitro and in planta, with our collaborators Jessica Bertheloot, Soulaiman Sakr from Institut de Recherche en Horticulture et Semences (IRHS) in Angers, we have shown that sucrose and auxin act antagonistically, dose-dependently, and non-linearly to modulate bud outgrowth. The Angers group provided experimental evidence that sucrose represses bud response to strigolactones but does not markedly affect the action of auxin on cytokinin levels. Using a modeling approach, we tested the ability of this complex regulatory network to explain the observed phenotypes. The computational model can account for various combinations of sucrose and hormones on bud outgrowth in a quantitative manner and makes it possible to express bud outgrowth delay as a simple function of auxin and sucrose levels in the stem. These results provide a simple auxin-sucrose-cytokinin-strigolactone network that accounts for plant adaptation to growing conditions. A paper relating this work is currently under review.

The fractal nature of plants. Inflorescence branching systems are complex and diverse. They result from the interaction between meristem growth and gene regulatory networks that control the flowering transition during morphogenesis. To study these systems, we focused on cauliflower mutants, in which the meristem repeatedly fails in making a complete transition to the flower and for which a complete mechanistic explanation is still lacking.

In collaboration with Eugenio Azpeitia and François Parcy's group in Grenoble, we have developed a first model of the control of floral initiation by genes, refining previous networks from the literature so that they can integrate our hypotheses about the emergence of cauliflower phenotypes. The complete network was validated by multiple analyses, including sensitivity analyses, stable state analysis, mutant analysis, among others. It was then coupled with an architectural model of plant development using L-systems. The coupled model was used to study how changes in gene dynamics and expression could impact in different ways the architectural properties of plants. The model was then used to study how changes in certain parameters could generate different curd morphologies, including the normal and the fractal-like Romanesco. A paper reporting this work is currently being written.

PLEIADE Team

7. New Results

7.1. Introgressions as a source of diversity

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

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

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

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

7.2. Geometric view of biodiversity and molecular taxonomy

For the geometric view on biodiversity and molecular based taxonomy (metabarcoding), 2018 has been characterized by significant progress in the connections between the questions posed by metabarcoding, and the implementaton of solutions by intensive computing in collaboration with HiePACS team. We are now able to run a Multidimensional Scaling (MDS) to build a point cloud where each point is a read (a short sequence from an environmental sample) on a sample of nearly 2×10^5 reads on one node of a cluster with shared memory. PLEIADE is member of two projects lead by Inria BSO to extend these possibilities to work with distributed memory (connection with the Chameleon dense linear algebra solver) with the possibility to reach results with one million sequences at hand. To our knowledge, such a connection between scientific questions related to biodiversity and HPC-HTC has up to now had no equivalent and is new.

In collaboration with HiePACS, the SED of Inria BSO, and the GRICAD (Mésocentre of Grenoble), PLEIADE has developed a scientific library of tools for dimension reductions in the framework of Big Data, specifically linear algebra of full matrices. A prototype was developed in numpy, then reimplemented in C++ with an external library for Random Projection, and made available in Julia. This library will be connected with the workflow developed in Regional project and ADT Gordon.

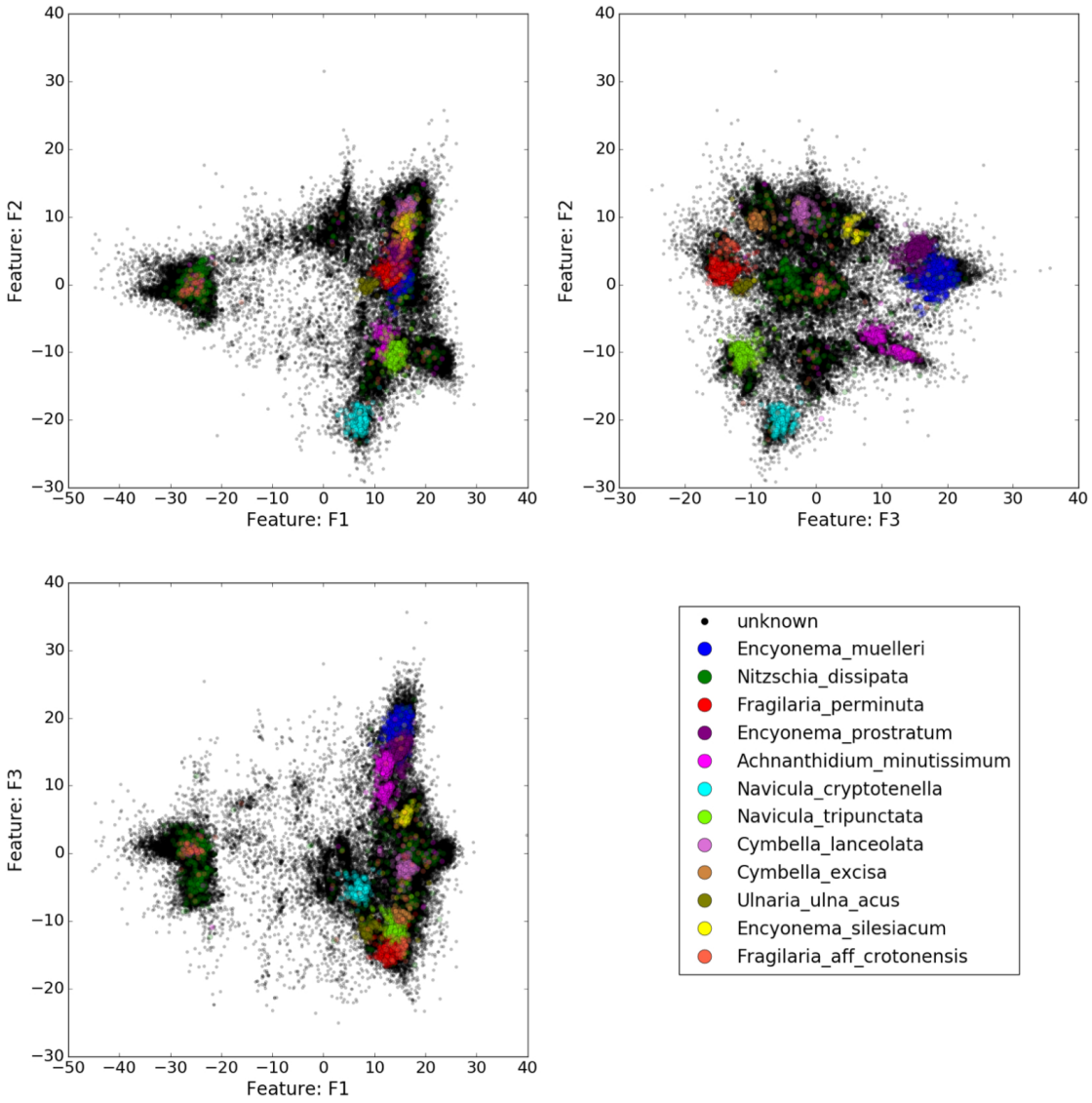


Figure 4. Validation of high density islands using supervised classification. Metagenomic reads from diatoms in Lake Geneva [26] were analyzed by the method from [7] and colored by species according to a reference database.

7.3. Tangible exploration of protein families

Protein families[3] are an effective way to compare the complete genomes of fungal species. In general these comparisons are very challanging due to the large evolutionary distances involved, the wide range of GC compositions observed from one species to the next, and the extensive map reshuffling that is characteristic of the yeasts in particular. Protein families are a classification of protein-coding gene sequences into phylogenetic groups, using clustering methods and semi-supervised classification. Members of a family are homologous and in many cases this homology is suggestive of functional similarity.

An intriguing feature of protein families is that the weighted graph constructed from their pairwise distance matrices has a structure that reflects the evolutionary history of the family. We developed software (family-3d, see above) that uses truncated distances to construct a weighted graph and to lay it out using an adaptation of the three-dimensional extension of the Kamada-Kawai force-directed layout. The resulting shapes for a set of protein families are then clustered manually by similarity (figure 5). Similarity in shape is highly suggestive of similarity in evolutionary scenarios, leading to hypotheses about the histories of individual protein families and the mechanisms by which functional diversity is obtained.

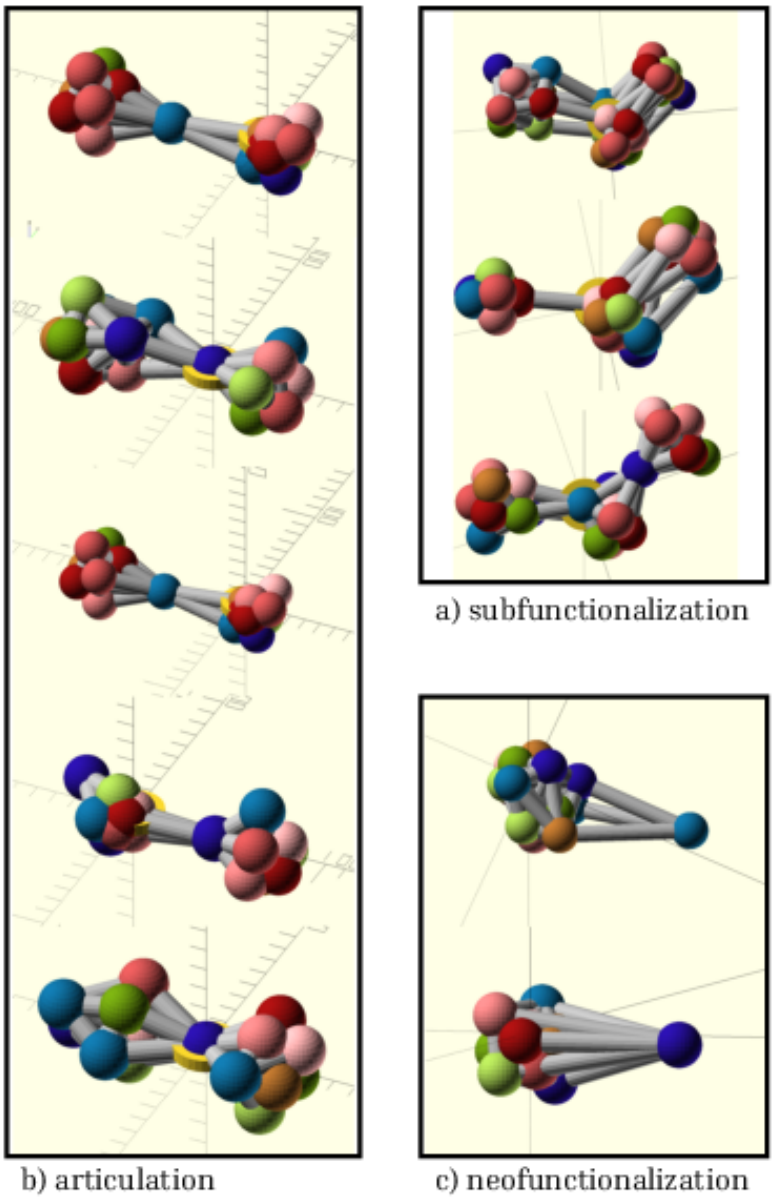


Figure 5. Three examples of classes of protein families with similar shapes: a) articulation, b) subfunctionalization, c) neofunctionalization

SERPICO Project-Team

7. New Results

7.1. A Monte-Carlo approach for missing wedge restoration in cryo-electron tomography

Participants: Emmanuel Moebel, Charles Kervrann.

We investigated a Monte-Carlo approach to restore spectral information in the missing wedge (MW) in cryo-electron tomography (CET). The MW is known to be responsible for several types of imaging artifacts, and arises because of limited angle tomography: it is observable in the Fourier domain and is depicted by a region where Fourier coefficient values are unknown (see Fig. 3). The proposed computational method tackles the restoration problem by filling up the MW by iterating the two 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. Also, specific constraint is added in the spectral domain by imposing the known Fourier coefficients to be unchanged through iterations. We justified this approach in the Monte-Carlo simulation and Bayesian framework. In practice, different denoising algorithms (BM3D, NL-Bayes, NL-means...) can be applied. In our experiments, several transforms have been tested in order to apply the constraint (Fourier transform, Cosine transform, pseudo-polar Fourier transform). Convincing results have been achieved (see Fig. 3) using the Fourier Shell Correlation (FSC) as an evaluation metric.

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

7.2. Algorithms for deconvolving and denoising fluorescence 2D-3D microsocopy images

Participants: Hoai-Nam Nguyen, Charles Kervrann.

In this work, we proposed a restoration method for 2D and 3D fluorescence imaging using a novel family of convex regularizers. The proposed regularization functionals are based on the concept of sparse variation, that consists in penalizing jointly the image intensity and gradient at each pixel to favor the co-localization of non-zero intensities and gradients, by considering eventually higher-order differentiation operators. By construction, these regularizers possess interesting mathematical properties, namely convexity, invariance to scale, rotation, and translation as the well-known total variation regularization approach. It enables therefore to design efficient algorithms to solve the underlying deconvolution problem, which is in general large-scale in the context of fluorescence microscopy. We reformulated denoising or deconvolution (given the point spread function) as a minimization problem of a convex energy function composed of a quadratic data fidelity term and a sparse-variation-based regularity term under the constraint of positivity and maximum intensity value. In order to minimize this energy, we considered a primal-dual (proximal) algorithm based on the full splitting technique, which only involves first-order operators to cope with the large-scale nature of the problem. Experimental results on both 2D and 3D synthetic and real fluorescence images demonstrated that our method was able to produce very competitive deconvolution results, when compared to several competing methods such as the Schatten norm of the Hessian matrix, in terms of quantitative performance as well as visual quality and computational time. The method is able to process a 512×512 image in 250 ms (in Matlab) with a non optimized implementation and can process 3D images in a few minutes (with no code optimization technique and multithreading).

Collaborators: Cyril Cauchois (Innopsys company).

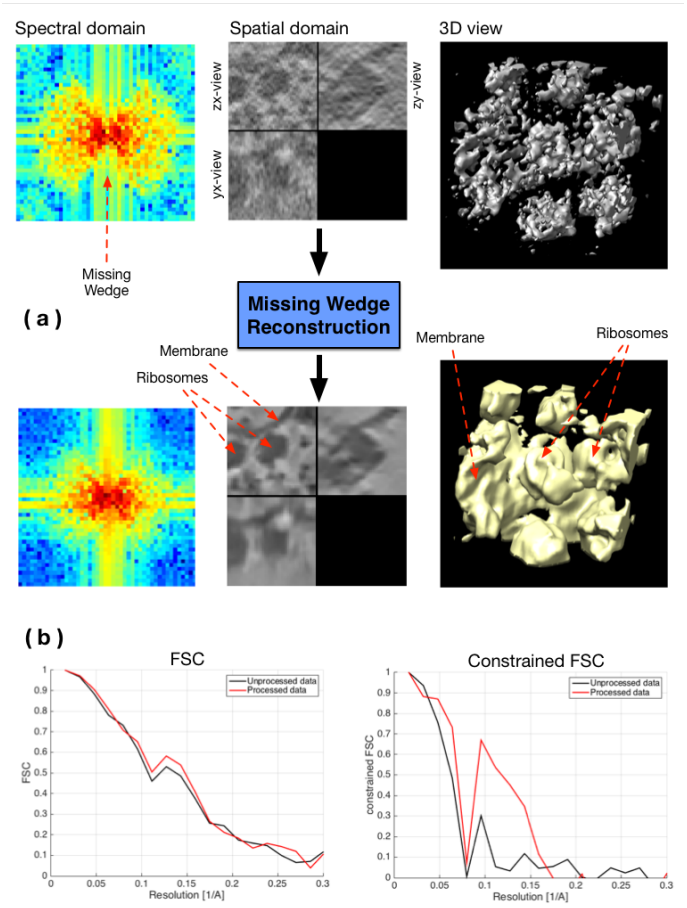


Figure 3. Experimental sub-tomogram containing ribosomes attached to a membrane. (a) Top row: original data in the spectral (left) and spatial (middle) domains and 3D view of the thresholded data (right). Bottom row: denoised data shown as above. (b) FSC and constrained FSC measures of the method input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

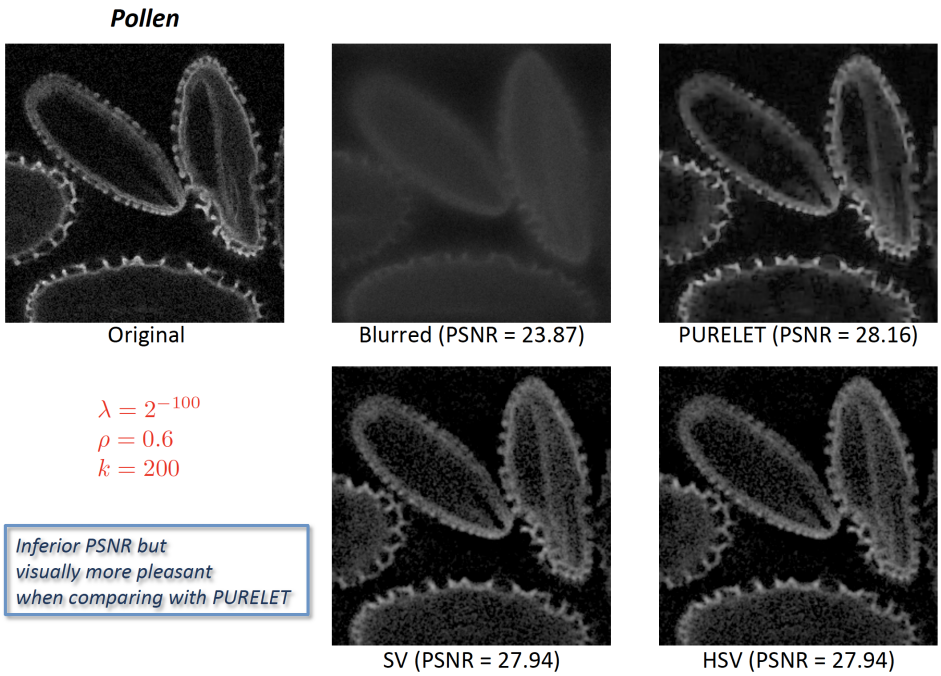


Figure 4. Comparison (2D view) of sparse deconvolution (SparsEvolution) method with Purelet method [42] on the artificial blurred and noisy (Gaussian-Poisson noise) 3D pollen image ($256 \times 256 \times 32$) *described in [42].*

7.3. Colocalization and co-orientation in fluorescence imaging

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

Collaborators: Frédéric Lavancier (University of Nantes, Laboratoire de Mathématiques Jean Leray), Thierry Pécot (Hollings Cancer Center at the Medical University of South Carolina), Jean Salamero and Liu Zengzhen (UMR 144 CNRS-Institut Curie).

7.4. Detection of transitions between diffusion models along biomolecule trajectories

Participants: Antoine Salomon, Charles Kervrann.

Recent advances in molecular biology and fluorescence microscopy imaging have made possible the inference of the dynamics of single molecules in living cells. When we observe a long trajectory (more than 100 points), it is possible that the particle switches mode of motion over time. Then, a goal is to estimate the temporal change-points, that is the distances at which a change of dynamics occurs. To address this issue, we proposed a non-parametric procedure based on test statistics [16], computed on local windows along the trajectory, to detect the change-points. This algorithm controls the number of false change-point detections in the case where the trajectory is fully Brownian. Our algorithm is user-friendly as there is only one parameter to tune, namely the sliding window size. A Monte Carlo study is proposed to demonstrate the performances of the method and also to compare the procedure to two competitive algorithms. Our method is much faster than previous methods which is an advantage when dealing with a large numbers of trajectories. With this computational approach, we analyzed real data depicting neuronal mRNPs (mRNAs in complex with mRNA-binding), and another very complex biological example, Gal-3 trafficking from the plasma membrane to different cellular compartments (acquired with Lattice Light Sheet microscopy). The analysis of multiple Gal-3 trajectories demonstrates nicely that there is not one typical signature. Biological trafficking events are very multifaceted. The algorithm was capable of identifying and characterizing the multistep biological movement, switching several times between subdiffusive, superdiffusive and Brownian motion.

Collaborators: Vincent Briane (UNSW Sydney, School of Medical Sciences, Australia), Myriam Vimond (CREST ENSAI Rennes), C.A. Valades Cruz and C. Wunder, (Institut Curie, PSL Research University, Cellular and Chemical Biology, U1143 INSERM / UMR 3666 CNRS).

7.5. Intracellular drift and diffusion coefficient estimation: a trajectory label-based approach

Participants: Antoine Salomon, Charles Kervrann.

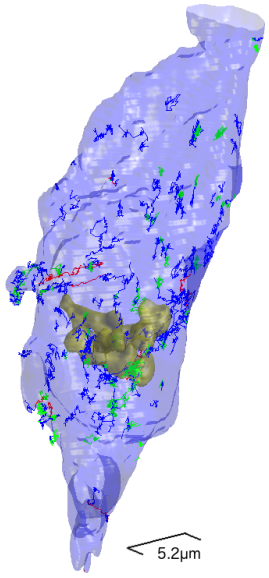


Figure 5. Change point detection over a set of three-dimensional trajectories of Galectin-3 proteins observed in HeLa cells with a Lattice Light Sheet microscope. The blue parts correspond to Brownian portions of the trajectory, red parts to superdiffusive portions, green parts to the subdiffusive portion. In light blue we plot the cell membrane and in yellow the Golgi apparatus for structural orientation.

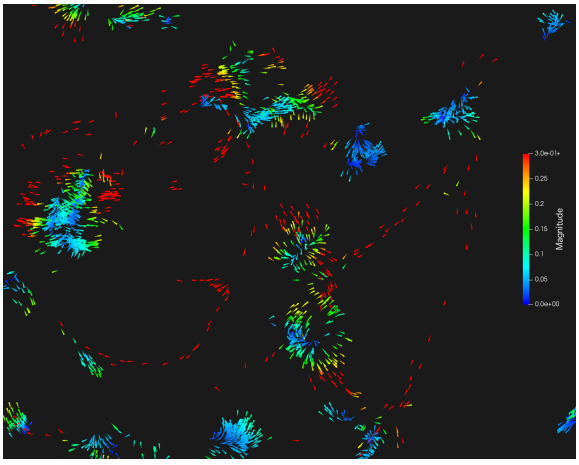


Figure 6. Estimation of a 2D drift field in 2D-TIRF microscopy of exocytosis biomolecules.

Dedicated computational methods for intracellular drift and diffusion map estimation rely on a scanning approach, using a sliding window (or cube, in the 3D case) in which all elementary particle movements taken from the trajectories inside the window/box are averaged [41]. This approach lacks precision, and a huge amount of trajectory data is needed to obtain significative results. In our new approach, we currently investigate several high-level features to obtain more satisfying results even from a small amount of data. First, we exploit classification of sub-trajectories (Brownian motion, superdiffusion, subdiffusion) obtained in the Lagrangian setting [16]. This classification is used to select trajectories of the same type in a local region and to guide weighting averaging. It allows us to calculate the drift separately on each type of movement, which avoids confusion between Brownian, confined and directed motions (see Fig. 6). Furthermore, the calculation is efficiently performed on trajectory-sliding kernels instead of scanning windows to save computing time in 2D and 3D. We considered square-box (sliding window), circle-box, cone-shaped, and Gaussian-shaped kernels.

Collaborators: Vincent Briane (UNSW Sydney, School of Medical Sciences, Australia),
Myriam Vimond (CREST ENSAI Rennes),
C.A. Valades Cruz and C. Wunder, (Institut Curie, PSL Research University, Cellular and Chemical Biology, U1143 INSERM / UMR 3666 CNRS).

7.6. 3D flow estimation in 3D fluorescence microscopy image sequences

Participants: Sandeep Manandhar, Patrick Bouthemy, Charles Kervrann.

Three-dimensional (3D) motion estimation for light-sheet microscopy is challenged by the heterogeneous scales and nature of intracellular dynamics. As typical examples in cell imaging, blebbing of a cell has small motion magnitude, while cell migration may show large displacement between frames. To tackle this problem, we have designed a two-stage 3D optical flow method. The first stage involves an extension of two-dimensional PatchMatch paradigm to 3D data, which in addition operates in a coarse-to-fine manner. We exploit multiple spatial scales to explore the possible range of intracellular motions. Our findings show that the metric based on Census transform is more robust to noise and to intensity variation between time steps. Only discrete displacements are estimated in this stage. Then, in a second stage, a 3D variational method enables to recover a sub-voxel dense 3D flow map. The variational approach still involves a data fidelity term based on the Census transform. The combination of the 3D PatchMatch and the 3D variational method is able to capture both large and small displacements. We assessed the performance of our method on data acquired with two different light sheet microscopes and compared it with a couple of other methods. The dataset depicts blebbing and migration of MV3 melanoma cells, and collagen network displacement induced by cell motility. As seen in Fig. 7 , our method is able to successfully estimate various range of motion during cell migration and blebbing. A straightforward way to visualize the resulting 3D flow field is to use 3D glyphs (arrows), which represent vector direction and magnitude. However, it may not lead to easy understanding for visualization in 3D and over time. Consequently, we propose to visualize 2D projections of 3D flow field in 3 orthogonal planes (see Figure 7 a.1.2 and b.1.2). Using the standard Middlebury-style color coding for 2D optic flow, the motion field becomes easier to understand. This work is carried out in collaboration with UTSW Dallas int the frame of the Inria associated team CytoDI.

Collaborators: Philippe Roudot and Erik Welf (UTSW, Dallas, USA).

7.7. Connecting trajectories in TIRF images of rod-shaped bacteria

Participants: Yunjiao Lu, Charles Kervrann.

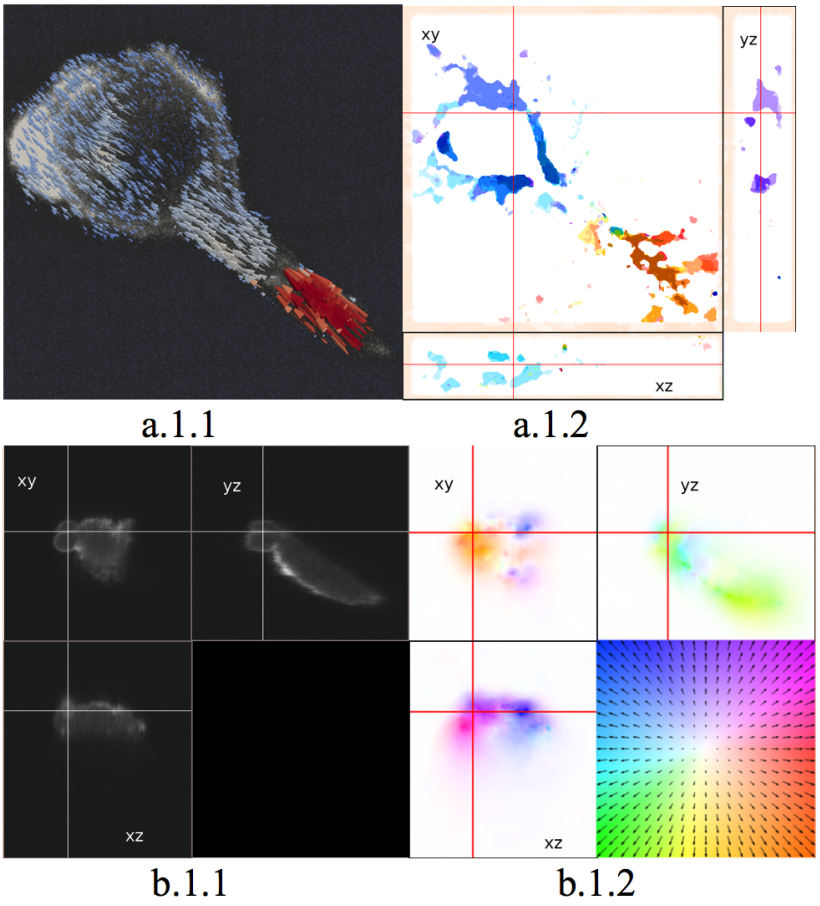


Figure 7. (a) Migration of MV3 melanoma cell in collagen; (a.1.1) 3D flow field in a slice of cell data represented with glyphs. Larger motion magnitude is coded in warm colors and smaller one in cold colors. (a.1.2) Motion map of collagen channel in 3 orthogonal planes (see b.1.2 for color code). (b) Blebbing of MV3 cell in a cover slip; (b.1.1) 3 orthogonal planes of cell data. (b.1.2) 3D computed flow field projected in 3 orthogonal planes. The motion map is color-coded as shown in lower-right corner of b.1.2. (input images by courtesy of Danuser lab, UTSW Dallas, USA).

In this work, we investigate the role of MreB protein involved in the cell wall construction in rod-shaped bacteria Bacillus. Previous work concerning the quantification of dynamics of MreB is performed from 2D TIRFM image sequences, ignoring the curvature of the rod-shaped cell wall and the thickness of TIRFM plane. To evaluate the effect of these approximations, we have developed a simulator to generate the trajectories on the surface of the cylinder. We assume that trajectories are modeled as:

$$dX = b(X)dt + \sqrt{2}B(X)dW,$$

where $b(X)$ is the drift tangent to the surface, W is a white noise, and $\sigma(X) = \frac{1}{2}B(X)B(X)^T$ is the diffusion tensor assumed to be isotropic. In our approach, the drift and diffusion tensor on the surface and on the projected plane are estimated respectively. Moreover, we propose to extrapolate the drift and diffusion tensor to the hidden region, to reconnect the segments we observe in the visible region, and then, to recover the entire trajectory on the surface of the cylinder in 3D (see Fig. 8).

Collaborators: Alain Trubuil and Pierre Hodara (INRA UR MAIAGE, Jouy-en-Josas),
Rut Carbllido-López and Cyrille Billaudeau (INRA, UR MICALIS, Jouy-en-Josas).

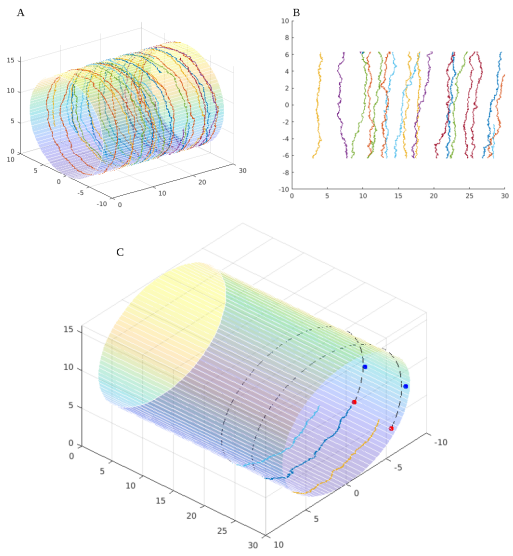


Figure 8. (A): Trajectories generated on the surface of the cylinder and (B): the view of TIRF microscope which is the projection of the dynamics on the cylinder near the support surface. The unity in the two figures is in pixel and in our TIRF images 1 pixel $\approx$ 64 nm. As the theoretical thickness of TIRF is 200 nm, the trajectories whose z coordinate is under 3.125 pixels are projected onto $x - y$ plan. Colors for different trajectories are random. (C): The connection result through the hidden region of three segments (light blue, dark blue and yellow in order of passage). Red points represent the extrapolated points using drift and diffusion tensor of segments on the surface, while blue points are the extrapolated points using projected drift and diffusion tensor on 2D plan.

7.8. 3D registration for correlative light-electron microscopy

Participants: Bertha Mayela Toledo Acosta, Patrick Bouthemy.

Correlative light and electron microscopy (CLEM) enables the study of cells and subcellular elements in complementary ways by combining information on the dynamics and on the structure of the cell, provided a reliable registration between light microscopy (LM) and electron microscopy (EM) images is efficiently achieved. We have developed a general automatic registration method. Due to large discrepancies in appearance, field-of-view, resolution and position, a pre-alignment stage is required before any 3D fine registration stage. We first compute a 2D maximum intensity projection (MPI) of the LM stack along the z-axis, and we match 2D EM regions of interest (ROI), extracted from different EM slices, into the 2D LM-MPI image. From the resulting candidates, we estimate, using a robust criterion, the 2D xy-shift to pre-align the LM and EM stacks. Afterwards, a 3D affine transformation between 3D-LM-ROI and 3D-EM–ROI can be estimated using mutual information. We carried out experimental results on different real datasets of 3D correlative microscopy, demonstrating computational efficiency and overlay accuracy.

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

7.9. 3D Convolutional Neural Networks for macromolecule localization in cryo-electron tomograms of intact cells

Participants: Emmanuel Moebel, Charles Kervrann.

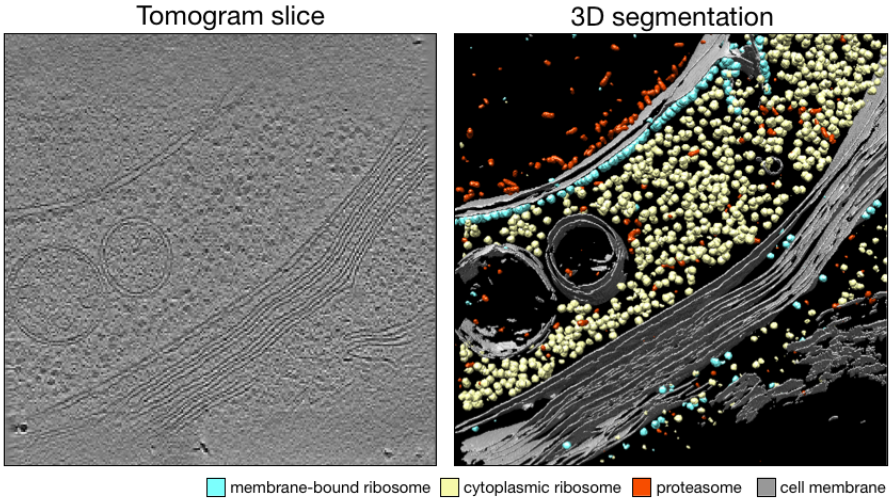


Figure 9. Experimental tomogram of a Chlamydomonas Reinhardtii cell (central slice) and the 3D segmentation obtained by our convolutional neural network. The network is able to identify 3 macromolecule classes (membrane-bound ribosome, cytoplasmic ribosome and proteasome) as well as the cell membrane (tomogram by courtesy of Max-Planck Institute, Martinsried, Germany).

In this study, we focus on macromolecule localization and classification in cryo-electron tomography images. Cryo-electron tomography (cryo-ET) allows one to capture 3D images of cells in a close to native state, at sub-nanometer resolution. However, noise and artifact levels are such that heavy computational processing is needed to access the image content. We propose a deep learning framework to accurately and jointly localize multiple types and states of macromolecules in cellular cryo-electron tomograms. We compare this framework to the commonly-used template matching method on both synthetic and experimental data. On synthetic image data, we show that our framework is very fast and produces superior detection results. On experimental data,

the detection results obtained by our method correspond to an overlap rate of 86% with the expert annotations, and comparable resolution is achieved when applying subtomogram averaging. In addition, we show that our method can be combined to template matching procedures to reliably increase the number of expected detections. In our experiments, this strategy was able to find additional 20.5% membrane-bound ribosomes that were missed or discarded during manual annotation.

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

7.10. Data assimilation and modeling of cell division mechanism

Participants: Ancageorgiana Caranfil, Charles Kervrann.

In this work, we focus on the dynamics of the spindle during cell division mechanism. We aim at understanding the role and interaction of the molecular key players at different scales, and their individual and collective impact on the global mechanism at the cell level. Our approach consists in creating a biophysical model for this mechanism, and uses data assimilation to adjust the model and optimally integrate the information from the observations. The overall spindle behavior is led by the spindle poles behavior. We thus proposed a new biophysical model for the spindle pole functioning during anaphase, that explains the oscillatory behavior with a minimum number of parameters. By mathematically analyzing our model, we confirmed some previous findings, such as the existence of a threshold number of active force-generator motors required for the onset of oscillations. We also confirmed that the monotonic increase of motor activity accounts for their build-up and die-down. We determined boundaries for the motor activity-related parameters for these oscillations to happen. This also allowed us to describe the influence of the number of motors, as well as physical parameters related to viscosity or string-like forces, on features such as the amplitude and frequency of oscillations. Lastly, by using a Bayesian approach to confront our model to experimental data, we were able to estimate distributions for our biological and biophysical parameters. A statistical reduction model approach was preliminary applied to select the most influential model parameters. These results give us insights on variations in spindle behavior during anaphase in asymmetric division, and provide means of prediction for phenotypes related to misguided asymmetric division. Data assimilation will be further used to properly combine the information given by our model and the experimental data.

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

7.11. Motion saliency in videos

Participants: Léo Maczyta, Patrick Bouthemy.

The problem we have addressed appertains to the domain of motion saliency in videos. More specifically, we aim to extract the temporal segments of the video where motion saliency is present. It is a prerequisite for computing motion saliency maps in relevant images. It turns out to be a frame classification problem. A frame is classified as dynamically salient if it contains local motion departing from its context. Temporal motion saliency detection is relevant for applications where one needs to trigger alerts or to monitor dynamic behaviors from videos. The proposed approach handles situations with a mobile camera, and involves a deep learning classification framework after camera motion compensation. We have designed and compared two methods respectively based on image warping, and on residual flow. A baseline that relies on a two-stream network to process temporal and spatial information, but that does not use camera motion compensation, was also defined. Experiments on real videos demonstrate that we can obtain an accurate classification in highly challenging situations, and get significant improvement over the baseline. In particular, we showed that the compensation of the camera motion produces a better classification. We also showed that for the limited training data available, providing the residual flow as input to the classification network produces better results than providing the warped images.

Collaborators: Olivier Le Meur (Percept team, Irisa).

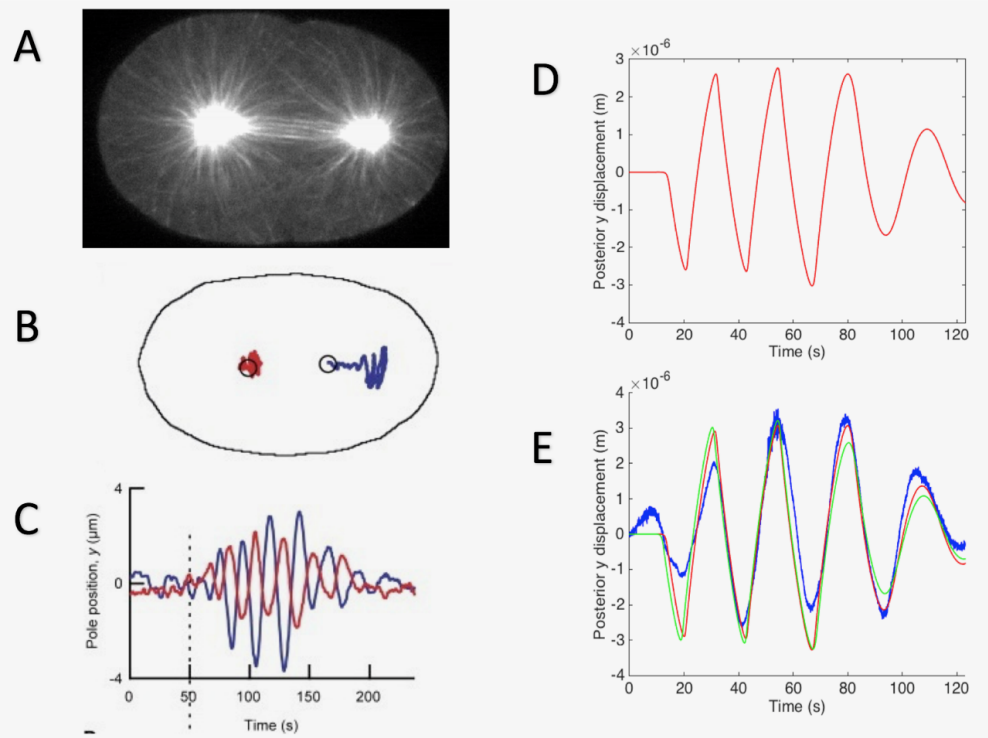


Figure 10. A. Cell division observed by fluorescence microscopy. B. and C. Tracking of the two spindle pole (anterior in red and posterior in blue). Oscillations of the poles during metaphase and anaphase. D. Simulation of oscillations using our model. E. Fitting of experimental data (in blue), and the two estimated curves with our method (minimum estimator in red, and mean estimator in green).

Method	Motion saliency timeline
Ground truth	
RFS-Motion2D	
RFS-DeepDOM	
WS-Motion2D	
WS-DeepDOM	

Figure 11. Timeline of the classification results supplied by the methods, for 12 videos of the testing set. Orange denotes salient frames, and blue non salient frames. Videos are separated by horizontal spaces. WS and RFS denote respectively the use of image warping and or residual flow in order to predict saliency. For dominant parametric motion estimation, the robust classical method Motion2D and the neural network DeepDOM were used.

ARAMIS Project-Team

7. New Results

7.1. Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data

Participants: Jorge Samper-González, Ninon Burgos, Simona Bottani, Sabrina Fontanella, Pascal Lu, Arnaud Marcoux, Alexandre Routier, Jérémy Guillon, Michael Bacci, Junhao Wen, Anne Bertrand, Hugo Bertin, Marie-Odile Habert, Stanley Durrleman, Theodoros Evgeniou, Olivier Colliot [Correspondant].

A large number of papers have introduced novel machine learning and feature extraction methods for automatic classification of Alzheimer's disease (AD). However, while the vast majority of these works use the public dataset ADNI for evaluation, they are difficult to reproduce because different key components of the validation are often not readily available. These components include selected participants and input data, image preprocessing and cross-validation procedures. The performance of the different approaches is also difficult to compare objectively. In particular, it is often difficult to assess which part of the method (e.g. preprocessing, feature extraction or classification algorithms) provides a real improvement, if any. We proposed a framework for reproducible and objective classification experiments in AD using three publicly available datasets (ADNI, AIBL and OASIS). The framework comprises: i) automatic conversion of the three datasets into a standard format (BIDS); ii) a modular set of preprocessing pipelines, feature extraction and classification methods, together with an evaluation framework, that provide a baseline for benchmarking the different components. We demonstrated the use of the framework for a large-scale evaluation on 1960 participants using T1 MRI and FDG PET data. In this evaluation, we assessed the influence of different modalities, preprocessing, feature types (regional or voxel-based features), classifiers, training set sizes and datasets. Performances were in line with the state-of-the-art. FDG PET outperformed T1 MRI for all classification tasks. No difference in performance was found for the use of different atlases, image smoothing, partial volume correction of FDG PET images, or feature type. Linear SVM and L2-logistic regression resulted in similar performance and both outperformed random forests. The classification performance increased along with the number of subjects used for training. Classifiers trained on ADNI generalized well to AIBL and OASIS. All the code of the framework and the experiments is publicly available: general-purpose tools have been integrated into the Clinica software (http://www.clinica.run/) and the paper-specific code is available at: https://gitlab.icm-institute.org/aramislab/AD-ML.

More details in [30].

7.2. An automated pipeline for the analysis of PET data on the cortical surface

Participants: Arnaud Marcoux, Ninon Burgos, Anne Bertrand, Marc Teichmann, Alexandre Routier, Junhao Wen, Jorge Samper-González, Simona Bottani, Stanley Durrleman, Marie-Odile Habert, Olivier Colliot [Correspondant].

We developed a fully automatic pipeline for the analysis of PET data on the cortical surface. Our pipeline combines tools from FreeSurfer and PETPVC, and consists of i) co-registration of PET and T1-w MRI (T1) images, ii) intensity normalization, iii) partial volume correction, iv) robust projection of the PET signal onto the subject's cortical surface, v) spatial normalization to a template, and vi) atlas statistics. We evaluated the performance of the proposed workflow by performing group comparisons and showed that the approach was able to identify the areas of hypometabolism characteristic of different dementia syndromes: Alzheimer's disease (AD) and both the semantic and logopenic variants of primary progressive aphasia. We also showed that these results were comparable to those obtained with a standard volume-based approach. We then performed individual classifications and showed that vertices can be used as features to differentiate cognitively normal and AD subjects. This pipeline is integrated into Clinica, an open-source software platform for neuroscience studies available at http://www.clinica.run/.

More details in [24].

7.3. Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region

Participants: Hossein Arabi, Jason A. Dowling, Ninon Burgos [Correspondant], Xiao Han, Peter B. Greer, Nikolaos Koutsouvelis, Habib Zaidi.

Magnetic resonance imaging (MRI)-guided radiation therapy (RT) treatment planning is limited by the fact that the electron density distribution required for dose calculation is not readily provided by MR imaging. We compare a selection of novel synthetic CT generation algorithms recently reported in the literature, including segmentation-based, atlas-based and machine learning techniques, using the same cohort of patients and quantitative evaluation metrics. Six MRI-guided synthetic CT generation algorithms were evaluated: one segmentation technique into a single tissue class (water-only), four atlas-based techniques, namely, median value of atlas images (ALMedian), atlas-based local weighted voting (ALWV), bone enhanced atlas-based local weighted voting (ALWV-Bone), iterative atlas-based local weighted voting (ALWV-Iter), and a machine learning technique using deep convolution neural network (DCNN). Organ auto-contouring from MR images was evaluated for bladder, rectum, bones, and body boundary. Overall, DCNN exhibited higher segmentation accuracy resulting in Dice indices while ALMedian showed the lowest accuracy. DCNN reached the best performance in terms of accurate derivation of synthetic CT values within each organ, followed by the advanced atlas-based methods. ALMedian led to the highest error. Considering the dosimetric evaluation results, ALWV-Iter, ALWV, DCNN and ALWV-Bone led to similar mean dose estimation within each organ at risk and target volume with less than 1% dose discrepancy. However, the two-dimensional gamma analysis demonstrated higher pass rates for ALWV-Bone, DCNN, ALMedian and ALWV-Iter at 1%/1 mm criterion. Overall, machine learning and advanced atlas-based methods exhibited promising performance by achieving reliable organ segmentation and synthetic CT generation. DCNN appears to have slightly better performance by achieving accurate automated organ segmentation and relatively small dosimetric errors (followed closely by advanced atlas-based methods, which in some cases achieved similar performance). However, the DCNN approach showed higher vulnerability to anatomical variation, where a greater number of outliers was observed with this method. Considering the dosimetric results obtained from the evaluated methods, the challenge of electron density estimation from MR images can be resolved with a clinically tolerable error.

More details in [4].

7.4. Double diffeomorphism: combining morphometry and structural connectivity analysis

Participants: Pietro Gori, Olivier Colliot, Linda Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman [Correspondant].

The brain is composed of several neural circuits which may be seen as anatomical complexes composed of grey matter structures interconnected by white matter tracts. Grey and white matter components may be modelled as 3D surfaces and curves respectively. Neurodevelopmental disorders involve morphological and organizational alterations which can not be jointly captured by usual shape analysis techniques based on single diffeomorphisms. We propose a new deformation scheme, called double diffeomorphism, which is a combination of two diffeomorphisms. The first one captures changes in structural connectivity, whereas the second one recovers the global morphological variations of both grey and white matter structures. This deformation model is integrated into a Bayesian framework for atlas construction. We evaluate it on a data-set of 3D structures representing the neural circuits of patients with Gilles de la Tourette syndrome (GTS). We show that this approach makes it possible to localise, quantify and easily visualise the pathological anomalies altering the morphology and organization of the neural circuits. Furthermore, results also indicate that the proposed deformation model better discriminates between controls and GTS patients than a single diffeomorphism.

More details in [15].

7.5. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms

Participants: Alexandre Bône, Olivier Colliot, Stanley Durrleman [Correspondant].

We propose a method to learn a distribution of shape tra-jectories from longitudinal data, i.e. the collection of individual objects repeatedly observed at multiple time-points. The method allows to compute an average spatiotemporal trajectory of shape changes at the group level, and the individual variations of this trajectory both in terms of geometry and time dynamics. First, we formulate a non-linear mixed-effects statistical model as the combination of a generic statistical model for manifold-valued longitudinal data, a deformation model defining shape trajectories via the action of a finite-dimensional set of diffeomorphisms with a man-ifold structure, and an efficient numerical scheme to compute parallel transport on this manifold. Second, we introduce a MCMC-SAEM algorithm with a specific approach to shape sampling, an adaptive scheme for proposal variances , and a log-likelihood tempering strategy to estimate our model. Third, we validate our algorithm on 2D simulated data, and then estimate a scenario of alteration of the shape of the hippocampus 3D brain structure during the course of Alzheimer's disease. The method shows for instance that hippocampal atrophy progresses more quickly in female subjects, and occurs earlier in APOE4 mutation carriers. We finally illustrate the potential of our method for classifying pathological trajectories versus normal ageing.

More details in [38].

7.6. Spatiotemporal Propagation of the Cortical Atrophy: Population and Individual Patterns

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

Repeated failures in clinical trials for Alzheimer's disease (AD) have raised a strong interest for the prodromal phase of the disease. A better understanding of the brain alterations during this early phase is crucial to diagnose patients sooner, to estimate an accurate disease stage, and to give a reliable prognosis. According to recent evidence, structural alterations in the brain are likely to be sensitive markers of the disease progression. Neuronal loss translates in specific spatiotemporal patterns of cortical atrophy, starting in the enthorinal cortex and spreading over other cortical regions according to specific propagation pathways. We developed a digital model of the cortical atrophy in the left hemisphere from prodromal to diseased phases, which is built on the temporal alignment and combination of several short-term observation data to reconstruct the long-term history of the disease. The model not only provides a description of the spatiotemporal patterns of cortical atrophy at the group level but also shows the variability of these patterns at the individual level in terms of difference in propagation pathways, speed of propagation, and age at propagation onset. Longitudinal MRI datasets of patients with mild cognitive impairments who converted to AD are used to reconstruct the cortical atrophy propagation across all disease stages. Each observation is considered as a signal spatially distributed on a network, such as the cortical mesh, each cortex location being associated to a node. We consider how the temporal profile of the signal varies across the network nodes. We introduce a statistical mixed-effect model to describe the evolution of the cortex alterations. To ensure a spatiotemporal smooth propagation of the alterations, we introduce a constrain on the propagation signal in the model such that neighboring nodes have similar profiles of the signal changes. Our generative model enables the reconstruction of personalized patterns of the neurodegenerative spread, providing a way to estimate disease progression stages and predict the age at which the disease will be diagnosed. The model shows that, for instance, APOE carriers have a significantly higher pace of cortical atrophy but not earlier atrophy onset.

More details in [19].

7.7. A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds

Participants: Maxime Louis, Benjamin Charlier, Paul Jusselin, Susovan Pal, Stanley Durrleman.

Parallel transport on Riemannian manifolds allows one to connect tangent spaces at different points in an isometric way and is therefore of importance in many contexts, such as for statistics on manifolds. The existing methods to compute parallel transport require either the computation of Riemannian logarithms, such as the Schild's ladder, or the Christoffel symbols. The Logarithm is rarely given in closed form, and therefore costly to compute whereas the number of Christoffel symbols explodes with the dimension of the manifold, making both these methods intractable. From an identity between parallel transport and Jacobi fields, we propose a numerical scheme to approximate the parallel transport along a geodesic. We find and prove an optimal convergence rate for the scheme, which is equivalent to Schild's ladder's. We investigate potential variations of the scheme and give experimental results on the Euclidean two-sphere and on the manifold of symmetric positive-definite matrices.

More details in [23].

7.8. Reduction of recruitment costs in preclinical AD trials. Validation of automatic pre-screening algorithm for brain amyloidosis.

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

We propose a method for recruiting asymptomatic Amyloid positive individuals in clinical trials, using a two-step process. We first select during a pre-screening phase a subset of individuals which are more likely to be amyloid positive based on the automatic analysis of data acquired during routine clinical practice, before doing a confirmatory PET-scan to these selected individuals only. This method leads to an increased number of recruitments and to a reduced number of PET-scans, resulting in a decrease in overall recruitment costs. We validate our method on 3 different cohorts, and consider 5 different classification algorithms for the pre-screening phase. We show that the best results are obtained using solely cognitive, genetic and socio-demographic features, as the slight increased performance when using MRI or longitudinal data is balanced by the cost increase they induce. We show that the proposed method generalizes well when tested on an independent cohort, and that the characteristics of the selected set of individuals are identical to the characteristics of a population selected in a standard way. The proposed approach shows how Machine Learning can be used effectively in practice to optimize recruitment costs in clinical trials.

More details in [3].

7.9. Multiplex core-periphery organization of the human connectome

Participants: Federico Battiston, Jeremy Guillon, Mario Chavez, Vito Latora, Fabrizio de Vico Fallani [Correspondant].

What is the core of the human brain is a fundamental question that has been mainly addressed by studying the anatomical connections between differently specialized areas, thus neglecting the possible contributions from their functional interactions. While many methods are available to identify the core of a network when connections between nodes are all of the same type, a principled approach to define the core when multiple types of connectivity are allowed is still lacking. Here, we introduce a general framework to define and extract the core–periphery structure of multi-layer networks by explicitly taking into account the connectivity patterns at each layer. We first validate our algorithm on synthetic networks of different size and density, and with tunable overlap between the cores at different layers. We then use our method to merge information from structural and functional brain networks, obtaining in this way an integrated description of the core of the human connectome. Results confirm the role of the main known cortical and subcortical hubs, but also suggest the presence of new areas in the sensori-motor cortex that are crucial for intrinsic brain functioning. Taken together these findings provide fresh evidence on a fundamental question in modern neuroscience and offer new opportunities to explore the mesoscale properties of multimodal brain networks.

More details in [6].

7.10. Integrating EEG and MEG signals to improve motor imagery classification in brain-computer interfaces

Participants: Marie-Constance Corsi, Mario Chavez, Denis Schwartz, Laurent Hugueville, Ankit Khambhati, Danielle Bassett, Fabrizio de Vico Fallani [Correspondant].

We adopted a fusion approach that combines features from simultaneously recorded electroencephalogram (EEG) and magnetoencephalogram (MEG) signals to improve classification performances in motor imagery-based brain–computer interfaces (BCIs). We applied our approach to a group of 15 healthy subjects and found a significant classification performance enhancement as compared to standard single-modality approaches in the alpha and beta bands. Taken together, our findings demonstrate the advantage of considering multimodal approaches as complementary tools for improving the impact of noninvasive BCIs.

More details in [10].

7.11. Role of inter-hemispheric connections in functional brain networks

Participants: Johann Martinez [Correspondant], Javier Buldu, David Papo, Fabrizio de Vico Fallani, Mario Chavez.

Today the human brain can be modeled as a graph where nodes represent different regions and links stand for statistical interactions between their activities as recorded by different neuroimaging techniques. Empirical studies have lead to the hypothesis that brain functions rely on the coordination of a scattered mosaic of functionally specialized brain regions (modules or sub-networks), forming a web-like structure of coordinated assemblies (a network of networks). The study of brain dynamics would therefore benefit from an inspection of how functional sub-networks interact between them. In this paper, we model the brain as an interconnected system composed of two specific sub-networks, the left (L) and right (R) hemispheres, which compete with each other for centrality, a topological measure of importance in a networked system. Specifically, we considered functional brain networks derived from high-density electroencephalographic (EEG) recordings and investigated how node centrality is shaped by interhemispheric connections. Our results show that the distribution of centrality strongly depends on the number of functional connections between hemispheres and the way these connections are distributed. Additionally, we investigated the consequences of node failure on hemispherical centrality, and showed how the abundance of inter-hemispheric links favors the functional balance of centrality distribution between the hemispheres.

More details in [25].

7.12. Statistical shape analysis of large datasets based on diffeomorphic iterative centroids

Participants: Claire Cury, Joan Glaunès, Olivier Colliot.

We proposed an approach for template-based shape analysis of large datasets, using diffeomorphic centroids as atlas shapes. Diffeomorphic centroid methods fit in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework and use kernel metrics on currents to quantify surface dissimilarities. The statistical analysis is based on a Kernel Principal Component Analysis (Kernel PCA) performed on the set of initial momentum vectors which parametrize the deformations. We tested the approach on different datasets of hippocampal shapes extracted from brain magnetic resonance imaging (MRI), compared three different centroid methods and a variational template estimation. The largest dataset is composed of 1,000 surfaces, and we are able to analyse this dataset in 26 h using a diffeomorphic centroid. Our experiments demonstrate that computing diffeomorphic centroids in place of standard variational templates leads to similar shape analysis results and saves around 70% of computation time. Furthermore, the approach is able to adequately capture the variability of hippocampal shapes with a reasonable number of dimensions, and to predict anatomical features of the hippocampus, only present in 17% of the population, in healthy subjects.

More details in [12].

7.13. Multi-modal brain fingerprinting: a manifold approximation based framework

Participants: Kuldeep Kumar, Olivier Colliot, Christian Desrosiers.

We proposed an efficient framework, based on manifold approximation, for generating brain fingerprints from multi-modal data. The proposed framework represents images as bags of local features, which are used to build a subject proximity graph. Compact fingerprints are obtained by projecting this graph in a low-dimensional manifold, using spectral embedding. Experiments using the T1/T2-weighted MRI, diffusion MRI, and resting state fMRI data of 945 Human Connectome Project subjects demonstrate the benefit of combining multiple modalities, with multi-modal fingerprints more discriminative than those generated from individual modalities. Results also highlight the link between fingerprint similarity and genetic proximity, monozygotic twins having more similar fingerprints than dizygotic or non-twin siblings. This link is also reflected in the differences of feature correspondences between twin/sibling pairs, occurring in major brain structures and across hemispheres. The robustness of the proposed framework to factors like image alignment and scan resolution, as well as the reproducibility of results on retest scans, suggest the potential of multi-modal brain fingerprinting for characterizing individuals in a large cohort analysis. In addition, taking inspiration from the computer vision community, the proposed rank retrieval evaluation based on the task of twin/sibling identification and using Mean Average Precision (MAP) can be used for a standardized comparison of future brain fingerprints.

More details in [20].

7.14. Structural, Microstructural, and Metabolic Alterations in Primary Progressive Aphasia Variants

Participants: Alexandre Routier [Correspondant], Marie-Odile Habert, Olivier Colliot, Marc Teichmann.

Neuroimaging studies have described the brain alterations in primary progressive aphasia (PPA) variants (semantic, logopenic, nonfluent/agrammatic). However, few studies combined T1, FDG-PET, and diffusion MRI techniques to study atrophy, hypometabolism, and tract alterations across the three PPA main variants. We therefore explored a large early-stage cohort of semantic, logopenic and nonfluent/agrammatic variants (N = 86) and of 23 matched healthy controls with anatomical MRI (cortical thickness), FDG PET (metabolism) and diffusion MRI (white matter tracts analyses), aiming at identifying cortical and sub-cortical brain alterations, and confronting these alterations across imaging modalities and aphasia variants. In the semantic variant, there was cortical thinning and hypometabolism in anterior temporal cortices, with left-hemisphere predominance, extending toward posterior temporal regions, and affecting tracts projecting to the anterior temporal lobes (inferior longitudinal fasciculus, uncinate fasciculus) and tracts projecting to or running nearby posterior temporal cortices: (superior longitudinal fasciculus, inferior frontal-occipital fasciculus). In the logopenic variant metabolic alterations were more extensive than atrophy affecting mainly the left temporal-parietal junction and extending toward more anterior temporal cortices. Metabolic and tract data were coherent given the alterations of the left superior and inferior longitudinal fasciculus and the left inferior frontal-occipital fasciculus. In the nonfluent/agrammatic variant cortical thinning and hypometabolism were located in the left frontal cortex but Broca's area was only affected on metabolic measures. Metabolic and tract alterations were coherent as reflected by damage to the left uncinate fasciculus connecting with Broca's area. Our findings provide a full-blown statistically robust picture of brain alterations in early-stage variants of primary progressive aphasia which has implications for diagnosis, classification and future therapeutic strategies. They demonstrate that in logopenic and semantic variants patterns of brain damage display a non-negligible overlap in temporal regions whereas they are substantially distinct in the nonfluent/agrammatic variant (frontal regions). These results also indicate that frontal networks (combinatorial syntax/phonology) and temporal networks (lexical/semantic representations) constitute distinct anatomo-functional entities with differential vulnerability to degenerative processes in aphasia variants. Finally, the identification of the specific damage

patterns could open an avenue for trans-cranial stimulation approaches by indicating the appropriate target-entry into the damaged language system.

More details in [29].

7.15. Neurite density is reduced in the presymptomatic phase of C9orf72 disease

Participants: Junhao Wen, Hui Zhang, Daniel Alexander, Stanley Durrleman, Olivier Colliot, Isabelle Le Ber, Anne Bertrand [Correspondant].

In this study, we aimed to assess the added value of neurite orientation dispersion and density imaging (NODDI) compared to conventional DTI and anatomical MRI to detect changes in presymptomatic carriers of chromosome 9 open reading frame 72 (C9orf72) mutation. The PREV-DEMALS study is a prospective, multicenter, observational study of first-degree relatives of individuals carrying the C9orf72 mutation. Sixty-seven participants (38 presymptomatic C9orf72 mutation carriers [C9+], 29 non carriers [C9-]) were included in the present cross-sectional study. Each participant underwent one single-shell, multi-shell diffusion MRI and 3DT1 MRI. Volumetric measures, DTI and NODDI metrics were calculated within regions of interest. Differences in white matter integrity, gray matter volume and free water fraction between C9+ and C9- individuals were assessed using linear mixed-effects models. Compared with C9-, C9+ demonstrated white matter abnormalities in 10 tracts with neurite density index, and only 5 tracts with DTI metrics. Effect size was significantly higher for the neurite density index than for DTI metrics in two tracts. No tract had a significantly higher effect size for DTI than for NODDI. For gray matter cortical analysis, free water fraction was increased in 13 regions in C9+, whereas 11 regions displayed volumetric atrophy. In conclusion, NODDI provides higher sensitivity and greater tissue-specificity compared to conventional DTI for identifying white matter abnormalities in the presymptomatic C9orf72 carriers. Our results encourage the use of neurite density as biomarker of the preclinical phase.

More details in [34].

7.16. Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training

Participants: Wen Wei, Emilie Poirion, Benedetta Bodini, Stanley Durrleman, Nicholas Ayache, Bruno Stankoff, Olivier Colliot [Correspondant].

Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS). A reliable measure of the tissue myelin content is therefore essential to understand the physiopathology of MS, track progression and assess treatment efficacy. Positron emission tomography (PET) with [11 C]PIB has been proposed as a promising biomarker for measuring myelin content changes in-vivo in MS. However, PET imaging is expensive and invasive due to the injection of a radioactive tracer. On the contrary, magnetic resonance imaging (MRI) is a non-invasive, widely available technique, but existing MRI sequences do not provide, to date, a reliable, specific, or direct marker of either demyelination or remyelination. In this work, we therefore propose Sketcher-Refiner Generative Adversarial Networks (GANs) with specifically designed adversarial loss functions to predict the PET-derived myelin content map from a combination of MRI modalities. The prediction problem is solved by a sketch-refinement process in which the sketcher generates the preliminary anatomical and physiological information and the refiner refines and generates images reflecting the tissue myelin content in the human brain. We evaluated the ability of our method to predict myelin content at both global and voxel-wise levels. The evaluation results show that the demyelination in lesion regions and myelin content in normal-appearing white matter (NAWM) can be well predicted by our method. The method has the potential to become a useful tool for clinical management of patients with MS.

More details in [40].

7.17. COGEVIS: A New Scale to Evaluate Cognition in Patients with Visual Deficiency

Participants: Claire Meyniel, Dalila Samri, Farah Stefano, Joel Crevoisier, Florence Bonté, Raffaela Migliaccio, Laure Delaby, Anne Bertrand, Marie-Odile Habert, Bruno Dubois, Baram Bodaghi, Stéphane Epelbaum [Correspondant].

We evaluated the cognitive status of visually impaired patients referred to low vision rehabilitation (LVR) based on a standard cognitive battery and a new evaluation tool, named the COGEVIS, which can be used to assess patients with severe visual deficits. We studied patients aged 60 and above, referred to the LVR Hospital in Paris. Neurological and cognitive evaluations were performed in an expert memory center. Thirty-eight individuals, 17 women and 21 men with a mean age of 70.3(SD=1.3 years) and a mean visual acuity of 0.12(SD=0.02), were recruited over a one-year period. Sixty-three percent of participants had normal cognitive status. Cognitive impairment was diagnosed in 37.5% of participants. The COGEVIS score cutoff point to screen for cognitive impairment was 24 (maximum score of 30) with a sensitivity of 66.7% and a specificity of 95%. Evaluation following 4 months of visual rehabilitation showed an improvement of Instrumental Activities of Daily Living ($p = 0.004$), National Eye Institute Visual Functioning Questionnaire ($p = 0.035$), and Montgomery-Åsberg Depression Rating Scale ($p = 0.037$). This study introduces a new short test to screen for cognitive impairment in visually impaired patients.

More details in [27].

7.18. Neural correlates of episodic memory in the Memento cohort

Participants: Stéphane Epelbaum [Correspondant], Vincent Bouteloup, Jean François Mangin, Valentina La Corte, Raffaela Migliaccio, Hugo Bertin, Marie Odile Habert, Clara Fischer, Chabha Azouani, Ludovic Fillon, Marie Chupin, Bruno Vellas, Florence Pasquier, Frederic Blanc, Audrey Gabelle, Mathieu Ceccaldi, Pierre Krolak-Salmon, Jacques Hugon, Olivier Hanon, Olivier Rouaud, Renaud David, Genevieve Chene, Bruno Dubois, Carole Dufouil.

The free and cued selective reminding test is used to identify memory deficits in mild cognitive impairment and demented patients. It allows assessing three processes: encoding, storage, and recollection of verbal episodic memory. We investigated the neural correlates of these three memory processes in a large cohort study. The Memento cohort enrolled 2323 outpatients presenting either with subjective cognitive decline or mild cognitive impairment who underwent cognitive, structural MRI and, for a subset, fluorodeoxyglucose-positron emission tomography evaluations. Encoding was associated with a network including parietal and temporal cortices; storage was mainly associated with entorhinal and parahippocampal regions, bilaterally; retrieval was associated with a widespread network encompassing frontal regions. The neural correlates of episodic memory processes can be assessed in large and standardized cohorts of patients at risk for Alzheimer's disease. Their relation to pathophysiological markers of Alzheimer's disease remains to be studied.

7.19. Cognitive and neuroimaging features and brain amyloidosis in individuals at risk of Alzheimer's disease

Participants: Bruno Dubois [Correspondant], Stéphane Epelbaum, Francis Nyasse, Hovagim Bakardjian, Geoffroy Gagliardi, Olga Uspenskaya, Marion Houot, Simone Lista, Federica Cacciamani, Marie Claude Potier, Anne Bertrand, Foudil Lamari, Habib Benali, Jean François Mangin, Olivier Colliot, Remy Genthon, Marie-Odile Habert, Harald Hampel.

Improved understanding is needed of risk factors and markers of disease progression in preclinical Alzheimer's disease. We assessed associations between brain amyloidosis and various cognitive and neuroimaging parameters with progression of cognitive decline in individuals with preclinical Alzheimer's disease. The INSIGHT-preAD is an ongoing single-centre observational study at the Salpêtrière Hospital, Paris, France. Eligible participants were age 70-85 years with subjective memory complaints but unimpaired cognition and memory (Mini-Mental State Examination [MMSE] score $\geq$ 27, Clinical Dementia Rating score 0, and Free and

Cued Selective Reminding Test [FCSRT] total recall score ≥ 41). We stratified participants by brain amyloid deposition on 18F-florbetapir PET (positive or negative) at baseline. All patients underwent baseline assessments of demographic, cognitive, and psychobehavioural, characteristics, APOE $\epsilon 4$ allele carrier status, brain structure and function on MRI, brain glucose-metabolism on 18F-fluorodeoxyglucose (18F-FDG) PET, and event-related potentials on electroencephalograms (EEGs). Actigraphy and CSF investigations were optional. Participants were followed up with clinical, cognitive, and psychobehavioural assessments every 6 months, neuropsychological assessments, EEG, and actigraphy every 12 months, and MRI, and 18F-FDG and 18F-florbetapir PET every 24 months. We assessed associations of amyloid deposition status with test outcomes at baseline and 24 months, and with clinical status at 30 months. Progression to prodromal Alzheimer's disease was defined as an amnestic syndrome of the hippocampal type. From May 25, 2013, to Jan 20, 2015, we enrolled 318 participants with a mean age of 76·0 years (SD 3·5). The mean baseline MMSE score was 28·67 (SD 0·96), and the mean level of education was high (score >6 [SD 2] on a scale of 1-8, where 1=infant school and 8=higher education). 88 (28showed amyloid deposition and the remainder did not. The amyloid subgroups did not differ for any psychobehavioural, cognitive, actigraphy, and structural and functional neuroimaging results after adjustment for age, sex, and level of education More participants positive for amyloid deposition had the APOE $\epsilon 4$ allele (33 [38%] vs 29 [13%], p<0·0001). Amyloid concentration in CSF significantly correlated with mean 18F-florbetapir uptake at baseline (r=-0·62, p<0·0001) and the ratio of amyloid to amyloid (r=-0·61, p<0·0001), and identified amyloid deposition status with high accuracy (mean area under the curve values 0·89, 95% CI 0·80-0·98 and 0·84, 0·72-0·96, respectively). No difference was seen in MMSE (28·3 [SD 2·0] vs 28·9 [1·2], p=0·16) and Clinical Dementia Rating scores (0·06 [0·2] vs 0·05 [0·3]; p=0·79) at 30 months (n=274) between participants positive or negative for amyloid. Four participants (all positive for amyloid deposition at baseline) progressed to prodromal Alzheimer's disease. They were older than other participants positive for amyloid deposition at baseline (mean 80·2 years [SD 4·1] vs 76·8 years [SD 3·4]) and had greater 18F-florbetapir uptake at baseline (mean standard uptake value ratio 1·46 [SD 0·16] vs 1·02 [SD 0·20]), and more were carriers of the APOE $\epsilon 4$ allele (three [75%] of four vs 33 [39%] of 83). They also had mild executive dysfunction at baseline (mean FCSRT free recall score 21·25 [SD 2·75] vs 29·08 [5·44] and Frontal Assessment Battery total score 13·25 [1·50] vs 16·05 [1·68]). Brain amyloidosis alone did not predict progression to prodromal Alzheimer's disease within 30 months. Longer follow-up is needed to establish whether this finding remains consistent.

More details in [13].

ATHENA Project-Team

6. New Results

6.1. Computational Diffusion MRI

6.1.1. Reducing the Number of Samples in Spatio-Temporal dMRI Acquisition Design

Participants: Patryk Filipiak, Rutger Fick [TheraPanacea, Paris], Alexandra Petiet [ICM, CENIR, Paris], Mathieu Santin [ICM, CENIR, Paris], Anne-Charlotte Philippe [ICM, CENIR, Paris], Stéphane Lehericy [ICM, CENIR, Paris], Philippe Ciuciu [CEA, Université Paris-Saclay], Demian Wassermann [Inria Parietal], Rachid Deriche.

Acquisition time is a major limitation in recovering brain white matter microstructure with diffusion magnetic resonance imaging. The aim of this work is to bridge the gap between growing demands on spatiotemporal resolution of diffusion signal and the real-world time limitations. We introduce an acquisition scheme that reduces the number of samples under adjustable quality loss. Finding a sampling scheme that maximizes signal quality and satisfies given time constraints is NP-hard. Therefore, a heuristic method based on a genetic algorithm is proposed in order to find suboptimal solutions in acceptable time. The analyzed diffusion signal representation is defined in the qτ space, so that it captures both spatial and temporal phenomena. The experiments on synthetic data and in vivo diffusion images of the C57Bl6 wild-type mouse corpus callosum reveal superiority of the proposed approach over random sampling and even distribution in the qτ space.

This work has been published in [12].

6.1.2. Dmipy, a Diffusion Microstructure Imaging toolbox in Python to improve research reproducibility

Participants: Abib Olushola Yessouffou Alimi, Rutger Fick [TheraPanacea, Paris], Demian Wassermann [Inria Parietal], Rachid Deriche.

The recovery of microstructure-related features of the brain's white matter is a current challenge in diffusion MRI (dMRI). In particular, multi-compartment (MC)-based models have been a popular approach to estimate these features. However, the usage of MC-models is often limited to those hard-coded in publicly available toolboxes.

In this work, we present Diffusion Microstructure Imaging in Python (Dmipy), a diffusion MRI toolbox which allows accessing any multi-compartment-based model and robustly estimates these important features from single-shell, multi-shell, and multi-diffusion time, and multi-TE data. Dmipy follows a *building block*-based philosophy to microstructure imaging, meaning an MC-model can be constructed and fitted to dMRI data using any combination of underlying tissue models, axon dispersion or diameter distributions, and optimization algorithms.using less than 10 lines of code, thus helps improve research reproducibility. In describing the toolbox, we show how Dmipy enables to easily design microstructure models and offers to the users the freedom to choose among different optimization strategies.We finally present three advanced examples of highly complex modeling approaches which are made easy using Dmipy.

This work has been published in [21], [30].

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

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

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

This work has been published in [11].

6.1.4. *Resolving the crossing/kissing fiber ambiguity using functionallyCOMMIT (Convex Optimization Modeling for Microstructure Informed Tractography)*

Participants: Matteo Frigo, Isa Costantini, Samuel Deslauriers-Gauthier, Rachid Deriche.

The architecture of the white matter is endowed with kissing and crossing bundles configurations. When these white matter tracts are reconstructed using diffusion MRI tractography, this systematically induces the reconstruction of many fiber tracts that are not coherent with the structure of the brain. The question on how to discriminate between true positive connections and false positive connections is the one addressed in this work. State-of-the-art techniques provide a partial solution to this problem by considering anatomical priors in the false positives detection process. We propose a novel model that tackles the same issue but takes into account both structural and functional information by combining them in a convex optimization problem. We validate it on two toy phantoms that reproduce the kissing and the crossing bundles configurations, showing that, through this approach, we are able to correctly distinguish true positives and false positives.

This work has been published in [25].

6.1.5. *Reducing false positive connection in tractograms using joint structure-function filtering*

Participants: Matteo Frigo, Guillermo Gallardo Diez, Isa Costantini, Alessandro Daducci [EPFL, Lausanne], Demian Wassermann [Inria Parietal], Samuel Deslauriers-Gauthier, Rachid Deriche.

Due to its ill-posed nature, tractography generates a significant number of false positive connections between brain regions. To reduce the number of false positives, Daducci et al. proposed the COMMIT framework, which has the goal of re-establishing the link between tractography and tissue microstructure. In this framework, the diffusion MRI signal is modeled as a linear combination of local models associated with streamlines where the weights are identified by solving a convex optimization problem. Streamlines with a weight of zero do not contribute to the diffusion MRI data and are assumed to be false positives. Removing these false positives yields a subset of streamlines supporting the anatomical data. However, COMMIT does not make use of the link between structure and function and thus weights all bundles equally. In this work, we propose a new strategy that enhances the COMMIT framework by injecting the functional information provided by functional MRI. The result is an enhanced tractogram filtering strategy that considers both functional and structural data.

This work has been published in [31].

6.1.6. *Combining Improved Euler and Runge-Kutta 4th order for Tractography in Diffusion-Weighted MRI*

Participants: Cherifi Dalila [IEEE University of Boumerdes, Algeria], Boudjada Messaoud [IEEE University of Boumerdes, Algeria], Morsli Abdelatif [IEEE University of Boumerdes, Algeria], Girard Gabriel [EPFL, Lausanne], Rachid Deriche.

In this work, we develop a general, deterministic tractography algorithm (CIERTE), which is a combination of Improved Euler and Range-Kutta fourth-order algorithm and test it on synthetic and real data. The proposed tractography method is validated using seven metrics of the tractometer evaluation system and positively compared to state-of-the-art tractography algorithms.

This work has been published in [9].

6.1.7. *Fiber orientation distribution function from non-negative sparse recovery with quantitative analysis of local fiber orientations and tractography using DW-MRI datasets*

Participants: Thinhinane Megherbi [USTHB, Algiers], Gabriel Girard [EPFL, Lausanne], Ghosh Aurobrata [AI Innovation Lab, Verisk Analytics], Fatima Oulebsir-Boumghar [USTHB, Algiers], Rachid Deriche.

In this work, we propose, evaluate and validate a new Diffusion Weighted MRI method to model and recover high quality tractograms even with multiple fiber populations in a voxel and from a limited number of acquisitions.

Our method relies on the estimation of the Fiber Orientation Distribution (FOD) function, parameterized as a non-negative sum of rank-1 tensors and the use of a non-negative sparse recovery scheme to efficiently recover the tensors, and their number. Each fiber population of a voxel is characterized by the orientation and the weight of a rank-1 tensor.

Using both deterministic and probabilistic tractography algorithms, we show that our method is able to accurately reconstruct narrow crossing fibers and obtain a high quality connectivity reconstruction even from a limited number of acquisitions. To this end, a validation scheme based on the connectivity recovered from tractography is developed to quantitatively evaluate and analyze the performance of our method. The tractometer tool is used to quantify the tractography obtained from a simulated DW-MRI dataset including a high angular resolution dataset of 60 gradient directions and a dataset of 30 gradient directions, each of them corrupted with Rician noise of SNR 10 and 20. The performance of our FOD model and its impact on the tractography results are also demonstrated and illustrated on in vivo DW-MRI datasets with high and low angular resolutions.

This work has been published in [15].

6.1.8. *Solving the Cross-Subject Parcel Matching Problem Using Optimal Transport*

Participants: Guillermo Gallardo Diez, Nathalie Gayraud, Maureen Clerc, Demian Wassermann [Inria Parietal], Samuel Deslauriers-Gauthier, Rachid Deriche.

Matching structural parcels across different subjects is an open problem in neuroscience. Even when produced by the same technique, parcellations tend to differ in the number, shape, and spatial localization of parcels across subjects. In this work, we propose a parcel matching method based on Optimal Transport. We test its performance by matching parcels of the Desikan atlas, parcels based on a functional criteria and structural parcels. We compare our technique against three other ways to match parcels which are based on the Euclidean distance, the cosine similarity, and the Kullback-Leibler divergence. Our results show that our method achieves the highest number of correct matches.

This work has been published in [32], [26].

6.1.9. *A Closed-Form Solution of Rotation Invariant Spherical Harmonic Features in Diffusion MRI*

Participants: Mauro Zucchelli, Samuel Deslauriers-Gauthier, Rachid Deriche.

Rotation invariant features are an indispensable tool for characterizing diffusion Magnetic Resonance Imaging (MRI) and in particular for brain tissue microstructure estimation. In this work, we propose a new mathematical framework for efficiently calculating a complete set of such invariants from any spherical function. Specifically, our method is based on the spherical harmonics series expansion of a given function of any order and can be applied directly to the resulting coefficients by performing a simple integral operation analytically. This enable us to derive a general closed-form equation for the invariants. We test our invariants on the diffusion MRI fiber orientation distribution function obtained from the diffusion signal both in-vivo and in synthetic data. Results show how it is possible to use these invariants for characterizing the white matter using a small but complete set of features.

This work has been published in [29].

6.1.10. Rational invariants of ternary forms under the orthogonal group

Participants: Paul Görlach [MPI for Mathematics in the Sciences], Evelyne Hubert [Inria, AROMATH], Théodore Papadopoulo, Rachid Deriche.

In [68], [69], [81] we started to explore the theory of tensor invariants as a mathematical framework for computing new biomarkers for HARDI. We pursued this work and, in collaboration with the project-team GALAAD/AROMATH, we succeeded to develop a complete set of rational invariants for ternary quartics [39]. Being rational, they are very close to the polynomial invariants developed in [69] but they constitute a complete set of invariants. They are also good tools to understand better the algebraic invariants of [81] and some others based on spherical harmonics decomposition [55]. We determined a generating set of rational invariants of minimal cardinality for the action of the orthogonal group $O(3)$ on the space $R[x,y,z]_{2d}$ of ternary forms of even degree 2d. The construction relies on two key ingredients. On one hand, the Slice Lemma allows us to reduce the problem to dermining the invariants for the action on a subspace of the finite subgroup $B(3)$ of signed permutations. On the other hand, our construction relies in a fundamental way on specific bases of harmonic polynomials. These bases provide maps with prescribed $B(3)$-equivariance properties. Our explicit construction of these bases should be relevant well beyond the scope of this work. The expression of the $B(3)$-invariants can then be given in a compact form as the composition of two equivariant maps. Instead of providing (cumbersome) explicit expressions for the $O(3)$-invariants, we provide efficient algorithms for their evaluation and rewriting. We also use the constructed $B(3)$-invariants to determine the $O(3)$-orbit locus and provide an algorithm for the inverse problem of finding an element in $R[x,y,z]_{2d}$ with prescribed values for its invariants. These are the computational issues relevant in brain imaging.

This work has been sumitted and is currently under review. A preprint is available in [39].

6.1.11. Edema-informed anatomically constrained particle filter tractography

Participants: Samuel Deslauriers-Gauthier, Drew Parker [UPenn, USA], François Rheault [SCIL, Sherbrooke University, CA], Steven Brem [UPenn, USA], Maxime Descoteaux [SCIL, Sherbrooke University, CA], Ragini Verma [UPenn, USA], Rachid Deriche.

In this work, we propose an edema-informed anatomically constrained tractography paradigm that enables reconstructing larger spatial extent of white matter bundles as well as increased cortical coverage in the presence of edema. These improvements will help surgeons to maximize the extent of the resection while minimizing the risk of cognitive deficits. The new paradigm is based on a segmentation of the brain into gray matter, white matter, corticospinal fluid, edema and tumor regions which utilizes a tumor growth model. Using this segmentation, a valid tracking domain is generated and, in combination with anatomically constrained particle filter tractography, allows streamlines to cross the edema region and reach the cortex. Using subjects with brain tumors, we show that our edema-informed anatomically constrained tractography paradigm increases the cortico-cortical connections that cross edema-contaminated regions when compared to traditional fractional anisotropy thresholded tracking.

This work has been published in [24].

6.1.12. Towards the assessment of myelination using time-dependent diffusion MRI indices

Participants: Abib Olushola Yessouffou Alimi, Alexandra Petiet [ICM, CENIR, Paris], Mathieu Santin [ICM, CENIR, Paris], Anne-Charlotte Philippe [ICM, CENIR, Paris], Stéphane Lehericy [ICM, CENIR, Paris], Demian Wassermann [Inria Parietal], Rachid Deriche.

In this work, we study the sensitivity of time-dependent diffusion MRI indices or qτ-indices to demyelination in the mouse brain. For this, we acquire in vivo four-dimentional diffusion-weighted images-varying over gradient strength, direction and diffusion time-and estimate the qτ-indices from the corpus callosum. First order Taylor approximation of each index gives fitting coefficients α and β whose variance we investigate. Results indicate that, cuprizone intoxication affects mainly index coefficient β by introducing inequality of variances between the two mice groups, most significantly in the splenium and that MSD increases and RTOP decreases over diffusion time τ.

This work has been published in [35].

6.1.13. An Analytical Fiber ODF Reconstruction in 3D Polarized Light Imaging

Participants: Abib Olushola Yessouffou Alimi, Yves Usson [UMR5525 TIMC-IMAG CNRS], Pierre-Simon Jouk [CHU Grenoble-Alpes], Gabrielle Michalowicz [CHU Grenoble-Alpes], Rachid Deriche.

Three dimensional polarized light imaging (3D-PLI) utilizes the birefringence in postmortem tissue to map its spatial fiber structure at a submillimeter resolution. In this work, we propose an analytical method to compute the fiber orientation distribution function (ODF) from high-resolution vector data provided by 3D-PLI. This strategy enables the bridging of high resolution 3D-PLI to diffusion magnetic resonance imaging with relatively low spatial resolution. First, the fiber ODF is modeled as a sum of K orientations on the unit sphere and expanded with a high order spherical harmonics series. Then, the coefficients of the spherical harmonics are derived directly with the spherical Fourier transform. We quantitatively validate the accuracy of the reconstruction against synthetic data and show that we can recover complex fiber configurations in the human heart at different scales.

This work has been published in [22].

6.1.14. fMRI Deconvolution via Temporal Regularization using a LASSO model and the LARS algorithm

Participants: Isa Costantini, Patryk Filipiak, Kostiantyn Maksymenko, Samuel Deslauriers-Gauthier, Rachid Deriche.

In the context of functional MRI (fMRI), methods based on the deconvolution of the blood oxygenated level dependent (BOLD) signal have been developed to investigate the brain activity, without a need of a priori knowledge about activations occurrence. In this work, we propose a novel temporal regularized deconvolution of the BOLD signal using the Least Absolute Shrinkage and Selection Operator (LASSO) model, solved by means of the Least-Angle Regression (LARS) algorithm. In this way, we were able to recover the underlying neurons activations and their dynamics.

This work has been published in [23], [37].

6.1.15. A Second Order Multi-Stencil Fast Marching Method with a Non-Constant Local Cost Model

Participants: Susana Merino-Caviedes [Universidad de Valladolid], Lucilio Cordero-Grande [King's College London], Maria Tereza Perez [Universidad de Valladolid], Pablo Casaseca-de-La-Higuera [Universidad de Valladolid], Marcos Martín-Fernández [Universidad de Valladolid], Carlos Alberola-Lopez [Universidad de Valladolid], Rachid Deriche.

The Fast Marching method is widely employed in several fields of image processing. Some years ago a Multi-Stencil version (MSFM) was introduced to improve its accuracy by solving the Eikonal equation for a set of stencils and choosing the best solution at each considered node. The following work proposes a modified numerical scheme for MSFM to take into account the variation of the local cost, which has proven to be second order. The influence of the stencil set choice on the algorithm outcome with respect to stencil orthogonality and axis swapping is also explored, where stencils are taken from neighborhoods of varying radius. The experimental results show that the proposed schemes improve the accuracy of their original counterparts, and that the use of permutation-invariant stencil sets provides robustness against shifted vector coordinates in the stencil set.

This work has been published in [16].

6.2. Unveiling brain activity using M/EEG

6.2.1. Data-driven cortical clustering to provide a family of plausible solutions to the M/EEG inverse problem

Participants: Maureen Clerc, Kostiantyn Maksymenko, Théodore Papadopoulo.

The Magneto/Electroencephalography (M/EEG) inverse problem consists in reconstructing cortical activity from M/EEG measurements. It is an ill-posed problem. Hence prior hypotheses are needed to constrain the solution space. In this work, we consider that the brain activity which generates the M/EEG signals is supported by single or multiple connected cortical regions. As opposed to methods based on convex optimization, which are forced to select one possible solution, we propose a cortical clustering based approach, which is able to find several candidate regions. These regions are different in term of their sizes and/or positions but fit the data with similar accuracy. We first show that even under the hypothesis of a single active region, several source configurations can similarly explain the data. We then use a multiple signal classification (MUSIC) approach to recover multiple active regions with our method. We validate our method on simulated and measured MEG data. Our results show that our method provides a family of plausible solutions which both accord with the priors and similarly fit the measurements.

This work is published in [41].

6.2.2. Fast approximation of EEG forward problem and application to tissue conductivity estimation

Participants: Maureen Clerc, Kostiantyn Maksymenko, Théodore Papadopoulo.

Bioelectric source analysis in the human brain from scalp electroencephalography (EEG) signals is sensitive to the conductivity of the different head tissues. Conductivity values are subject dependent, so non-invasive methods for conductivity estimation are necessary to suitably tune the EEG models. To do so, the EEG forward problem solution (so-called lead field matrix) must be computed for a large number of conductivity configurations. Computing one lead field requires a matrix inversion which is computationally intensive for realistic head models. Thus, the required time for computing a large number of lead fields can become impractical. In this work, we propose to approximate the lead field matrix for a set of conductivity configurations, using the exact solution only for a small set of basis points in the conductivity space. Our approach accelerates the computing time, while controlling the approximation error. Our method is tested for brain and skull conductivity estimation, with simulated and measured EEG data, corresponding to evoked somato-sensory potentials. This test demonstrates that the used approximation does not introduce any bias and runs significantly faster than if exact lead field were to be computed.

This work has been submitted to a journal and is available as a preprint [40].

6.2.3. Model based optimal multipolar stimulation without a priori knowledge of nerve structure

Participants: Maureen Clerc, Mélissa Dali [Inria Camin], David Guiraud [Inria Camin], Jérémy Laforêt [Inria Camin], Olivier Rossel [Inria Camin].

Multipolar cuff electrode can selectively stimulate areas of peripheral nerves and therefore enable to control independent functions. However, the branching and fascicularization are known for a limited set of nerves and the specific organization remains subject-dependent. This work presents general modeling and optimization methods in the context of multipolar stimulation using a cuff electrode without a priori knowledge of the nerve structure. Vagus nerve stimulation experiments based on the optimization results were then investigated.

The model consisted of two independent components: a lead field matrix representing the transfer function from the applied current to the extracellular voltage present on the nodes of Ranvier along each axon, and a linear activation model. The optimization process consisted in finding the best current repartition (ratios) to reach activation of a targeted area depending on three criteria: selectivity, efficiency and robustness.

The results showed that state-of-the-art configurations (tripolar transverse, tripolar longitudinal) were part of the optimized solutions but new ones could emerge depending on the trade-off between the three criteria and the targeted area. Besides, the choice of appropriate current ratios was more important than the choice of the stimulation amplitude for a stimulation without a priori knowledge of the nerve structure. We successfully assessed the solutions in vivo to selectively induce a decrease in cardiac rhythm through vagus nerve stimulation while limiting side effects. Compared to the standard whole ring configuration, a selective solution found by simulation provided on average 2.6 less adverse effects.

The preliminary results showed the correctness of the simulation, using a generic nerve geometry. It suggested that this approach will have broader applications that would benefit from multicontact cuff electrodes to elicit selective responses. In the context of the vagus nerve stimulation for heart failure therapy, we show that the simulation results were confirmed and improved the therapy while decreasing the side effects.

This work has been published in [10].

6.3. Combined M/EEG and dMRI

6.3.1. Linking resting-state functional connectivity and the structural connectome – investigation of an eigen-structure model

Participants: Rebecca Bonham-Carter, Samuel Deslauriers-Gauthier, Rachid Deriche.

Resting-state functional connectivity (rs-FC) dynamics are not random but rather structured with common dominant patterns called resting-state networks (RSNs). These dynamics are influenced by the underlying network of white-matter connections. Specifically, temporal correlations in resting-state BOLD fMRI signals have been correlated with the structural network determined via diffusion weighted imaging (DWI). The literature on this structure-function relationship encompasses generative non-linear models and a variety of linear models. The objective of this study is to provide new validation and understanding of two linear models. Both models enforce that the structural network Laplacian and rs-FC share a common eigen-structure. In contrast to previous work, in this work two linear models of resting-state functional connectivity (rs-FC), developed by Abdelnour et al., are validated on simulated BOLD fMRI data generated using The Virtual Brain18 (TVB) and 49 HCP subjects real structural connectomes. Both consider rs-FC as a diffusion process on the structural network. The mean correlations between rs-FC matrices we obtain 0.699 ± 0.086 and 0.518 ± 0.095, and between rs-FC eigenvalues 0.981 ± 0.013, agree with the original model implementations on empirical data. Using The Virtual Brain simulator together with real structural data is shown to offer a new and efficient test and validation framework for approaches predicting rs-FC from structure.

This work is under review.

6.3.2. White Matter Information Flow Mapping from Diffusion MRI and EEG

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

The human brain can be described as a network of specialized and spatially distributed regions. The activity of individual regions can be estimated using electroencephalography and the structure of the network can be measured using diffusion magnetic resonance imaging. However, the communication between the different cortical regions occurring through the white matter, coined information flow, cannot be observed by either modalities independently. Here, we present a new method to infer information flow in the white matter of the brain from joint diffusion MRI and EEG measurements. This is made possible by the millisecond resolution of EEG which makes the transfer of information from one region to another observable. A subject specific Bayesian network is built which captures the possible interactions between brain regions at different times. This network encodes the connections between brain regions detected using diffusion MRI tractography derived white matter bundles and their associated delays. By injecting the EEG measurements as evidence into this model, we are able to estimate the directed dynamical functional connectivity whose delays are supported by the diffusion MRI derived structural connectivity. We present our results in the form of information flow diagrams that trace transient communication between cortical regions over a functional data window. The performance of our algorithm under different noise levels is assessed using receiver operating characteristic curves on simulated data. In addition, using the well-characterized visual motor network as grounds to test our model, we present the information flow obtained during a reaching task following left or right visual stimuli. These promising results present the transfer of information from the eyes to the primary motor cortex. The information flow obtained using our technique can also be projected back to the anatomy and animated to produce videos of the information path through the white matter, opening a new window into multi-modal dynamic brain connectivity.

This work is under review.

6.3.3. Bridging Brain Structure and Function by Correlating Structural Connectivity and Cortico-Cortical Transmission

Participants: Fabien Almairac [CHU Nice], Patryk Filipiak, Lavinia Slabu, Maureen Clerc, Théodore Papadopoulo, Denys Fontaine [CHU Nice], Lydiane Mondot [CHU Nice], Stéphan Chanalet [CHU Nice], Demian Wassermann [Inria Parietal], Rachid Deriche.

Elucidating the structure-function relationship of the brain is one of the main open questions in neuroscience. The capabilities of diffusion MRI-based (dMRI) techniques to quantify the connectivity strength between brain areas, namely structural connectivity, in combination with modalities such as electrocorticography (ECoG) to quantify brain function have enabled advances in this field. In this work, we aim to establish a relationship between: i) dMRI structural connectivity measures, ii) direct measures of electrical properties of the human brain cortex obtained with ECoG, iii) response elicited by direct electrostimulation of the brain (DES).

The results of this multi-modal approach combining structure and function explorations of the brain should: i) help to elucidate the relationship between non-invasive (dMRI) structural connectivity measures and cortico-cortical transmission properties (delays, transfer functions), ii) help in understanding the organization of the brain for cognitive functions as well as neurosurgical planning for resection of brain tumors and drug-resistant epilepsy

This work has been presented in [36].

6.4. Brain Computer Interfaces

6.4.1. Online enhancement of visuospatial attention performance

Participants: Maureen Clerc, Thomas Brochier [Institut des Neurosciences de la Timone], Romain Trachel.

This study on real-time decoding of visuospatial attention has two objectives: first, to reliably decode self-directed shifts of attention from electroencephalography (EEG) data, and second, to analyze whether this information can be used to enhance visuospatial performance. Visuospatial performance was measured in a target orientation discrimination task, in terms of reaction time, and error rate. Our experiment extends the Posner paradigm by introducing a new type of ambiguous cues to indicate upcoming target location. The cues are designed so that their ambiguity is imperceptible to the user. This entails endogenous shifts of attention

which are truly self-directed. Two protocols were implemented to exploit the decoding of attention shifts. The first 'adaptive' protocol uses the decoded locus to display the target. In the second 'warning' protocol, the target position is defined in advance, but a warning is flashed when the target mismatches the decoded locus. Both protocols were tested in an online experiment involving ten subjects. The reaction time improved in both the adaptive and the warning protocol. The error rate was improved in the adaptive protocol only. This proof of concept study brings evidence that visuospatial brain–computer interfaces (BCIs) can be used to enhance improving human–machine interaction in situations where humans must react to off-center events in the visual field.

This work has been published in [8].

6.4.2. Review of classification methods for EEG-based Brain-Computer Interfaces: A 10-year update

Participants: Maureen Clerc, Laurent Bougrain [Neurosys, Inria Nancy], Fabien Lotte [Potioc, Inria Bordeaux], Alain Rakotomamonjy [Université de Rouen].

Most current Electroencephalography (EEG)-based Brain-Computer Interfaces (BCIs) are based on machine learning algorithms. There is a large diversity of classifier types that are used in this field, as described in the 2007 review paper [75]. Now, approximately 10 years after this review publication, many new algorithms have been developed and tested to classify EEG signals in BCIs. The time is therefore ripe for an updated review of EEG classification algorithms for BCIs. We surveyed the BCI and machine learning literature from 2007 to 2017 to identify the new classification approaches that have been investigated to design BCIs. We synthesize these studies in order to present such algorithms, to report how they were used for BCIs, what were the outcomes, and to identify their pros and cons. We found that the recently designed classification algorithms for EEG-based BCIs can be divided into four main categories: adaptive classifiers, matrix and tensor classifiers, transfer learning and deep learning, plus a few other miscellaneous classifiers. Among these, adaptive classifiers were demonstrated to be generally superior to static ones, even with unsupervised adaptation. Transfer learning can also prove useful although the benefits of transfer learning remain unpredictable. Riemannian geometry-based methods have reached state-of-the-art performances on multiple BCI problems and deserve to be explored more thoroughly, along with tensor-based methods. Shrinkage linear discriminant analysis and random forests also appear particularly useful for small training samples settings. On the other hand, deep learning methods have not yet shown convincing improvement over state-of-the-art BCI methods. This paper provides a comprehensive overview of the modern classification algorithms used in EEG-based BCIs, presents the principles of these Review of Classification Algorithms for EEG-based BCI 2 methods and guidelines on when and how to use them. It also identifies a number of challenges to further advance EEG classification in BCI.

This work has been published in [14].

6.4.3. Automatizing calibration

Participants: Maureen Clerc, Federica Turi, Nathalie Gayraud.

Brain Computer Interfaces (BCIs) based on visual evoked potentials (VEP) allow for spelling from a keyboard of flashing characters. Among VEP BCIs, code-modulated visual evoked potentials (c-VEPs) are designed for high-speed communication. In c-VEPs, all characters flash simultaneously. In particular, each character flashes according to a predefined 63-bit binary sequence (m-sequence), circular-shifted by a different time lag. For a given character, the m-sequence evokes a VEP in the electroencephalogram (EEG) of the subject, which can be used as a template. This template is obtained during a calibration phase at the beginning of each session. Then, the system outputs the desired character after a predefined number of repetitions by estimating its time lag with respect to the template. Our work avoids the calibration phase, by extracting from the VEP relative lags between successive characters, and predicting the full word using a dictionary.

This work has been published in [28].

BIOVISION Project-Team

6. New Results

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

6.1.1. Improving social interaction through augmented reality

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

Today's visual enhancement systems for low-vision people consist of dedicated augmented reality hardware allowing to magnify or enhance the overall scene, independently of the image content or patient needs. For example, for patients with central vision loss, interacting with others may become a painful activity when faces and expressions can hardly be recognized. In [17], we introduce a new augmented reality system allowing to selectively enhance faces, using two image processing techniques [44], [50]. Our system is based on a Fove 0 head-mounted display (FOVE Inc, San Mateo, CA, USA). It has the capacity to adjust the enhancement to the detected faces' size and distance, hence maintaining a constant boost in the critical range of spatial frequency. It offers a binocular and large Field-of-View and performs at near real-time with a modest laptop computer using multithreading. Preliminary experiments with three patients with central vision loss suggest that the enhancements chosen strongly depends on each patient's condition and lead to improved recognition abilities when patients find their optimal settings.

This work is presented in [17].

6.1.2. Text auto-illustration for improving reading accessibility to low-vision people

Participants: Paula Pawlowski, Pierre Kornprobst, Elena Cabrio [Inria, EPI WIMMICS], Marco Benzi [Université Côte d'Azur (France)].

We have started to explore how to make reading more efficient and more enjoyable for low-vision patients through text auto-illustration. Text auto-illustration consists in automatically extracting images from the web which are related to a given text, using natural language processing methods.

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

6.2.1. Retinal waves

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

Retinal waves are bursts of activity occurring spontaneously in the developing retina of vertebrate species, contributing to the shaping of the visual system organization: retina circuitry shaping, retinotopy, eye segregation [51], [36], [46], [37]. They stop a few weeks after birth. Wave activity begins in the early development, long before the retina is responsive to light. It was recently found that they can be reinitiated pharmacologically in the adult mammalian retina [35]. This could have deep consequences on therapy for several degenerative retinal diseases. The mechanism of their generation, in developing, or adult retinas, remains however incompletely understood [52].

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

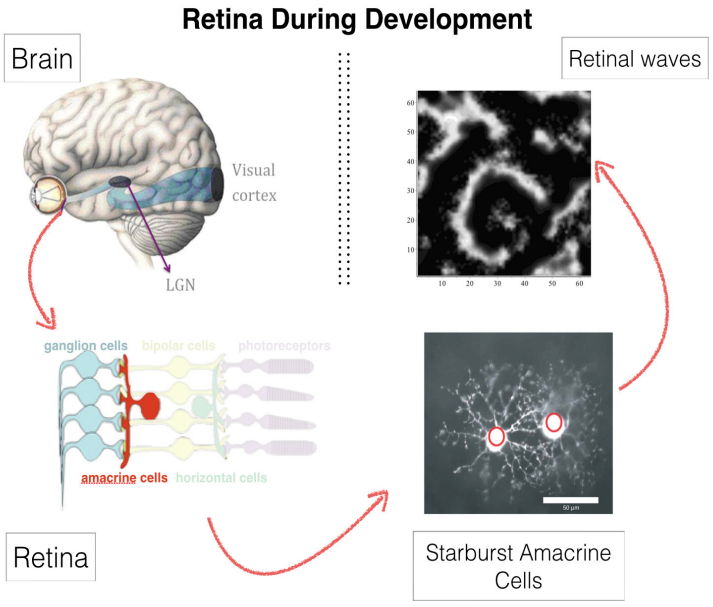


Figure 1. Multiscale dynamics of retinal waves

This work has been presented in [11], [13], [14], [16], [20], [12], [24].

6.2.2. Trajectory anticipation, from retina to V1

Participants: Bruno Cessac, Selma Souihel, Frédéric Chavane, Alain Destexhe, Matteo Di Volo, Olivier Marre.

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

at multiple scales (from individual cells to large networks) and multiple stages (retina, primary visual cortex V1).

In the context of the ANR Trajectory, we are working on such an integrated approach. We aim at modelling the population responses at two key stages of visual motion encoding: the retina and V1 based on simultaneous micro- and mesoscopic recordings made by our partners Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France), Institut de la Vision (Paris, France) and Unité de Neurosciences Information et Complexité, Gif sur Yvette, France. We are designing a simulator of retinal output + V1, reproducing both the retinal anticipation and the cortical response measured by optical imaging. We are also analyzing the effects of lateral connectivity in the retina, via amacrine cells, in processing motion. This lateral connectivity is accountable for the correlated activity of RGCs in experimental data. We are measuring these correlations to add further biological plausibility to our model. This study is a step toward understanding mechanisms of motion coding and anticipation with strong impact on our understanding of the visual system.

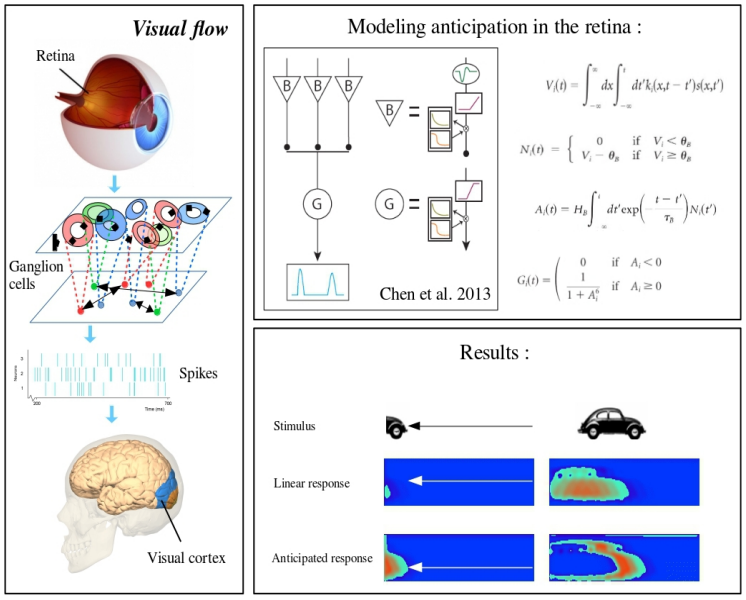


Figure 2. Motion anticipation

These results have been presented in [26], [27]

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

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

Retinal spike response to stimuli is constrained, on one hand by short range correlations (receptive field overlap) and on the other hand by lateral connectivity (cells connectivity). This last effect is difficult to handle from statistics because it requires to consider spatio-temporal correlations with a time delay long enough to take into account the time of propagation along synapses. Although MaxEnt models are useful

to fit optimal model (maximizing entropy) under the constraints of reproducing observed correlations, they do address spatio-temporal correlations in their classical form (Ising or higher order interactions but without time delay). Binning in such models somewhat integrates propagation effects, but in an implicit form, and increasing binning severely bias data. To resolve this issue we have considered spatio-temporal MaxEnt model formerly developed e.g. by Vasquez et al. [49]. The price to pay, however is a huge set of parameters that must be fitted to experimental data to explain the observed spiking patterns statistics. There is no a priori knowledge of which parameters are relevant and which ones are contributing to overfitting. We propose here a method of dimension reduction, i.e. a projection on a relevant subset of parameters, relying on the so-called Susceptibility matrix closely related to the Fisher information. In contrast to standard methods in information geometry though, this matrix handles space and time correlations. We have applied this method for retina data obtained in a diurnal rodent (Octodon degus, having 30% of cones photoreceptors) and a 252-MEA system. Three types of stimuli were used: spatio-temporal uniform light, white noise and a natural movie. We show the role played by time-delayed pairwise interactions in the neural response to stimuli both for close and distant cells. Our conclusion is that, to explain the population spiking statistics we need both short-distance interactions as well as long-distance interactions, meaning that the relevant functional correlations are mediated not only by common input (i.e. receptive field overlap, electrical coupling; spillover) but also by long range connections.

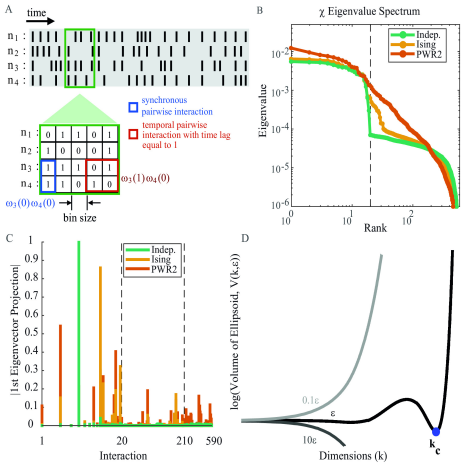


Figure 3. Dimension reduction method

This work has been submitted to Plos Comp Bio [22].

6.2.4. Linear response for spiking neuronal networks with unbounded memory

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

The activity of a neuronal network, characterized by action potentials (spikes), is constrained by the intrinsic properties of neurons and their interactions. When a neuronal network is submitted to external stimuli, the statistics of spikes changes, and it is difficult to disentangle the influence of the stimuli from the intrinsic dynamics. We have established a general linear response relation for spiking neuronal networks, based on

chains with unbounded memory. This relation allows quantifying the influence of a weak amplitude external stimuli on spatio-temporal spike correlations, in a general context where the memory in spike dynamics can go arbitrarily far in the past. With this approach, we show how linear response is explicitly related to neuron dynamics with an example, the gIF model, introduced by M. Rudolph and A. Destexhe [91]. This illustrates the effect of the stimuli, intrinsic neuronal dynamics, and network connectivity on spike statistics.

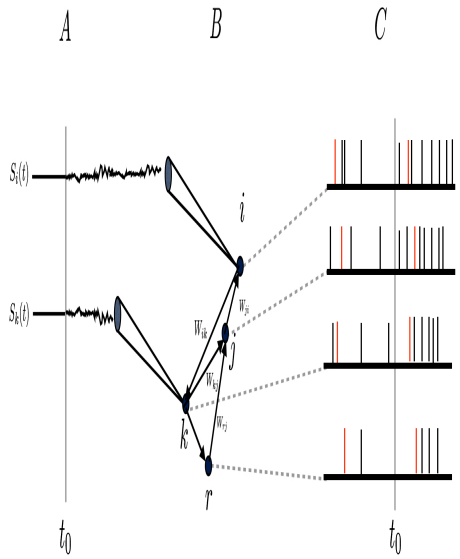


Figure 4. Linear response. The excitation of some neurons in a network will induce a variation in spike correlations between all connected neurons, even those which are not excited. We have computed this effect in [19].

This work has been submitted to Journal of Mathematical Neurosciences [19].

CAMIN Team

6. New Results

6.1. A sensor fusion approach for inertial sensors based 3D kinematics and pathological gait assessments

Participants: Benoît Sijobert, Christine Azevedo, Jérôme Froger, Francois Feuvrier.

Pathological gait assessment and assistive control based on functional electrical stimulation (FES) in post-stroke individuals, brings out a common need to robustly quantify kinematics facing multiple constraints.

Through an experimental study(Figure 5), we proposed a novel approach using inertial sensors to compute dorsiflexion angles and spatio-temporal parameters, in order to be later used as inputs for online close-loop control of FES. 26 post-stroke subjects were asked to walk on a pressure mat (GaitRite®) equipped with inertial measurement units (IMU) and passive reflective markers (Vicon®). A total of 930 strides were individually analyzed and results between IMU-based algorithms and reference systems compared. The novel methods integrated two aspects: 1) robust stance phase detection based on acceleration and angular rate combination and 2) estimation of joint angles based on an Attitude and Heading Reference System (non linear observer) algorithm and gravity cancellation for reconstructing 3D trajectory of individual steps. Mean absolute (MA) errors of dorsiflexion angles were found to be less than 4°, while stride lengths were robustly segmented and estimated with a MA error less than 10 cm [30]. These results open new doors to rehabilitation using adaptive FES closed-loop control strategies.

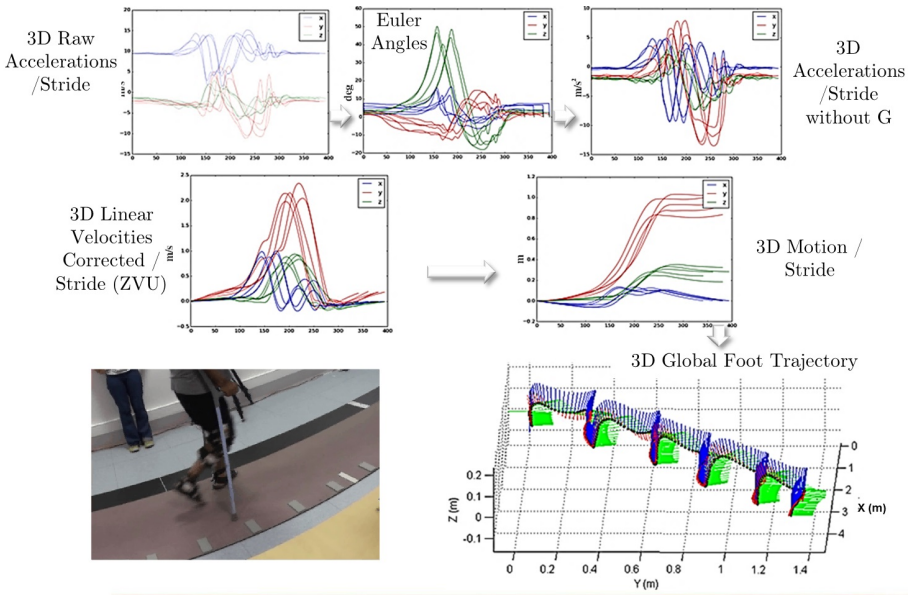


Figure 5. Illustration of the global approach needed to obtain gait trajectories from raw magneto-inertial data of a foot-mounted Inertial Measurement Unit to assess a circumduction gait.

6.2. FES-based online closed-loop control of knee joint to reduce stance phase asymmetry in post-stroke gait

Participants: Benoît Sijobert, Christine Azevedo, Charles Fattal.

Numerous stimulation strategies have been investigated over the past thirty years to assist or restore gait (Figure 6). The studies using FES to restore gait have been mostly conducted in post-stroke individuals and focused on correcting the drop foot syndrome by supplementing the absence of dorsiflexion. The state-of-the-art reflects a real lack of interest in using FES to improve the paretic knee rehabilitation, which however plays a key role in post-stroke gait recovery.
An experimental protocol (#RCB 2017-A03611-52, CRF La Chataigneraie, Menucourt, France) was designed with the main purpose of proposing a novel approach using a FES-based control of knee joint to reduce stance phase asymmetry and study the feasibility of using such FES systems in clinical rehabilitation, compared to classical knee orthosis. Secondary objectives aimed at improving gait quality, walking range and comfortable speed using the same modality. The main hypothesis was to determine if using FES to real-time control the paretic knee angle could reduce the time needed to recover a normal balance while providing a secure stance phase. To monitor weight bearing and stance time asymmetry the participants were equipped with instrumented insoles. The subjects were also equipped with 2 inertial measurement units located on the thigh and the tibia, wired to a Raspberry Pi3. Each IMU embedded a high speed based processor and a Kalman Filter directly providing quaternion estimation needed to compute knee angles. One IMU was installed in the back of the participants at the second sacral vertebra level to estimate vertical trunk displacement. Stimulation was sent via a two-channel wireless stimulator to the quadriceps and hamstrings via surface electrodes. A specific hardware architecture has been developed for this protocol. When required and depending on the participant's gait pattern, a "pre-stance" stimulation could be triggered either via an online detection of peak knee flexion or when the sagittal angular speed recorded via the gyroscope crossed zero. In stance phase, stimulation was triggered either to quadriceps or hamstrings, depending on the paretic knee angle (PKA) estimation relatively to the knee angle setpoint (KAS) defined by the practitioner as the optimal flexion during stance phase (around 5^{O}). A proportional (P) controller adjusted the pulse width depending on the error ϵ between PKA and KAS. Equipped with their usual technical aids (cane, ankle foot orthosis...) participants were asked to perform a 10m-path walking at a self-selected pace. An oral instruction was given at the beginning of each trial to encourage the participants to transfer their weight onto the paretic leg. This experiment is an ongoing experiment but preliminary tests have already validated the technical feasibility of the approach.

6.3. IMU-based FES cycling in individuals with SCI

Participants: Benoît Sijobert, Christine Azevedo, Ronan Le Guillou, Charles Fattal, Emerson Fachin Martin, Henrique Resende.

It has been shown that FES-cycling of subjects with Spinal Cord Injuries (SCI) results in physiological and psychological positive effects such as cardiovascular training, decrease in pressure sores occurrence and self-esteem improvements. However, the use of this technology has often remained restricted to indoor and stationary ergometers in clinical contexts, partly due to the small amount (10–25 W) of power produced and the requirement of experimented users to finely tuned the stimulation patterns needed to stimulate lower limb muscles with an adequate modality. In order to promote the research around this topic and more broadly the development of assistive technology for people with physical disabilities, we participated to the first Cybathlon in October 2016 (FreeWheels project), using a stimulation pattern based on crank angle. Taking part to this event highlighted the need for a simpler automated stimulation pattern generator, able to adapt the stimulation to the environment, to the muscle fatigue or to the individual (e.g. position on the bike, number of stimulable muscles, etc...). In order to further investigate control solutions, we first needed to be able to accurately quantify the influence of each parameter preliminarily used (stimulation pattern, stimulation parameters, fixed-wheel or free-wheel, individualized quadriceps, pilot position, etc...) on power produced and endurance and observe if other variables could be used as an input instead of the crank angle. The decision was made to develop an instrumented home trainer specifically designed to record a weak power (<200 W) while ensuring

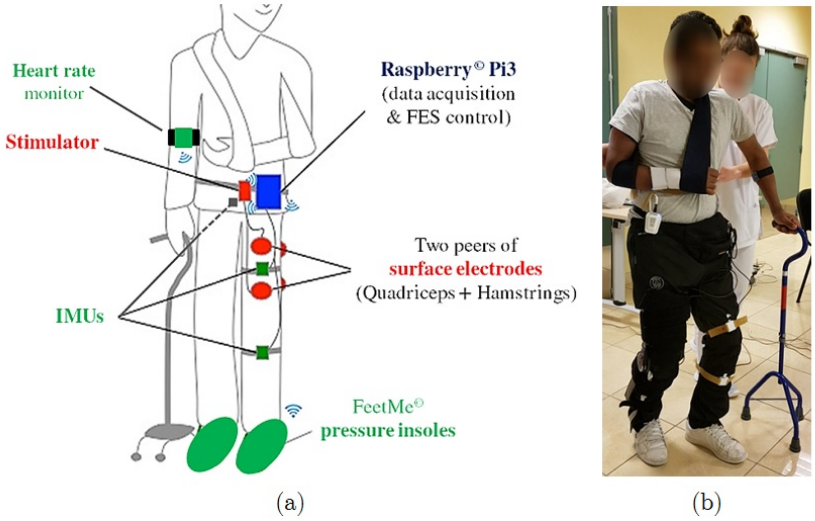


Figure 6. Experimental setup diagram (a) and picture (b). The participants are equipped with Bluetooth pressure insoles, 2 wired IMU on the leg and 1 wireless IMU in the back. A Raspberry records and processes the sensors data to send an appropriate command to a wireless stimulator to stimulate the quadriceps and hamstrings via surface electrodes

a minimum accuracy of 0.5 %: a rotating torquemeter (Scaime TSR 2300) was installed between the rear wheel and a flywheel thanks to a mechanical assembly built in collaboration with the National Engineering School of Saint-Étienne (ENISE, Loire, France)(Figure 7). The software part was developed as part of ADT STIMBIO.

Instead of using the crank angle, undergoing researches have also investigated the ability of using inertial sensors to automatically design a stimulation pattern on the bike depending on the knee angles. Based on the joint angle computation presented in section 6.1 and experimentally validated, a similar control modality have been studied and implemented. Using the online peak knee flexion algorithm developed in the study presented in section 6.2 to continuously detect this event, we developed a novel approach in order to trigger the quadriceps stimulation at the beginning of the pushing phase. This would enable to take into account a possible sliding in seat position without requiring an accurate placement of the IMUs or a geometrical model of the individual. A study has been initiated with the University of Brasilia (UnB, District Federal, Brazil) as part of the CACAO collaboration, to explore advanced control approaches [31]. Experimental data have been recorded and are investigated in order to compare the different control approaches (Figure 8).

6.4. Respiratory detection and monitoring

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

This work is conducted within a CIFRE phd thesis. The general subject is the respiration induced by implanted stimulation for the tetraplegic and syndrome of Ondine. In France, every year, there is approximately 90 new spinal cord injuries who have a ventilatory dependence due to a high cervical involvement. The prevalence of syndrome of Ondine (central sleep apnea) would be 25.5 per million inhabitants. Because of many disadvantages of mechanical ventilation, the technique of implanted electrical stimulation to restore the respiratory function of the patients can be proposed. But existing systems are based on open-loop controllers,

Figure 7. Experimental setup using IMUs to record joint angles and a specifically designed home trainer to monitor power output.

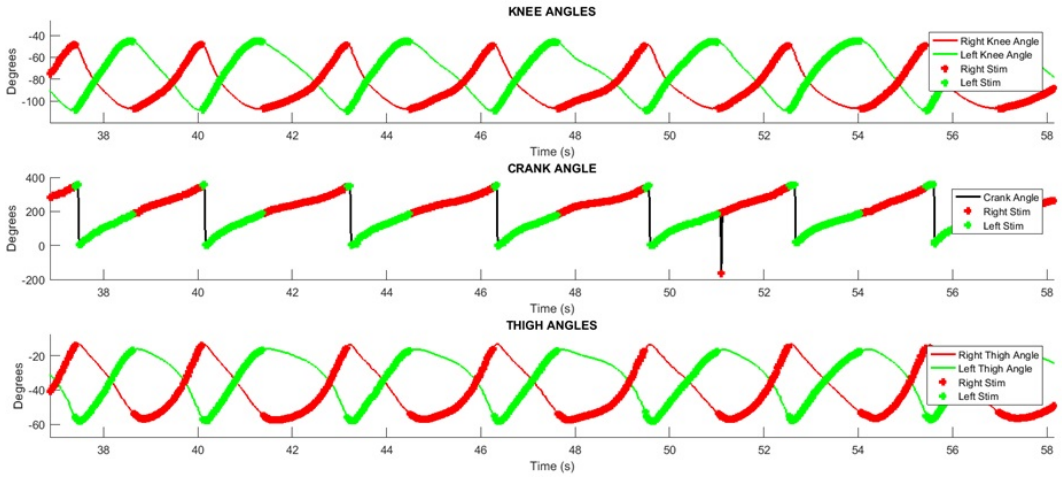


Figure 8. Comparison of stimulation patterns between crank angle and joint angle based triggering method.

ie the phrenic nerve is stimulated with the same intensity, at the same frequency for the whole time, even when the patients can breathe spontaneously. The principle aim of the work is to develop a respiratory detection/monitoring module in this context.

A solution based on tracheal sounds analysis has been developed. Tracheal sounds are recorded by microphone which is inserted into a support and stuck on the neck of subject like showed in Figure 9 .i. All the materials are showed in Figure 9 .ii: microphones (yellow), analog ampli-filtering card (red), the acquisition machine POWERLAB (green), the numeric development card NUCLEO (blue).

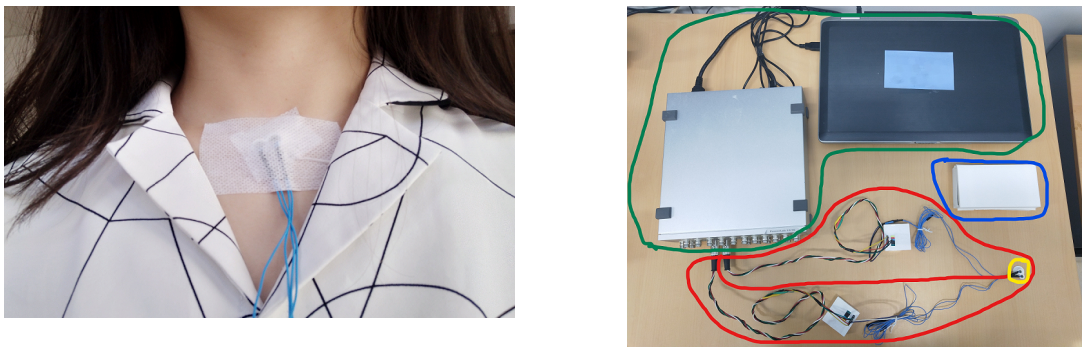


Figure 9. i)The placement of microphone ii)Recording materials

The signal is processed in its envelop (temporal domain) and frequency power (frequency domain). A threshold detection applied to detect respiration. An example of detection result is illustrated in Figure 10 . Heart beating sounds can also be extracted to calculate cardiac rhythm.

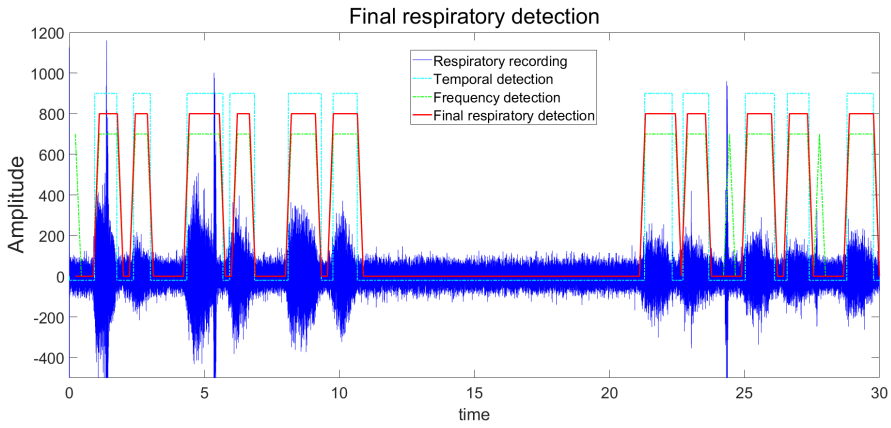


Figure 10. The detection result

Preliminary recordings on healthy individuals have been performed. Recordings on patients are in planning. Publications for conferences and journals are in preparation. The variation of cardiac amplitude will also be analyzed to give a secondary breathing detection. Advance signal processing techniques are now under study.

6.5. Attenuation and delay of remote potentials evoked by direct electrical stimulation during brain surgery.

Participants: Anthony Boyer, Sofiane Ramdani [LIRMM], Hugues Duffau [CHU Montpellier], David Guiraud, François Bonnetblanc.

Direct electrical stimulation (DES) is used during awake brain surgery for functional mapping as it generates transient behavioural disturbances, allowing the identification of both cortical areas and subcortical white matter pathways which are essential to the function. However, the electrophysiological effects of DES remain by far unknown. DES may be coupled with the measurement of Evoked Potentials (EPs) to study the conductive and integrative properties of activated neural ensembles and probe the spatiotemporal dynamics of short- and long- range networks. We recorded ECoG signals on two patients undergoing awake brain surgery and measured EPs on functional sites after cortical stimulations, using combinations of stimulation parameters (Figure 11). We were more particularly interested in the generation of evoked potentials (EPs) triggered by both close and remote stimulations. Obtained EPs were very similar in shape, suggesting a stereotyped electrophysiological response, but delayed in time and attenuated in amplitude when elicited from a different gyrus or remotely from the recording site. We were also able to observe the bidirectional nature of the arcuate fasciculus triggering EPs on 2 anatomically connected sites. We propose different activation and electrophysiological propagation mechanisms following DES based on recruited neural elements. The variations in amplitude and delay of EPs are most likely due to different propagation mechanisms, which can be intra- or sub- cortical, and correspond to commonly described DCRs and CCEPs.

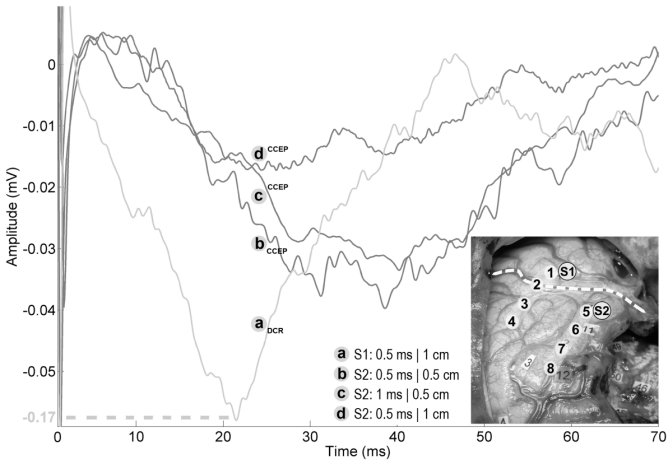


Figure 11. Differential recordings between electrodes 2 and 3 while stimulating S1 and S2. The picture illustrates the stimulation sites (S1, S2) and ECoG positioning with respect to the initial 60 Hz cortical brain mapping. Experimental DES was applied on: (1) the Wernicke's area (S1), associated with complete anomia; (2) the ventral premotor cortex (S2), which led to movement and counting interruptions. Tumor was about 164 cm3. The Sylvian fissure is highlighted by a thick dashed line.

6.6. High Frequency stimulation used for efficient and fiber type selective stimulation

Participants: David Guiraud, Mélissa Dali, Olivier Rossel, Thomas Guiho, Pawel Maciejasz.

In neural electrical stimulation, limiting the charge delivered during a stimulus pulse is essential to avoid nerve tissue damage and to save power. Previous experimental and modeling studies indicated that waveforms such as nonrectangular continuous pulses or rectangular chopped pulse were able to improve stimulation efficiency. The goal of this study is to evaluate if non-rectangular chopped pulses such as quarter sine and ramp are more charge efficient than rectangular chopped pulse. We performed in vivo study on 17 lumbricus terrestris and compared the charge per stimulating phase needed to activate lateral giant fibers (LGF) and medial giant fiber (MGF) using chopped non-rectangular pulses and rectangular pulse, varying stimulation duration parameters. Results indicated that non rectangular chopped pulses activated MGF and LGF with less charge than rectangular chopped pulses. For MGF (respectively LGF), the gain of charge was up to 33.9% (resp. 17.8%) using chopped ramp, and up to 22.8% (resp. 18.1%) using chopped quarter sine.

6.7. Early detection of stroke during the accute phase

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

Cerebral infarctions can now be treated with new techniques using intravenous thrombolysis and thrombectomy. Their proven efficacy is directly correlated to the time lapse between the start of symptoms and the initiation of treatment. Currently, a definitive diagnosis can only be made once the patient has performed a radiological imaging (CT scan or MRI) on a medical center equipped with these expensive devices, thus enabling the medical team to initiate the appropriate treatment. Transit times during the pre-hospitalization phase before diagnosis are therefore often longer and have the greatest negative impact on the patient's prognosis. In collaboration with the interventional neuroradiology department of Gui de Chauliac Hospital, I2FH and Euromov, the EleVANT project is aiming to prospectively evaluate new techniques to assess a diagnosis of acute cerebral ischaemia. This low cost technology could be used in a mobile way for the very early diagnosis of cerebral infarction and thus reduce treatment delays, opening the way to a new generation of diagnostic tools. The concept consist on evaluating the cerebral near-infrared spectroscopy (NIRS) response to different stimulus, and to evaluate its lateralization. Recently, we tested our device on healthy volunteers. Method: Left and right hemisphere reactivity index are recorded by NIRS and normalized (Figure 12). Result: The experiments present a suitable feasibility and repeatability. In healthy subjects, a good response to the stimulus is recorded, and no significant differences between hemispheres are observed. The confidence level is acceptable since the amplitude response in above the standard deviation level.

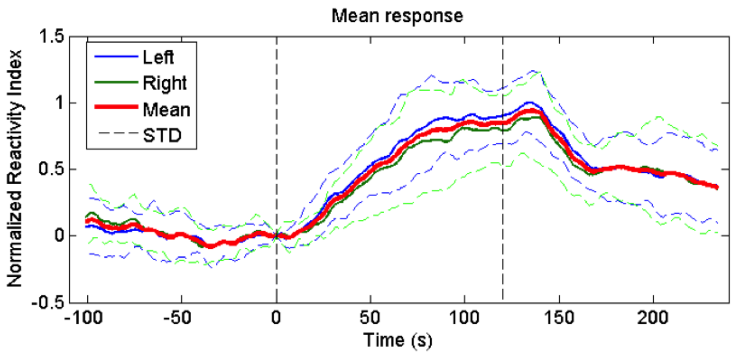


Figure 12. NIRS reactivity Index in response to a stimulus (bounded by the dashed lines)

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

6.8. Real-time simulation of stimulation systems

Participant: Daniel Simon.

RT-STIM (Real-time FES simulation) is a C/C++ framework able to carry out realistic simulations of a fully featured functional electro-stimulation system. It allows for the temporally consistent co-simulation of both the continuous model of skeletal joints and muscles on one hand, and of numerical resources such as control tasks, schedulers and communication protocols on the other hand (Figure 13). Initial software-in-the-loop simulations can be seamlessly extended towards hardware-in-the-loop simulation by a progressive integration of real components such as a Raspberry portable control board or gateways towards Vivaltis stimulators and HiKoB sensors.

It is intended to be a support for the design and implementation of safe stimulation feedback controllers in the team. Hence, the simulation software is designed around the bio-mechanical models of joints and muscles excited using electro-stimulation developed in the Demar and Camin teams during the past years. To cope with the objectives of the team which targets the restoration of grasping for tetraplegia, a model of a human hand, currently using 23 joints and 23 muscles, has been integrated. It is expected to be a central tool for the Agilis project starting in 2019.

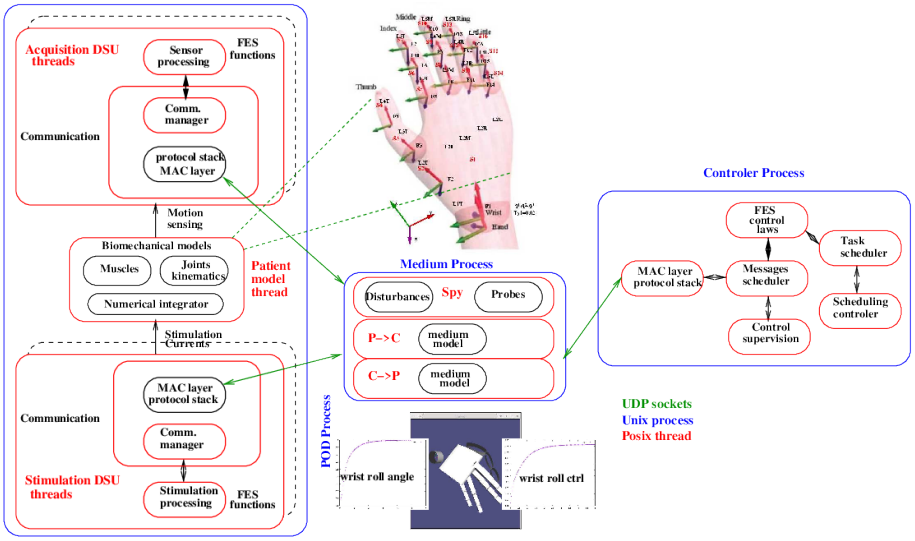


Figure 13. Simulation of a hand under FES

6.9. Real-time control for distributed stimulation systems

Participants: Daniel Simon, Ashwini Patil, Ronan Le Guillou.

Feedback control is needed to control complex movement, such as precise grasping, involving several muscles and nerves. Moreover, the components of the control loops (i.e., sensors, electrodes and micro-controllers) are distributed over communication links which induce data delivery scheduling, delays, and occasional data loss. A new approach gathering control and computation related design and implementation constraints was developed during the past years. From the feedback provided by experiments in real-time robot control,

Considering the non-linear and time varying models, simple controllers cannot provide reliable and robust solutions. the approach is developed along two main directions, control aware computing, e.g., using feedback schedulers, and real-time aware control, e.g. using feedback controllers designed to be robust and/or adaptive w.r.t. timing deviations [32].

Beyond simple control loops, a Model Predictive Control approach for FES using an adaptive horizon is under evaluation. Even if no conclusion about the control approach can be currently carried out, it was an occasion to positively evaluate the Julia programming language as an alternative to others high level languages for control design and real-time implementation.

A software control structure, primarily implemented and evaluated on a portable Rasperry3 micro-controller, is currently documented to become the root of a generic real-time control software template, usable even by non-specialists in the Camin team.

EPIONE Project-Team

6. New Results

6.1. Medical Image Analysis

6.1.1. Learning a Probabilistic Model for Diffeomorphic Registration

Participants: Julian Krebs [Correspondant], Hervé Delingette, Tommaso Mansi [Siemens Healthineers, Princeton, NJ, USA], Nicholas Ayache.

This work is funded by Siemens Healthineers, Princeton, NJ, USA

deformable registration, probabilistic modeling, deep learning, latent variable model, deformation transport, disease clustering

We developed a probabilistic approach for deformable image registration in 3-D using deep learning methods [30]. This method includes:

- A probabilistic formulation of the registration problem through unsupervised learning of an encoded deformation model (Fig. 4).
- A differentiable exponentiation layer and an user-adjustable smoothness layer that ensure the outputs of neural networks to be regular and diffeomorphic.
- An analysis of size and structure of a latent variable space for registration.
- Experiments on deformation transport and disease clustering.

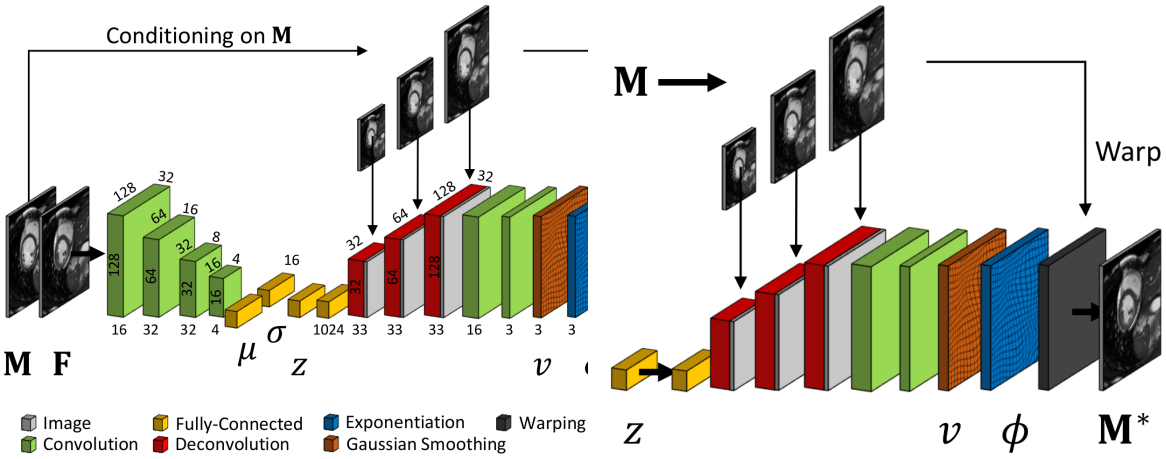


Figure 4. (Left) Probabilistic registration network including a diffeomorphic layer (exponentiation). Deformations are encoded in z from which velocities are decoded while being conditioned on the moving image. (Right) Decoder network for sampling and deformation transport: Apply z-code conditioned on any new image M.

6.1.2. Learning Myelin Content in Multiple Sclerosis from Multimodal MRI

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

This work is done in collaboration with the Aramis-Project team of Inria in Paris and the researchers at the Brain and Spinal Cord Institute (ICM) located in Paris.

Multiple Sclerosis, MRI, PET, GANs

- We predict myelin content from multiparametric MRI [36].
- We design an adaptive loss and a sketch-refinement process for GANs, decomposing the problem into anatomy/physiology and myelin content prediction (Fig. 5).
- We show similar results to the PET-derived gold standard.

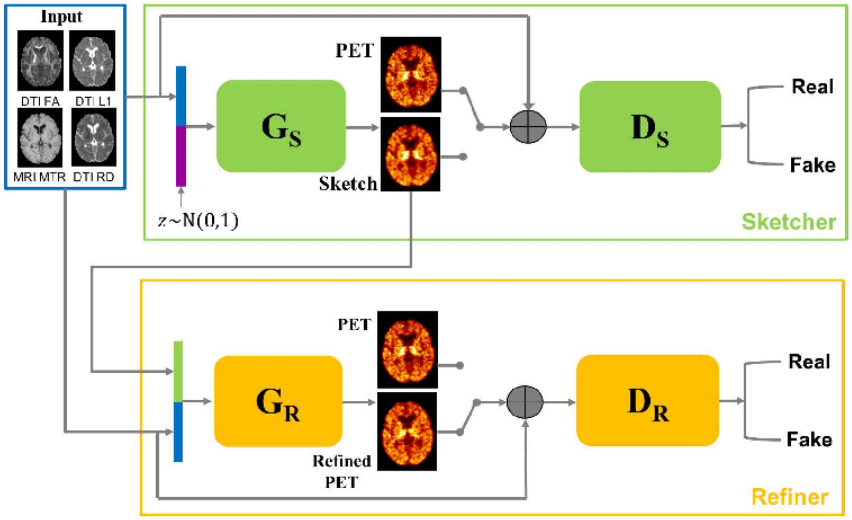


Figure 5. The sketcher receives MR images and generates the preliminary anatomy and physiology information. The refiner receives MR images IM and the sketch IS. Then it refines and generates PET images.

6.1.3. Consistent and Robust Segmentation of Cardiac Images with Propagation

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

This project is funded by European Research Council (MedYMA ERC-AdG-2011-291080).

Cardiac segmentation, deep learning, neural network, 3D consistency, spatial propagation

We propose a method based on deep learning to perform cardiac segmentation on short axis MRI image stacks iteratively from the top slice (around the base) to the bottom slice (around the apex) [26][62]. At each iteration, a novel variant of U-net is applied to propagate the segmentation of a slice to the adjacent slice below it (Fig. 6).

- 3D-consistency is hence explicitly enforced.
- Robustness and generalization ability to unseen cases are demonstrated.
- Results comparable or even better than the state-of-the-art are achieved.

The corresponding open source software, CardiacSegmentationPropagation, is available in https://team.inria.fr/epione/en/software/.

6.1.4. Deep Learning for Tumor Segmentation

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

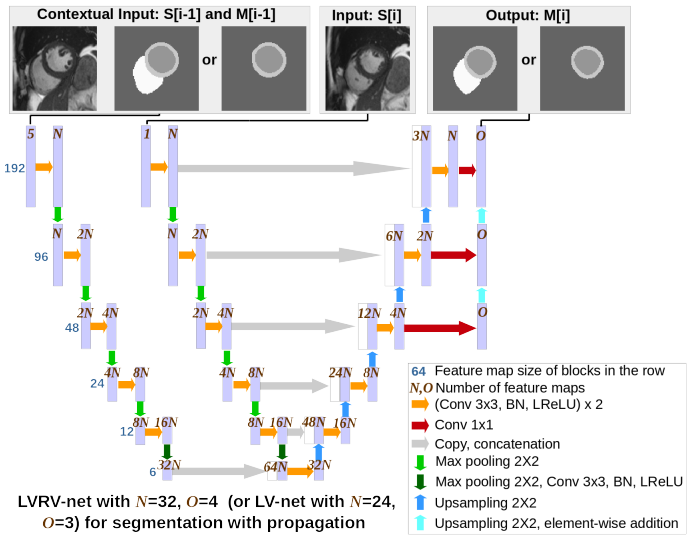


Figure 6. Propagation of cardiac segmentation by a neural network.

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

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

- We proposed a model for tumor segmentation which is able to analyze a very large spatial context by combining 2D and 3D CNNs [56] (Fig. 7). Top-3 performance was obtained on BRATS 2017 challenge.
- We proposed an approach to train CNNs for tumor segmentation with a mixed level of supervision [55]. Our approach significantly improves segmentation accuracy compared to standard supervised learning.
- We designed a system for segmentation of organs at risk for protontherapy. Promising preliminary results were obtained.

6.2. Imaging & Phenomics, Biostatistics

6.2.1. Radiomic analysis to improve diagnosis and therapy in oncology

Participants: Fanny Orlhac [Correspondant], Nicholas Ayache, Charles Bouveyron, Jacques Darcourt [CAL], Hervé Delingette, Olivier Humbert [CAL], Pierre-Alexandre Mattei [Copenhague University], Thierry Pourcher [CEA], Fanny Vandenbos [CHU Nice].

Inria postdoctoral fellowship for 16 months

Radiomics, Statistical learning, Metabolomics

- We proposed a modeling which extends the High-Dimensional Discriminant Analysis (HDDA) model by incorporating a sparsity pattern for each class, called sparse HDDA (sHDDA) [21].
- We demonstrated its efficacy in identifying lung lesions based on CT radiomic features (see Figure 8) [21] or triple-negative breast lesions from PET radiomic features and metabolomic data [34], [32], [33]. Thanks to the class-specific variable selection, the final model can be easily interpreted by physicians.
- We also demonstrated the capacity of the ComBat method to harmonize radiomic features extracted from PET images acquired with different imaging protocols [40], [39].

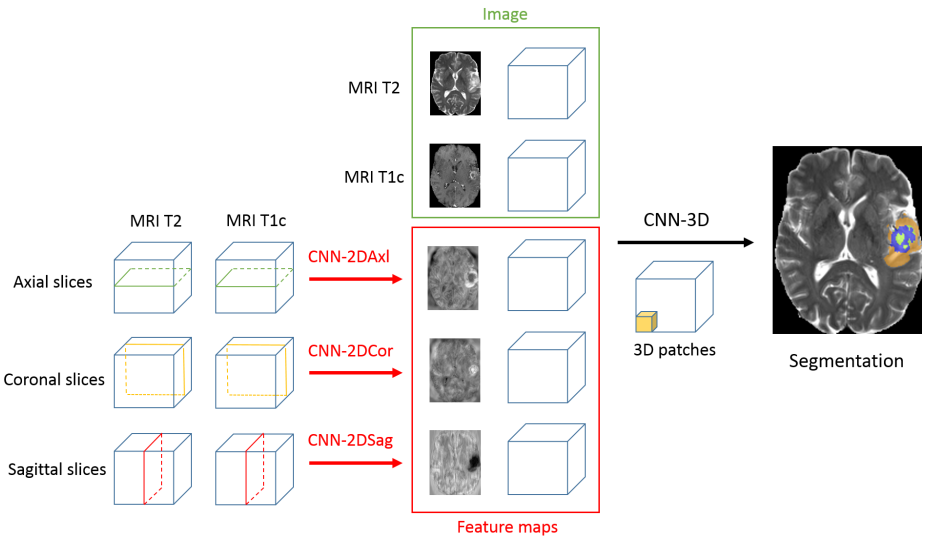


Figure 7. Illustration of our 2D-3D model for brain tumor segmentation.

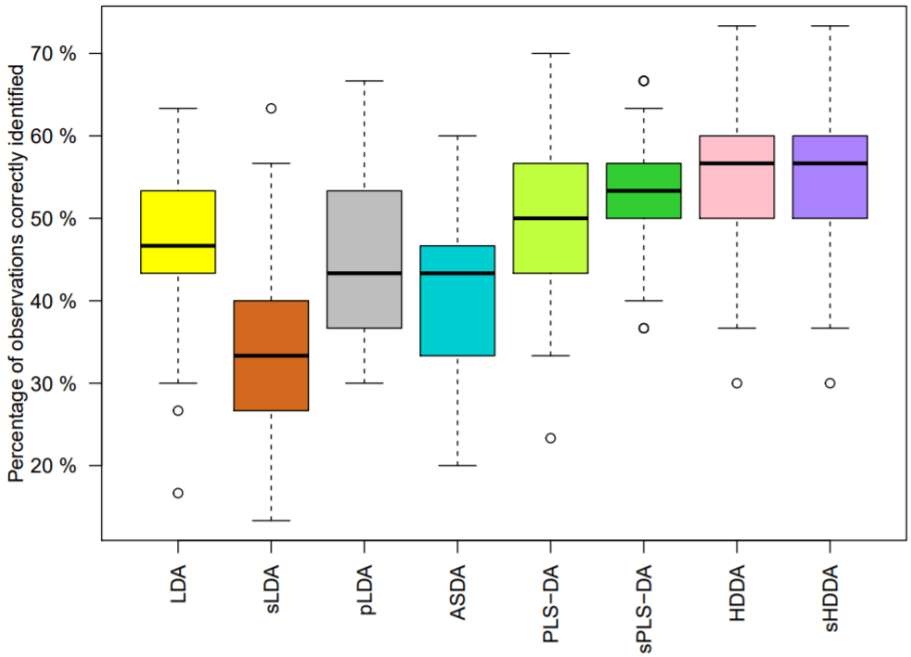


Figure 8. Classification accuracy of the eight statistical methods, including HDDA in pink and sHDDA in purple, to identifying lung lesions based on radiomic features extracted from CT images.

6.2.2. Statistical learning on large databases of heterogeneous imaging, cognitive and behavioral data

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

Supported by the French government, through the UCA^{JEDI} Investments in the Future project managed by the National Research Agency (ANR) ref. num. ANR-15-IDEX-01, our research is within the MNC3 initiative (Médecine Numérique: Cerveau, Cognition, Comportement), in collaboration with the Institut Claude Pompidou (CHU of Nice). Computational facilities are funded by the grant AAP Santé 06 2017-260 DGA-DSH, and by the Inria Sophia Antipolis - Méditerranée, "NEF" computation cluster.

statistical learning, joint analysis, neuroimaging

The aim of our work is to build scalable learning models for the joint analysis of heterogeneous biomedical data, to be applied to the investigation of neurological and neuropsychiatric disorders from collections of brain imaging, body sensors, biological and clinical data available in current large-scale databases such as ADNI[0] and local clinical cohorts.

We developed a computationally efficient formulation of probabilistic latent variable models [37]. This approach is capable to highlight meaningful relationships among biomarkers in the context of Alzheimer's disease (Figure 9) that can be used to develop optimal strategies for disease quantification and prediction.

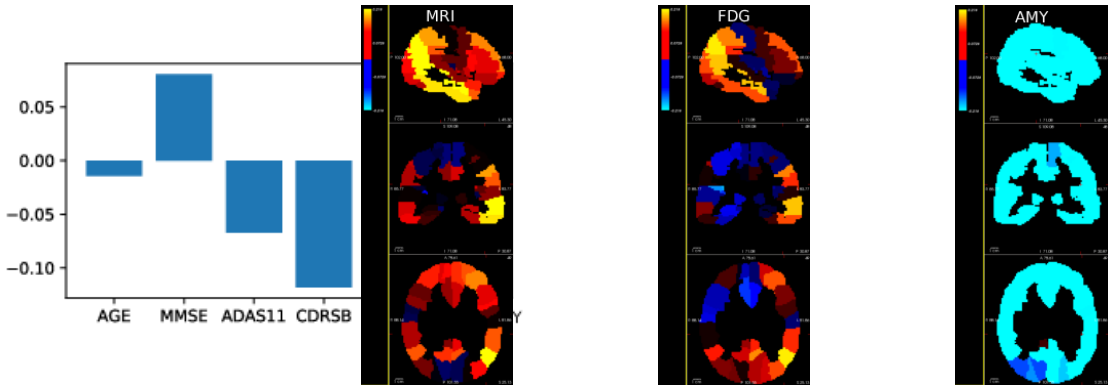


Figure 9. Joint relationships among Alzheimer's disease biomarkers discovered by our multi-channel model in the ADNI dataset. Relationships in line with the current literature discoveries. Clinical biomarkers on the left; brain imaging biomarkers on the right: gray matter density (MRI), glucose uptake (FDG), amyloid uptake (AMY).

6.2.3. Joint Biological & Imaging markers for the Diagnosis of severe lung diseases

Participants: Benoît Audelan [Correspondant], Hervé Delingette, Nicholas Ayache.

Lung cancer, Early detection, Sparse Bayesian Learning

Lung cancer is among the most common cancer and is considered to be one of the most important public health problem. The aim of this work is to improve the detection of lung cancer by combining imaging and biological markers. Exploratory analysis have been conducted to discriminate cancer patients versus controls from circulating miRNAs data using sparse Bayesian learning and to automatically pre-process lung CT images (Fig. 10).

[0] http://adni.loni.usc.edu/

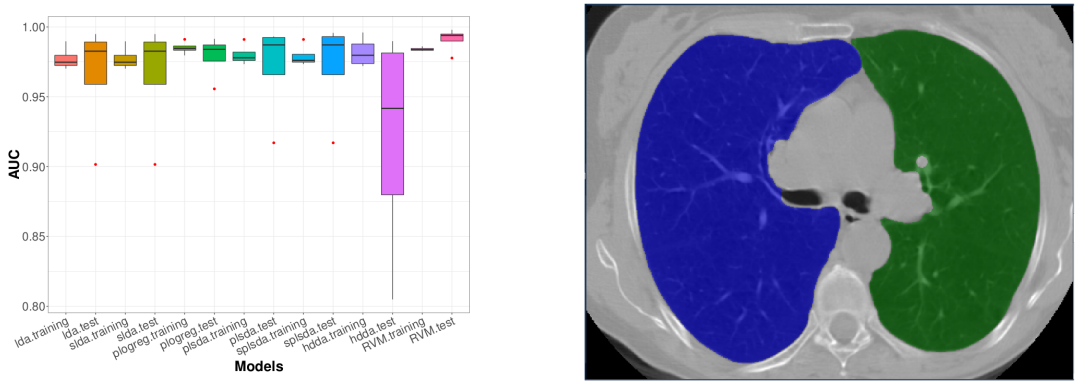


Figure 10. Comparison of statistical methods for classifying lung cancer patient from miRNAs data and lung segmentation

6.2.4. A data-driven model of mechanistic brain atrophy propagation in dementia

Participants: Sara Garbarino [Correspondant], Marco Lorenzi.

Sara Garbarino acknowledges financial support from the French government managed by L'Agence Nationale de la Recherche under Investissements d'Avenir UCA JEDI (ANR-15-IDEX-01) through the project "AtroPro-Dem: A data-driven model of mechanistic brain Atrophy Propagation in Dementia".

Gaussian Processes, Bayesian non–parametric modelling, neuroimaging data, protein dynamics, brain network

Models of misfolded proteins aim at discovering the bio-mechanical properties of neurological diseases by identifying plausible associated dynamical systems. Solving these systems along the full disease trajectory is usually challenging, due to the lack of a well defined time axis for the pathology. This issue is solved by disease progression models where long-term progression trajectories are estimated via time reparametrization of individual observations. However, due to their loose assumptions on the dynamics, they do not provide insights on the bio-mechanical properties of protein propagation.

In this project we propose a unified model of spatio-temporal protein dynamics based on the joint estimation of long-term protein dynamics and time reparameterization of individuals observations (Figure 11). The model is expressed within a Gaussian Process regression setting, where constraints on the dynamics are imposed through non–linear dynamical systems.

6.2.5. Federated Learning in Distributed Medical Databases: Meta-Analysis of Large-Scale Subcortical Brain Data

Participants: Santiago Silva [Correspondant], Marco Lorenzi, Boris Gutman, Andre Altman, Eduardo Romero, Paul M. Thompson.

This work was supported by the French government, through the UCAJEDI Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-15-IDEX-01 (project Meta-ImaGen).

Federated learning, distributed databases, PCA, SVD, meta-analysis, brain disease.

We proposed a federated learning framework for securely accessing and meta-analyzing any biomedical data without sharing individual information.

- A frontend pipeline for preprocessing and analyzing data was proposed, including: standardization, confounders correction, and variability analysis via federated PCA.

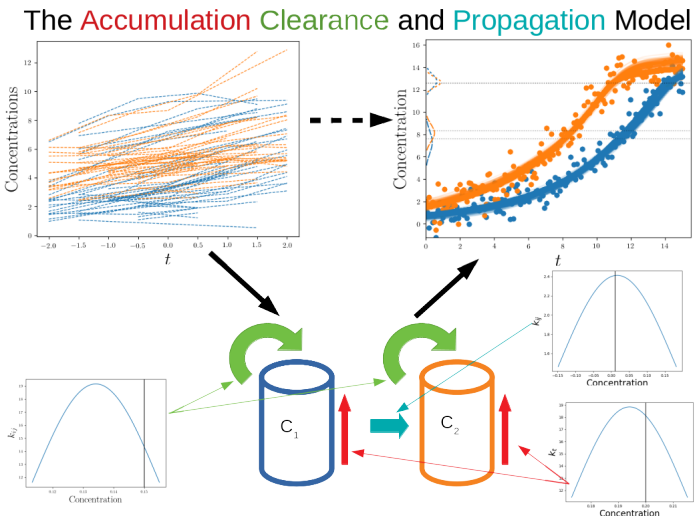


Figure 11. Schematic representation of our framework. Here we have two brain regions whose concentrations are collected for many subjects over a short term time span. The dynamics of such concentrations is described in terms of accumulation, clearance and propagation parameters. The proposed Bayesian framework estimates the distribution of such parameters and the long term trajectories with respect to the estimated disease time axis.

- Tested on multi-centric and multi-diagnosis databases (ADNI, PPMI and UK-Biobank) showed a clear differentiation between control and Alzheimer's subjects (Figure 12).
- Further developments of this study will extend the proposed analysis to large-scale imaging genetics data, such as in the context of the ENIGMA meta-study.

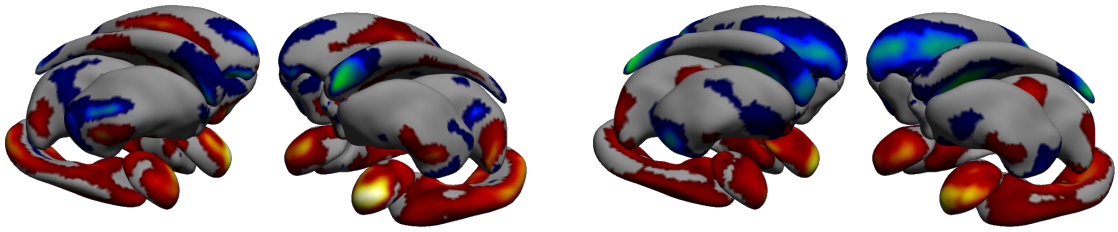


Figure 12. First principal component estimated with the proposed federated framework. The component maps prevalently hippocampi and amigdalae. ***Left:*** *Thickness.* ***Right:*** *Log-Jacobians.*

6.3. Computational Anatomy

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

Participants: Clement Abi Nader [Correspondant], Nicholas Ayache, Philippe Robert, Marco Lorenzi.

Gaussian Process, Alzheimer's Disease, Disease Progression Modelling

The aim of this project is to develop a spatio-temporal model of Alzheimer's Disease (AD) progession [47]. We assume that the brain progression is characterized by independent spatio-temporal sources that we want to separate. We estimate brain structures involved in the disease progression at different resolutions thus dealing with the non-stationarity of medical images, while assigning to each of them a monotonic temporal progression using monotonic Gaussian processes (Figure 13 , left-middle panel). We also compute an individual time-shift parameter to assess the disease stage of each subject (Figure 13 , right panel).

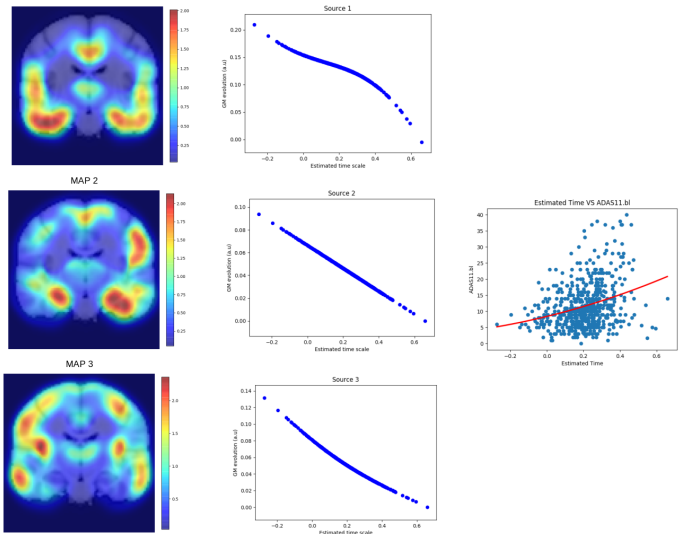


Figure 13. Left-Middle: Brain structures involved in AD along with their temporal evolution. Right: Correlation between ADAS11 cognitive score and the individual time-shift.

6.3.2. A model of brain morphological evolution

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

Longitudinal modeling, Deformation framework, Brain morphology, Alzheimer's disease, Aging.

We proposed a deformation-based generative model of the brain morphological evolution that can jointly describes the effect of aging and Alzheimer's disease. It relies on longitudinal description of the aging and disease consequences and can be use to compute image-based cross-sectional progression markers. This approach is able to propose a description of the disease evolution, population and subject-wise(see Figure 14) and open the way to a better modeling of the disease progression.

6.3.3. Geometric statistics

Participant: Xavier Pennec [Correspondant].

This work is partially funded by the ERC-Adv G-Statistics.

Statistics on manifolds, Differential geometry,

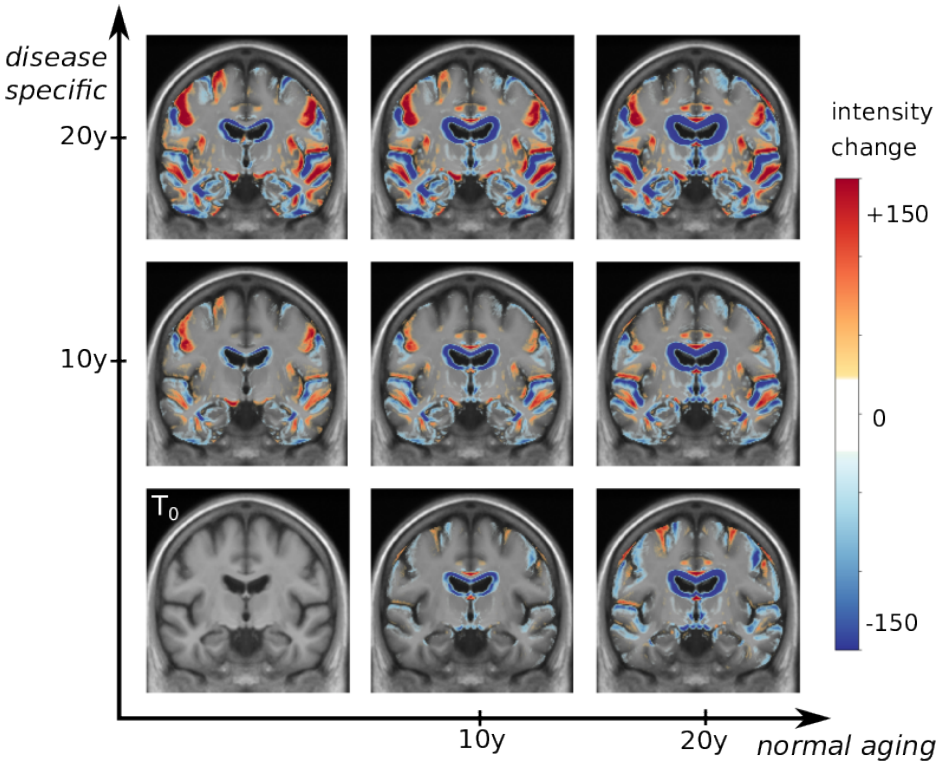


Figure 14. Representation of the 2D parametric template subspace generated by the model. In these images, the bottom row correspond to a healthy evolution, and the diagonal (from bottom left to top right) to a typical pathological evolution. The colors represent the voxel-wise intensity differences between the images and the reference T_0 to highlight the boundary shifts between tissues and CSF.

Beyond the mean value, Principal Component Analysis (PCA) is often used to describe the main modes of variability and to create low dimensional models of the data. Generalizing these tools to manifolds is a difficult problem. In order to define low dimensional parametric subspaces in manifolds, we proposed in [22] to use the locus of points that are weighted means of a number of reference points. These barycentric subspaces locally define submanifolds which can naturally be nested to provide a hierarchy of properly embedded subspaces of increasing dimension (a flag) approximating the data better and better. This defines a generalization of PCA to manifolds called Barycentric Subspace Analysis (BSA) which provides a new perspective for dimension reduction. It appears to be well suited for implicit manifolds such as the ones defined by multiple registrations in longitudinal or cross-section image analysis. An example of such an application was provided in [23] for 4D cardiac image sequences.

In classical estimation problems, the number of samples is always finite. The variability that this induces on the estimated empirical mean is a classical result of the law of large numbers in the asymptotic regime. In manifolds, it is not clear how the curvature influences the estimation of the empirical Fréchet mean with a finite number of samples. Preliminary results showed that there is an unexpected bias inversely proportional to the number of samples induced by the gradient of the curvature and a correction term of the same order on the covariance matrix slowering or accelerating the effective convergence rate towards the Fréchet mean of the underlying distribution. These preliminary results were derived using a new simple methodology that is also extending the validity from Riemannian manifolds to affine connection spaces [45].

6.3.4. Brain template as a Fréchet mean in quotient spaces

Participants: Nina Miolane [Correspondant], Xavier Pennec.

Computational anatomy, Morphological brain template, Hierarchical modeling.

Geometrically, the procedure used to construct the reference anatomy for normalizing the measurements of individual subject in neuroimaging studies can be summarized as the Fréchet mean of the images projected in a quotient space. We have previously shown that this procedure is asymptotically biased, therefore inconsistent. In [15], we presented a methodology that quantifies spatially the brain template's asymptotic bias. We identify the main variables controlling the inconsistency. This leads us to investigate the topology of the template's intensity levels sets, represented by its Morse-Smale complex. We have proposed a topologically constrained adaptation of the template computation that constructs a hierarchical template with bounded bias. We apply our method to the analysis of a brain template of 136 T1 weighted MR images from the Open Access Series of Imaging Studies (OASIS) database.

6.3.5. Cardiac Motion Evolution Modeling from Cross-Sectional Data using Tensor Decomposition

Participants: Kristin Mcleod [Simula Research Laboratory], Maxime Sermesant, Xavier Pennec.

Cardiac motion tracking, modeling cardiac motion evolution over time

Cardiac disease can reduce the ability of the ventricles to function well enough to sustain long-term pumping efficiency. We proposed in [14] a cardiac motion tracking method to study and model cohort effects related to age with respect to cardiac function. The proposed approach makes use of a Polyaffine model for describing cardiac motion of a given subject, which gives a compact parameterisation that reliably and accurately describes the cardiac motion across populations. Using this method, a data tensor of motion parameters is extracted for a given population. The partial least squares method for higher-order arrays is used to build a model to describe the motion parameters with respect to age, from which a model of motion given age is derived. Based on cross-sectional statistical analysis with the data tensor of each subject treated as an observation along time, the left ventricular motion over time of Tetralogy of Fallot patients is analysed to understand the temporal evolution of functional abnormalities in this population compared to healthy motion dynamics (see Figure 15).

6.3.6. Challenging cardiac shape and motion statistics

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

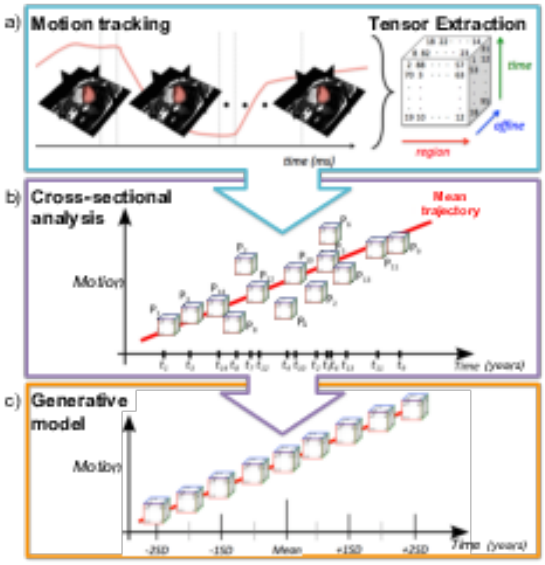


Figure 15. Building a generative model of the long-term motion changes in a population (c) by a) building a data tensor of polyaffine motion parameters that represent the motion over the cardiac cycle for each subject in the population using cross-sectional statistical analysis of polyaffine tensors and b) performing cross-sectional statistical analysis of the data tensors.

Shape statistics, Non-rigid registration, Deep learning, Cardiac shape and motion

Two of the methods previously developed by Marc-Michel Rohé in his PhD were benchmarked against other state of the art methods in two successive MICCAI challenges. First, the SVF-net developed to perform a very-fast inter-subject heart registration based on convolutional neural networks was embedded into a multi-atlas segmentation pipeline and tested against other deep learning techniques for the automatic MRI cardiac multi-structures segmentation [2]. Second, a combination of polyaffine cardiac motion tracking and supervised learning was used to predict myocardial infarction [25]. Both challenges demonstrate the good performances of the tested methods.

6.4. Computational Physiology

6.4.1. CIMPLE : Cochlear Implantation Modeling, PLanning & Evaluation

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

This work is funded by the Provence-Alpes-Côte-d'Azur region, the Université Côte d'Azur and Oticon Medical.

Statistic Learning, Image segmentation, Cochlea Implantation

The work aims to establish an effective, quantitative and rapid assessment method for human cochlear implantation planning and adjustment.

During last one year, our team explored a cost-effectively and practically algorithm to achieve the goal.

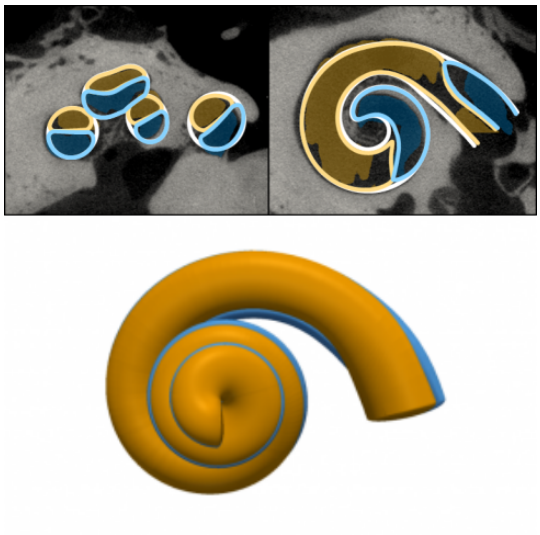


Figure 16. The figure shown a example segmentation on micro-CT image. Lower is the parametric model that quantitative mesure the cochea shape.

6.5. Computational Cardiology & Image-Based Cardiac Interventions

6.5.1. Population-based priors for group-wise Personalisation

Participants: Roch Molléro [Correspondant], Hervé Delingette, Xavier Pennec, Nicholas Ayache, Maxime Sermesant.

The authors acknowledge the partial funding by the MD-Paedigree EU Project.

Personalised cardiac model, Parameter observability, Statistical modeling, Dimensionality reduction, Heterogeneous clinical data, Imputation

Personalised cardiac models have a large number of parameters while the available data for a given patient is typically limited to a small set of measurements, thus the parameters cannot be estimated uniquely. This is a practical obstacle for clinical applications, where accurate parameter values can be important. Here we explore an original approach based on an algorithm called Iteratively Updated Priors (IUP), in which we perform successive personalisations of a full database through Maximum A Posteriori (MAP) estimation, where the prior probability at an iteration is set from the distribution of personalised parameters in the database at the previous iteration (Figure 17). At the convergence of the algorithm, estimated parameters of the population lie on a linear subspace of reduced (and possibly sufficient) dimension in which for each case of the database, there is a (possibly unique) parameter value for which the simulation fits the measurements. We first show how this property can help the modeler select a relevant parameter subspace for personalisation. In addition, since the resulting priors in this subspace represent the population statistics in this subspace, they can be used to perform consistent parameter estimation for cases where measurements are possibly different or missing in the database, which we illustrate with the personalisation of a heterogeneous database of 811 cases [18].

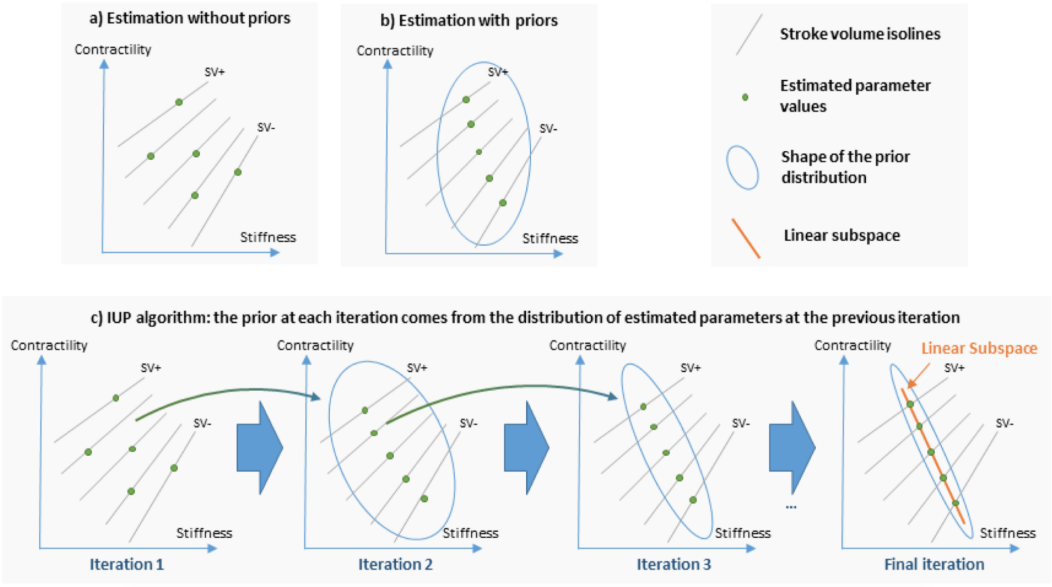


Figure 17. Schematic representation of parameter estimation problem: both contractility and stiffness are estimated from the stroke volume (SV). Both have an influence on the stroke volume (SV) so there are isolines of stroke volume (in grey) for varying parameters. (a) estimation without priors, the estimated values (green) for each case can be anywhere on an isoline (grey). (b) estimation performed with a prior (Gaussian covariance in blue), estimated values are grouped closer to prior mean. (c) Iteratively Updated Priors (IUP) algorithm performs successive estimations where the prior is set from the distribution of estimated parameters at the previous iteration. This leads the parameters to lie on a reduced linear subspace (orange).

6.5.2. Fast Personalized Computer Simulation of Electrical Activation from CT Imaging in Post-infarction Ventricular Tachycardia

Participants: Nicolas Cedilnik [Correspondant], Hubert Cochet [IHU Liryc, Bordeaux], Maxime Sermesant.

This work is funded by the IHU Liryc, in Bordeaux.

Cardiac modeling, Personalised simulation, ablation, intervention guidance.

In the vast majority of post-MI VT ablation procedures, VT is either non inducible or non mappable. We introduce a fast and robust model of cardiac electrophysiology that can be directly parameterized from CT images to predict activation maps (Figure 18). The model is based on the Eikonal equation for wave propagation, where local conduction velocities are estimated from CT. A fully automated method is used to segment the LV wall and assign local conduction velocity according to local LV thickness. Then, a "channelness" filter automatically detects potential VT isthmuses as channels of preserved thickness within severely thinned scar. The model can then be paced within each channel to produce simulated activation maps within seconds. A neural network for automated LV wall segmentation was trained on 450 CTs segmented by experts. Validation, performed on another 50 cases, showed excellent accuracy (Dice score vs. expert 0.95). In 11 patients undergoing post-MI VT ablation (age 58±13, 9 men), simulated activation maps were validated vs. 25 high density maps acquired in Rhythmia (16 paced, 9 VT). Quantitative differences between predicted and measured local activation times remained substantial, particularly in dense scar (> 50ms). Nonetheless, activation patterns were well predicted in most cases (22/25), the 3 poor correlations being observed in patients with fewer scar. Personalized simulation of activation maps from CT scan is feasible and reliably reproduces activation patterns in post-MI VT. The method is fast enough to be used clinically in an interactive fashion for procedural planning [4].

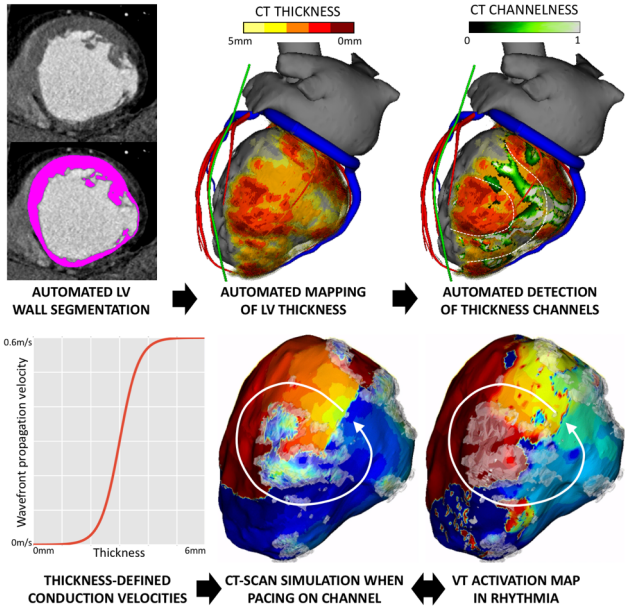


Figure 18. Our image-based model personalization pipeline

6.5.3. Cardiac Modeling, Medical Imaging and Machine Learning for Electrostructural Tomography

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

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

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

Electrocardiographic imaging (ECGI) aims at reconstructing cardiac activity from torso measurements.To achieve this one has to solve the ill-posed inverse problem of the torso propagation. We propose a novel application for Deep Learning Networks to learn spatio-temporal correlations on ECGI (Figure 19). We developped a conditional variational auto-encoder (CVAE).The input are activation maps and the model takes two conditions: on one hand the cardiac shape(cardiac segmentation) and the other one the ECG signals.

The model currently involves simulated data: 120 activations maps were simulated from one cardiac geometry along with simulated body surface potential maps.80% of the data was used for training and the remaining 20% for testing. As a result we were able to observe a good prediction of the activation pattern.

Next, we will test the model with real data provided by the IHU Liryc.

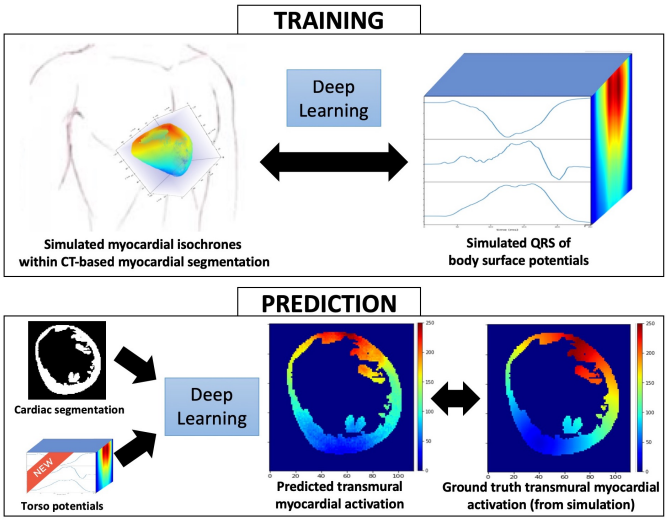


Figure 19. Setup of the conditional variational auto-encoder based on synthetic data

6.5.4. Discovering the link between cardiovascular pathologies and neurodegeneration through biophysical and statistical models of cardiac and brain images

Participants: Jaume Banus Cobo [Correspondant], Maxime Sermesant, Marco Lorenzi.

Université Côte d'Azur (UCA)

Lumped models - Biophysical simulation - Statistical learning

The project aims at developing a computational model of the relationship between cardiac function and brain damage from large-scale clinical databases of multi-modal and multi-organ medical images. The model is based on advanced statistical learning tools for discovering relevant imaging features related to cardiac dysfunction and brain damage; these features are combined within a unified mechanistic framework to providing a novel understanding of the relationship between cardiac function, vascular pathology and brain damage (Fig. 20). We are also testing data-driven statistical learning models for the discovery of associations between cardiac function and brain damage. For example, by applying CCA we identified a first component that shows a positive correlation between the volume of white matter hyper intensities (WMHs), the number of WMHs lesions, the brain ventricles volume and high blood pressure values. On the other side we observed a second component, inversely associated to the first one, in which we observe a strong correlation between ejection fraction (EF) and total brain volume (white matter plus grey matter).

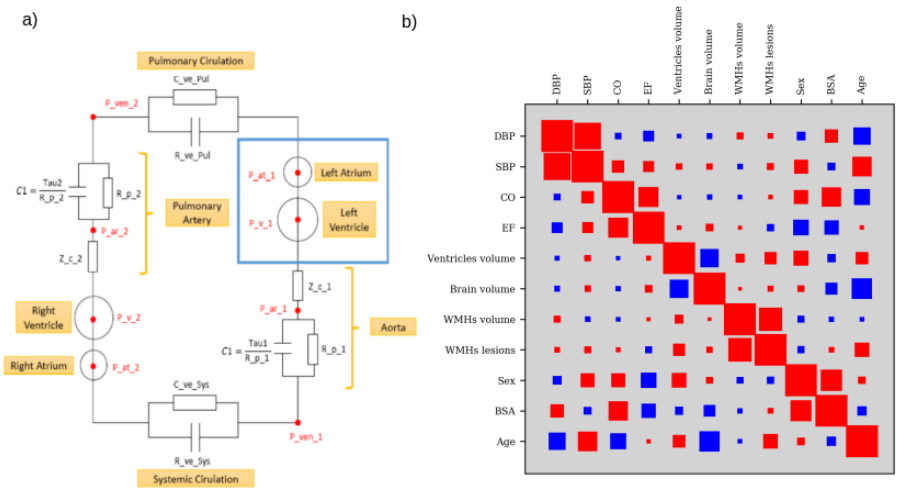


Figure 20. a) Schematic representation of the 0D model used to simulate the whole body circulation, its parameters are optimized to fit the available clinical measurements b) Partial correlations between cardiovascular, brain and demographic variables. Red represents positive correlation and blue negative correlation.

6.5.5. Automatic Image Segmentation of cardiac structures with Adapted U-Net

Participants: Shuman Jia [Correspondant], Antoine Despinasse, Zihao Wang, Hervé Delingette, Xavier Pennec, Hubert Cochet, Maxime Sermesant.

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

We proposed automated, two-stage, three-dimensional U-Nets with a contour loss, to segment the left atrium, as explained in [28], which obtained state-of-the-art results in the STACOM international challenge (Figure 21). Using similar method, we participated in Data Challenge organised at Journées Francophones de Radiologie and obtained a second place for renal cortex segmentation.

6.5.6. Parallel Transport of Surface Deformation

Participants: Shuman Jia [Correspondant], Nicolas Duchateau, Pamela Moceri, Xavier Pennec, Maxime Sermesant.

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

We looked into normalization of temporal deformation and proposed a more symmetric mapping scheme for pole ladder, which relies on geodesic symmetries around mid-points, as illustrated in [29] (Figure 22)). This modified parallel transport method method was shown to be of order 4 on general manifolds and exact in symmetric spaces [58].

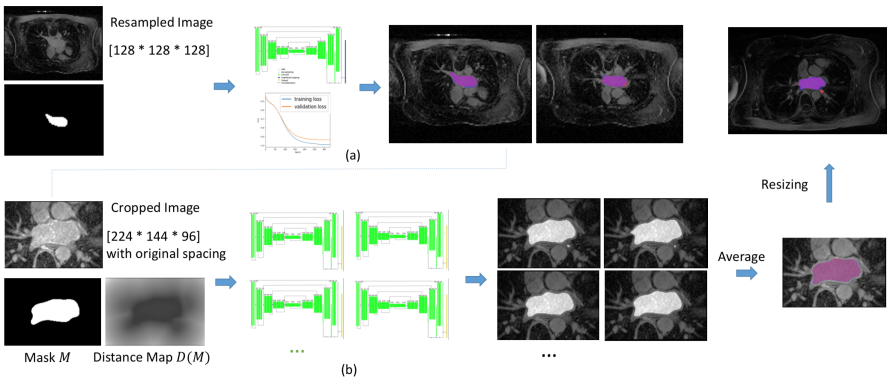


Figure 21. The framework of successive U-Nets training. (a) The fist U-Net - cropping; (b) the second U-Net - segmenting, with ensemble prediction models. We show here axial slices of MR images, overlapped with manual segmentation of the left atrium in blue, our segmentation in red, intersection of the two in purple.

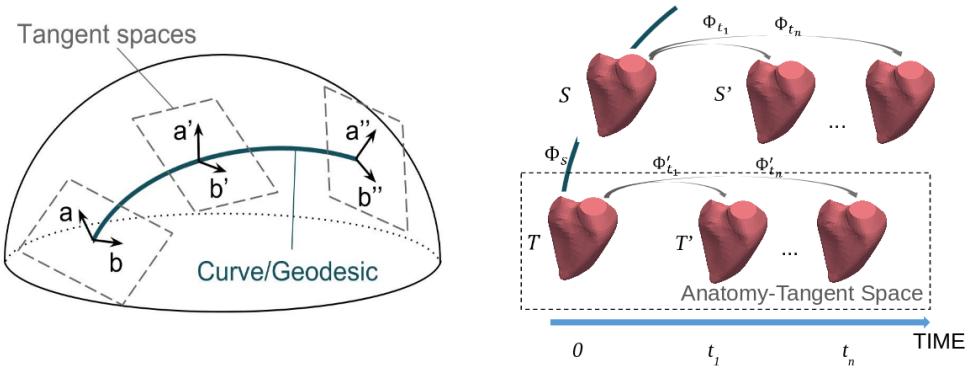


Figure 22. Illustration of parallel transport of vectors a and b along a curve (left) and its application to cardiac imaging (right) with a focus on surfaces.

GALEN-POST Team

7. New Results

7.1. Invertible Deep Networks

Participant: Edouard Oyallon (in collaboration with J.H. Jacobsen and A. Smeulders, Instituut voor Informatica)

It is widely believed that the success of deep convolutional networks is based on progressively discarding uninformative variability about the input with respect to the problem at hand. This is supported empirically by the difficulty of recovering images from their hidden representations, in most commonly used network architectures. In this paper we show via a one-to-one mapping that this loss of information is not a necessary condition to learn representations that generalize well on complicated problems, such as ImageNet. Via a cascade of homeomorphic layers, we build the i-RevNet, a network that can be fully inverted up to the final projection onto the classes, i.e. no information is discarded. Building an invertible architecture is difficult, for one, because the local inversion is ill-conditioned, we overcome this by providing an explicit inverse. An analysis of i-RevNet's learned representations suggests an alternative explanation for the success of deep networks by a progressive contraction and linear separation with depth. To shed light on the nature of the model learned by the i-RevNet we reconstruct linear interpolations between natural image representations [28].

7.2. Compression of CNNs inputs

Participant: Edouard Oyallon (in collaboration with E. Belilovsky, DIRO, Montréal, S. Zagoruyko, WILLOW, Inria Paris and M. Valko, SEQUEL, Inria Lille)

Typical inputs of CNNs are highly redundant and could be potentially reduced. We study the first-order scattering transform as a candidate for reducing the signal processed by a convolutional neural network (CNN). We study this transformation and show theoretical and empirical evidence that in the case of natural images and sufficiently small translation invariance, this transform preserves most of the signal information needed for classification while substantially reducing the spatial resolution and total signal size. We show that cascading a CNN with this representation performs on par with ImageNet classification models commonly used in downstream tasks such as the ResNet-50. We subsequently apply our trained hybrid ImageNet model as a base model on a detection system, which has typically larger image inputs. On Pascal VOC and COCO detection tasks we deliver substantial improvements in the inference speed and training memory consumption compared to models trained directly on the input image [34].

7.3. Interstitial lung disease segmentation

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

Interstitial lung diseases (ILD) encompass a large spectrum of diseases sharing similarities in their physiopathology and computed tomography (CT) appearance. In the work [42], we propose the adaption of a deep convolutional encoder-decoder (CED) that has shown high accuracy for image segmentation. Such architectures require annotation of the total region with pathological findings. This is difficult to acquire, due to uncertainty in the definition and extent of disease patterns and the need of significant human effort, especially for large datasets. Therefore, often current methods use patch-based implementations of convolutional neural networks, which however tend to produce spatially inhomogeneous segmentations due to their local contextual view. We exploit the advantages of both architectures by using the output of a patch-based classifier as a prior to a CED.

Moreover, in order to deal with the limited available datasets that are available, in [41], we introduce a novel multi-network architecture that exploits domain knowledge to address those challenges. The proposed architecture consists of multiple deep neural networks that are trained after co-aligning multiple anatomies through multi-metric deformable registration. This multi-network architecture can be trained with fewer examples and leads to better performance, robustness and generalization through consensus. Comparable to human accuracy, highly promising results on the challenging task of interstitial lung disease segmentation demonstrate the potential of our approach.

7.4. Image Registration with 3D Convolutional Neural Networks

Participants: Stergios Christodoulidis, Mihir Sahasrabudhe, Guillaume Chassagnon, Maria Vakalopoulou (in collaboration with Stavroula Mougiakakou and Marie-Pierre Revel and Nikos Paragios: University of Bern; AP-HP - Hopital Cochin Broca Hotel Dieu; Therapanacea)
Image registration and in particular deformable registration methods are pillars of medical imaging. Inspired by the recent advances in deep learning, we propose in this paper, [25] we proposed a new deep learning based and unsupervised method for image registration. In partucular, a novel convolutional neural network architecture that couples linear and deformable registration within a unified architecture endowed with near real-time performance. Our framework is modular with respect to the global transformation component, as well as with respect to the similarity function while it guarantees smooth displacement fields. We evaluate the performance of our network on the challenging problem of MRI lung registration, and demonstrate superior performance with respect to state of the art elastic registration methods. The proposed deformation (between inspiration & expiration) was considered within a clinically relevant task of interstitial lung disease (ILD) classification and showed promising results.

7.5. Radiomics for response to immunotherapy

Participants: Roger Sun, Maria Vakalopoulou (in collaboration with Elaine Johanna Limkin, Laurent Dercle, Stéphane Champiat, Shan Rong Han, Loic Verlingue, David Brandao, Andrea Lancia, Samy Ammari, Antoine Hollebecque, Jean-Yves Scoazec, Aurélien Marabelle, Christophe Massard, Jean-Charles Soria, Charlotte Robert, Nikos Paragios, Eric Deutsch, Charles Ferté: Institute Gustave Roussy; Therapanacea)
Because responses of patients with cancer to immunotherapy can vary in success, innovative predictors of response to treatment are urgently needed to improve treatment outcomes. In this retrospective multicohort work [17], we used four independent cohorts of patients with advanced solid tumours to develop and validate a radiomic signature predictive of immunotherapy response by combining contrast-enhanced CT images and RNA-seq genomic data from tumour biopsies to assess CD8 cell tumour infiltration. To develop the radiomic signature of CD8 cells, we used the CT images and RNA sequencing data of 135 patients with advanced solid malignant tumours who had been enrolled into the MOSCATO trial between May 1, 2012, and March 31, 2016, in France (training set). The genomic data, which are based on the CD8B gene, were used to estimate the abundance of CD8 cells in the samples and data were then aligned with the images to generate the radiomic signatures. The concordance of the radiomic signature (primary endpoint) was validated in a Cancer Genome Atlas [TGCA] database dataset including 119 patients who had available baseline preoperative imaging data and corresponding transcriptomic data on June 30, 2017. From 84 input variables used for the machine-learning method (78 radiomic features, five location variables, and one technical variable), a radiomics-based predictor of the CD8 cell expression signature was built by use of machine learning (elastic-net regularised regression method). Two other independent cohorts of patients with advanced solid tumours were used to evaluate this predictor. The immune phenotype internal cohort (n=100), were randomly selected from the Gustave Roussy Cancer Campus database of patient medical records based on previously described, extreme tumour-immune phenotypes: immune-inflamed (with dense CD8 cell infiltration) or immune-desert (with low CD8 cell infiltration), irrespective of treatment delivered; these data were used to analyse the correlation of the immune phenotype with this biomarker. Finally, the immunotherapy-treated dataset (n=137) of patients recruited from Dec 1, 2011, to Jan 31, 2014, at the Gustave Roussy Cancer Campus, who had been treated with anti-PD-1 and anti-PD-L1 monotherapy in phase 1 trials, was used to assess the predictive value of this biomarker in terms of clinical outcome.

7.6. Semantic Segmentation Techniques for Brain Tumor Patients

Participants: Siddhartha Chandra, Théo Estienne, Roger Sun, Enzo Battistella, Maria Vakalopoulou (in collaboration with Charlotte Robert, Nikos Paragios, Eric Deutsch: Institute Gustave Roussy; Therapanacea)
In this work [23] we propose a novel deep learning based pipeline for the task of brain tumor segmentation. Our pipeline consists of three primary components: (i) a preprocessing stage that exploits histogram standardization to mitigate inaccuracies in measured brain modalities, (ii) a first prediction stage that uses the V-Net deep learning architecture to output dense, per voxel class probabilities, and (iii) a prediction refinement stage that uses a Conditional Random Field (CRF) with a bilateral filtering objective for better context awareness. Additionally, we compare the V-Net architecture with a custom 3D Residual Network architecture, trained on a multi-view strategy, and our ablation experiments indicate that V-Net outperforms the 3D ResNet-18 with all bells and whistles, while fully connected CRFs as post processing, boost the performance of both networks. We report competitive results on the BraTS 2018 validation and test set as also summarized on [52].

7.7. Demystification of AI-driven medical image interpretation

Participants: Maria Vakalopoulou (in collaboration with P. Savadjiev, J. Chong, A. Dohan, C. Reinhold, B. Gallix: McGill University; Therapanacea)
The recent explosion of 'big data' has ushered in a new era of artificial intelligence (AI) algorithms in every sphere of technological activity, including medicine, and in particular radiology. However, the recent success of AI in certain flagship applications has, to some extent, masked decades-long advances in computational technology development for medical image analysis. In this work [16], we provide an overview of the history of AI methods for radiological image analysis in order to provide a context for the latest developments. We review the functioning, strengths and limitations of more classical methods as well as of the more recent deep learning techniques. We discuss the unique characteristics of medical data and medical science that set medicine apart from other technological domains in order to highlight not only the potential of AI in radiology but also the very real and often overlooked constraints that may limit the applicability of certain AI methods. Finally, we provide a comprehensive perspective on the potential impact of AI on radiology and on how to evaluate it not only from a technical point of view but also from a clinical one, so that patients can ultimately benefit from it.

7.8. Semantic Segmentation for Remote Sensing Data

Participants: Maria Papadomanolaki, Maria Vakalopoulou (in collaboration with Christina Karakizi, Georgia Antoniou, Konstantinos Karantzalos, Nikos Paragios; National Technical University of Athens, Therapanacea)
Detailed, accurate and frequent land cover mapping is a prerequisite for several important geospatial applications and the fulfilment of current sustainable development goals. This work [9] introduces a methodology for the classification of annual high-resolution satellite data into several detailed land cover classes. In particular, a nomenclature with 27 different classes was introduced based on CORINE Land Cover (CLC) Level-3 categories and further analysing various crop types. Without employing cloud masks and/or interpolation procedures, we formed experimental datasets of Landsat-8 (L8) images with gradually increased cloud cover in order to assess the influence of cloud presence on the reference data and the resulting classification accuracy. The performance of shallow kernel-based and deep patch-based machine learning classification frameworks was evaluated. Quantitatively, the resulting overall accuracy rates differed within a range of less than 3%; however, maps produced based on Support Vector Machines (SVM) were more accurate across class boundaries and the respective framework was less computationally expensive compared to the applied patch-based deep Convolutional Neural Network (CNN). Further experimental results and analysis indicated that employing all multitemporal images with up to 30% cloud cover delivered relatively higher overall accuracy rates as well as the highest per-class accuracy rates. Moreover, by selecting 70% of the top-ranked features after applying a feature selection strategy, slightly higher accuracy rates were achieved. A detailed discussion of the quantitative and qualitative evaluation outcomes further elaborates on the performance of all considered classes and highlights different aspects of their spectral behaviour and separability.

Moreover, semantic segmentation is a mainstream method in several remote sensing applications based on very-high-resolution data, achieving recently remarkable performance by the use of deep learning and more specifically, pixel-wise dense classification models. In this work [36], we exploit the use of a relatively deep architecture based on repetitive downscale upscale processes that had been previously employed for human pose estimation. By integrating such a model, we are aiming to capture low-level details, such as small objects, object boundaries and edges. Experimental results and quantitative evaluation has been performed on the publicly available ISPRS (WGIII/4) benchmark dataset indicating the potential of the proposed approach.

7.9. BRANE Clust: Cluster-Assisted Gene Regulatory Network Inference Refinement

Participants: Jean-Christophe Pesquet (in collaboration with Aurélie Pirayre, IFP Energies nouvelles, Camille Couprie, Facebook Research, Laurent Duval, IFP Energies nouvelles)
Discovering meaningful gene interactions is crucial for the identification of novel regulatory processes in cells. Building accurately the related graphs remains challenging due to the large number of possible solutions from available data. Nonetheless, enforcing a priori on the graph structure, such as modularity, may reduce network indeterminacy issues. BRANE Clust (Biologically-Related A priori Network Enhancement with Clustering) refines gene regulatory network (GRN) inference thanks to cluster information. It works as a post-processing tool for inference methods (i.e. CLR, GENIE3). In BRANE Clust, the clustering is based on the inversion of a system of linear equations involving a graph-Laplacian matrix promoting a modular structure. Our approach [14] is validated on DREAM4 and DREAM5 datasets with objective measures, showing significant comparative improvements. We provide additional insights on the discovery of novel regulatory or co-expressed links in the inferred Escherichia coli network evaluated using the STRING database. The comparative pertinence of clustering is discussed computationally (SIMoNe, WGCNA, X-means) and biologically (RegulonDB).

7.10. Proximity Operators of Discrete Information Divergences

Participants: Jean-Christophe Pesquet (in collaboration with Mireille El Gheche, EPFL, Giovanni Chierchia, ESIEE Paris)
Information divergences allow one to assess how close two distributions are from each other. Among the large panel of available measures, a special attention has been paid to convex ϕ-divergences, such as Kullback-Leibler, Jeffreys-Kullback, Hellinger, Chi-Square, Renyi, and I_α divergences. While ϕ-divergences have been extensively studied in convex analysis, their use in optimization problems often remains challenging. In this regard, one of the main shortcomings of existing methods is that the minimization of ϕ-divergences is usually performed with respect to one of their arguments, possibly within alternating optimization techniques. In this paper, we overcome this limitation by deriving new closed-form expressions for the proximity operator of such two-variable functions. This makes it possible to employ standard proximal methods for efficiently solving a wide range of convex optimization problems involving ϕ-divergences. In addition, we show that these proximity operators are useful to compute the epigraphical projection of several functions of practical interest. The proposed proximal tools are numerically validated in the context of optimal query execution within database management systems, where the problem of selectivity estimation plays a central role. Experiments are carried out on small to large scale scenarios [6].

7.11. Stochastic quasi-Fejèr block-coordinate fixed point iterations with random sweeping

Participants: Jean-Christophe Pesquet (in collaboration with Patrick Combettes, North Caroline State University)

Our previous work investigated the almost sure weak convergence of block-coordinate fixed point algorithms and discussed their applications to nonlinear analysis and optimization. This algorithmic framework features random sweeping rules to select arbitrarily the blocks of variables that are activated over the course of the iterations and it allows for stochastic errors in the evaluation of the operators. The present paper establishes results on the mean-square and linear convergence of the iterates. Applications to monotone operator splitting and proximal optimization algorithms are presented.

7.12. Rational optimization for nonlinear reconstruction with approximate ℓ_0 penalization

Participants: Marc Castella, Arthur Marmin, Jean-Christophe Pesquet

Recovering nonlinearly degraded signal in the presence of noise is a challenging problem. In this work, this problem is tackled by minimizing the sum of a non convex least-squares fit criterion and a penalty term. We assume that the nonlinearity of the model can be accounted for by a rational function. In addition, we suppose that the signal to be sought is sparse and a rational approximation of the ℓ_0 pseudo-norm thus constitutes a suitable penalization. The resulting composite cost function belongs to the broad class of semi-algebraic functions. To find a globally optimal solution to such an optimization problem, it can be transformed into a generalized moment problem, for which a hierarchy of semidefinite programming relaxations can be built. Global optimality comes at the expense of an increased dimension and, to overcome computational limitations concerning the number of involved variables, the structure of the problem has to be carefully addressed. A situation of practical interest is when the nonlinear model consists of a convolutive transform followed by a componentwise nonlinear rational saturation. We then propose to use a sparse relaxation able to deal with up to several hundreds of optimized variables. In contrast with the naive approach consisting of linearizing the model, our experiments show that the proposed approach offers good performance [53].

7.13. Representation Learning on Real-World Graphs

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat (in collaboration with Duong Nguyen, UC San Diego)

Network representation learning (NRL) methods aim to map each vertex into a low dimensional space by preserving both local and global structure of a given network. In recent years, various approaches based on random walks have been proposed to learn node embeddings – thanks to their success in several challenging problems. In this work, we have introduced two methodologies to compute latent representations of nodes based on random walks.

In particular, we have proposed BiasedWalk, an unsupervised Skip-gram-based network embedding algorithm which can preserve higher-order proximity information, as well as capture both the homophily and role equivalence relationships between nodes [33]. BiasedWalk relies on a novel node sampling procedure based on biased random walks, that can behave as actual depth-first-search and breath-first-search explorations – thus, forcing the sampling scheme to capture both role equivalence and homophily relations between nodes. Furthermore, BiasedWalk is scalable on large scale graphs, and is able to handle different types of networks structures, including (un)weighted and (un)directed ones.

Furthermore, we have introduced TNE (Topical Node Embeddings), a general framework to enhance node embeddings acquired by means of the random walk-based approaches [45]. Similar to the notion of *topical word embeddings* in the domain of Natural Language Processing, the proposed framework assigns each vertex to a topic with the favor of various statistical models and community detection methods, and then generates enhanced community representations.

We have evaluated our methods on two downstream tasks: node classification and link prediction in social, information and biological networks. The experimental results demonstrate that the biased random walks as well as the incorporation of vertex and topic embeddings outperform widely-known baseline NRL methods.

7.14. Anonymity on Directed Networks

Participants: Fragkiskos Malliaros (in collaboration with Jordi Casas-Roma and Julián Salas, Universitat Oberta de Catalunya; Michalis Vazirgiannis, École Polytechnique)

In recent years, a huge amount of social and human interaction networks have been made publicly available. Embedded within this data, there is user's private information that must be preserved before releasing the data to third parties and researchers. In this work, we have considered the problem of anonymization on directed networks. Although there are several anonymization methods for networks, most of them have explicitly been designed to work with undirected networks and they can not be straightforwardly applied when they are directed. Moreover, ignoring the direction of the edges causes important information loss on the anonymized networks in the best case. In the worst case, the direction of the edges may be used for reidentification, if it is not considered in the anonymization process. Here, we have proposed two different models for k-degree anonymity on directed networks, and we also present algorithms to fulfill these k-degree anonymity models [4]. Given a network G, we construct a k-degree anonymous network by the minimum number of edge additions. Our algorithms use multivariate micro-aggregation to anonymize the degree sequence, and then they modify the graph structure to meet the k-degree anonymous sequence. We apply our algorithms to several real datasets and demonstrate their efficiency and practical utility.

7.15. Influence Maximization in Complex Networks

Participants: Fragkiskos Malliaros (in collaboration with Michalis Vazirgiannis, George Panagopoulos, Maria-Evgenia Rossi, Bowen Shi, Christos Giatsidis, École Polytechnique; Nikolaos Tziortziotis, Université Paris-Sud)

Influence maximization in complex networks has attracted a lot of attention due to its numerous applications, including diffusion of social movements, the spread of news, viral marketing and outbreak of diseases. The objective is to discover a group of users that are able to maximize the spread of influence across a network. The seminal *greedy* algorithm developed by Kempe, Kleinberg and Tardos progressively adds new nodes to the seed set, maximizing the expected influence spread; the algorithm gives a solution to the influence maximization problem while having a good approximation ratio.

Nevertheless, one of the bottlenecks of the greedy algorithm is that it does not scale well on large scale datasets. In our work, we have proposed Matrix Influence (MATI), an efficient algorithm that can be used under both the Linear Threshold and Independent Cascade diffusion models [15]. MATI is based on the precalculation of the influence by taking advantage of the simple paths in the node's neighborhood. An extensive empirical analysis has been performed on multiple real-world datasets showing that MATI has competitive performance when compared to other well-known algorithms with regards to running time and expected influence spread.

Furthermore, the previously described greedy algorithm focuses solely on static networks. However, with the emergence of several complementary data, such as the network's temporal changes and the diffusion cascades taking place over it, novel methods have been proposed with promising results. In our work, we have introduced a simple yet effective algorithm (called DiffuGreedy) that combines the algorithmic methodology of the greedy approach with diffusion cascades [35]. We have compared it with four different prevalent influence maximization approaches, on a large scale Chinese microblogging dataset. More specifically, for comparison, we have employed methods that derive the seed set using the static network, the temporal network, the diffusion cascades, and their combination. A set of diffusion cascades from the latter part of the dataset is set aside for evaluation. The experimental evaluation has shown that the proposed DiffuGreedy outperforms widely used baseline methods in both quality of the seed set and computational efficiency.

7.16. Graph-based Text Analytics

Participants: Fragkiskos Malliaros (in collaboration with Konstantinos Skianis and Michalis Vazirgiannis, École Polytechnique)

Text categorization is a core task in a plethora of text mining applications. In our work, contrary to the traditional *Bag-of-Words* approach, we have considered the *Graph-of-Words* model in which each document is represented by a graph that encodes relationships between the different terms. Based on this formulation, we treat the term weighting task as a node ranking problem; the importance of a term is determined by the importance of the corresponding node in the graph, using node centrality criteria. We have also introduced novel graph-based weighting schemes by enriching graphs with word-embedding distances, in order to reward or penalize the importance of semantically close terms [39]. Our methods produce more discriminative feature weights for text categorization, outperforming existing frequency-based criteria – highlighting also the importance of graph-based methods in text analytics and natural language processing in general.

7.17. Auxiliary Variable Method for MCMC Algorithms in High Dimension

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Yosra Marnissi, SAFRAN TECH, and Amel Benazza-Benyahia, SUP'COM Tunis)
In this work, we are interested in Bayesian inverse problems where either the data fidelity term or the prior distribution is Gaussian or driven from a hierarchical Gaussian model. Generally, Markov chain Monte Carlo (MCMC) algorithms allow us to generate sets of samples that are employed to infer some relevant parameters of the underlying distributions. However, when the parameter space is high-dimensional, the performance of stochastic sampling algorithms is very sensitive to existing dependencies between parameters. In particular, this problem arises when one aims to sample from a high-dimensional Gaussian distribution whose covariance matrix does not present a simple structure. Another challenge is the design of Metropolis–Hastings proposals that make use of information about the local geometry of the target density in order to speed up the convergence and improve mixing properties in the parameter space, while not being too computationally expensive. These two contexts are mainly related to the presence of two heterogeneous sources of dependencies stemming either from the prior or the likelihood in the sense that the related covariance matrices cannot be diagonalized in the same basis. In this work, we address these two issues. Our contribution consists of adding auxiliary variables to the model in order to dissociate the two sources of dependencies. In the new augmented space, only one source of correlation remains directly related to the target parameters, the other sources of correlations being captured by the auxiliary variables. Experimental results conducted on two practical image restoration problems indicate that adding the proposed auxiliary variables makes the sampling problem simpler, and thus the computational cost of each iteration of the Gibbs sampler is significantly reduced while ensuring good mixing properties [11].

7.18. Generation of patient-specific cardiac vascular networks

Participant: Hugues Talbot (in collaboration with C. Jaquet, L. Najman, ESIEE Paris; L. Grady, M. Schaap, B. Spain, H. Kim, C. Taylor, HeartFlow; I. Vignon-Clementel, REO)
In this work, we have proposed a blood-vessel generation procedure for extending known patient vasculature over and within the heart ventricle [8]. It is patient-specific, in the sense that it extend the known, segmented patient vasculature, and it is consistent with physics-based blood vessels characteristics (i.e. derived from CFD) and known vessel physiology. The generated vascular network bridges the gap between the vasculature that can be imaged and assessed via classical means (CT or MRI) and perfusion maps that can be imaged with specific modalities (radiotracer injected scintigraphy or PET). One objective of this work is to eventually propose a forward model for perfusion map generation, that can be used to solved the associated inverse problem of finding the cause of observed perfusion deficits associated with coronary diseases that cannot be imaged directly.

7.19. Curvilinear structure analysis using path operators

Participant: Hugues Talbot (in collaboration with O. Merveille, N. Passat, CRESTIC, and L. Najman, ESIEE Paris)

In this work, we propose mathematical morphology based operators that use paths as families of structuring elements [12]. Structuring elements are like the windows of linear operators, they define the extent of the related operators (convolutions in the linear case, openings and closings in the morphology case). When dealing with thin objects (e.g. fibres, blood vessels, textures, etc), a compact, isotropic window is usually inappropriate because no such window can fit in these objects. This is more critical for morphology, which is concerned with preserving shapes, than with linear operators. Thin windows must therefore be devised, but there are a large number of potentially interesting thin windows at each point in an image. In this article, we leverage the definition of noise-resistant, path operators to define a non-linear notion of vesselness, that can be used for thin object detection, filtering and segmentation in 2D and 3D.

7.20. High throughput automated detection of axial malformations in fish embryo

Participant: Hugues Talbot (in collaboration with D. Genest, M. Léonard, N. De Crozé, L'Oréal, and E. Puybareau, J. Cousty, LIGM)

Fish embryos are used throughout the cosmetics industry to assess the toxicity of the components of their products, as well as more generally in waterways pollution measurements. Indeed pollution is often detectable in trace amounts when they hinder, stop or cause malformations during fish embryo development. In this work, we propose a high-throughput procedure for detecting tail malformation in fish embryo, based on image analysis and machine learning [5]. These malformation are among the most difficult to assess but very common in various degrees of severity. Our procedure provide similar error rate as trained and careful humans operators, as assessed on thousands of images acquired in partneship with L'Oréal. We also show that our procedure is much faster and more consistent than human operators. It is now used in production by our partner.

MATHNEURO Team

4. New Results

4.1. Neural Networks as dynamical systems

4.1.1. Latching dynamics in neural networks with synaptic depression

Participants: Elif Köksal Ersöz, Carlos Aguilar [Université de Nice - BCL], Pascal Chossat [Université de Nice - LJAD, Inria MathNeuro], Martin Krupa [Université de Nice - LJAD, UCA, Inria MathNeuro], Frédéric Lavigne [Université de Nice - BCL].

Prediction is the ability of the brain to quickly activate a target concept in response to a related stimulus (prime). Experiments point to the existence of an overlap between the populations of the neurons coding for different stimuli, and other experiments show that prime-target relations arise in the process of long term memory formation. The classical modelling paradigm is that long term memories correspond to stable steady states of a Hopfield network with Hebbian connectivity. Experiments show that short term synaptic depression plays an important role in the processing of memories. This leads naturally to a computational model of priming, called latching dynamics; a stable state (prime) can become unstable and the system may converge to another transiently stable steady state (target). Hopfield network models of latching dynamics have been studied by means of numerical simulation, however the conditions for the existence of this dynamics have not been elucidated. In this work we use a combination of analytic and numerical approaches to confirm that latching dynamics can exist in the context of a symmetric Hebbian learning rule, however lacks robustness and imposes a number of biologically unrealistic restrictions on the model. In particular our work shows that the symmetry of the Hebbian rule is not an obstruction to the existence of latching dynamics, however fine tuning of the parameters of the model is needed.

A natural follow-up of the work which has lead to the article [1] has been initiated through the postdoc project of Elif Köksal Ersöz. The objective is to extend the previous results in several ways. First, to gain more robustness in the heteroclinic chains sustained by the network model. Second, to be able to simulate much larger networks and exhibit heteroclinic dynamics in them. Third, to link with experimental data. The postdoc of Elif Köksal Ersöz, which finished at the end of December 2018, has been funded by the "tail" of the ERC Advanced Grant NerVi held by Olivier Faugeras.

4.1.2. Pseudo-simple heteroclinic cycles in $\mathbb{R}^4$

Participants: Pascal Chossat [Université de Nice - LJAD, Inria MathNeuro], Alexander Lohse [Universität Hamburg, Germany], Olga Podvigina [Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russia].

We study pseudo-simple heteroclinic cycles for a Γ-equivariant system in $\mathbb{R}^4$ with finite $\Gamma \subset O(4)$, and their nearby dynamics. In particular, in a first step towards a full classification – analogous to that which exists already for the class of simple cycles – we identify all finite subgroups of $O(4)$ admitting pseudo-simple cycles. To this end we introduce a constructive method to build equivariant dynamical systems possessing a robust heteroclinic cycle. Extending a previous study we also investigate the existence of periodic orbits close to a pseudo-simple cycle, which depends on the symmetry groups of equilibria in the cycle. Moreover, we identify subgroups $\Gamma \subset O(4)$, $\Gamma\neg \subset SO(4)$, admitting fragmentarily asymptotically stable pseudo-simple heteroclinic cycles (It has been previously shown that for $\Gamma \subset SO(4)$ pseudo-simple cycles generically are completely unstable). Finally, we study a generalized heteroclinic cycle, which involves a pseudo-simple cycle as a subset.

This work has been published in Physica D and is available as [13].

4.1.3. Qualitative stability and synchronicity analysis of power network models in port-Hamiltonian form

Participants: Volker Mehrmann [Technical University of Berlin, Germany], Riccardo Morandin [Technical University of Berlin, Germany], Simona Olmi, Eckehard Schöll [Technical University of Berlin, Germany].

In view of highly decentralized and diversified power generation concepts, in particular with renewable energies, the analysis and control of the stability and the synchronization of power networks is an important topic that requires different levels of modeling detail for different tasks. A frequently used qualitative approach relies on simplified nonlinear network models like the Kuramoto model with inertia. The usual formulation in the form of a system of coupled ordinary differential equations is not always adequate. We present a new energy-based formulation of the Kuramoto model with inertia as a polynomial port-Hamiltonian system of differential-algebraic equations, with a quadratic Hamiltonian function including a generalized order parameter. This leads to a robust representation of the system with respect to disturbances: it encodes the underlying physics, such as the dissipation inequality or the deviation from synchronicity, directly in the structure of the equations, and it explicitly displays all possible constraints and allows for robust simulation methods. The model is immersed into a system of model hierarchies that will be helpful for applying adaptive simulations in future works. We illustrate the advantages of the modified modeling approach with analytics and numerical results.

This work has been published in Chaos and is available as [18].

4.1.4. Collective behavior of oscillating electric dipoles

Participants: Simona Olmi, Matteo Gori [Centre de Physique Théorique, Marseille], Irene Donato [Centre de Physique Théorique, Marseille], Marco Pettini [Centre de Physique Théorique, Marseille].

We investigate the dynamics of a population of identical biomolecules mimicked as electric dipoles with random orientations and positions in space and oscillating with their intrinsic frequencies. The biomolecules, beyond being coupled among themselves via the dipolar interaction, are also driven by a common external energy supply. A collective mode emerges by decreasing the average distance among the molecules as testified by the emergence of a clear peak in the power spectrum of the total dipole moment. This is due to a coherent vibration of the most part of the molecules at a frequency definitely larger than their own frequencies corresponding to a partial cluster synchronization of the biomolecules. These results can be verified experimentally via spectroscopic investigations of the strength of the intermolecular electrodynamic interactions, thus being able to test the possible biological relevance of the observed macroscopic mode.

This work has been published in Scientific Reports and is available as [19].

4.1.5. Controlling seizure propagation in large-scale brain networks

Participants: Simona Olmi, Spase Petkoski [Institut de Neurosciences des Systèmes, Marseille], Maxime Guye [Centre d'Exploration Métabolique par Résonance Magnétique, Marseille], Fabrice Bartolomei [Epilepsies, Lésions Cérébrales et Systèmes Neuraux de la Cognition, Marseille], Viktor Jirsa [Institut de Neurosciences des Systèmes, Marseille].

Information transmission in the human brain is a fundamentally dynamic network process. In partial epilepsy, this process is perturbed and highly synchronous seizures originate in a local network, the so-called epileptogenic zone (EZ), before recruiting other close or distant brain regions. We studied patient-specific brain network models of 15 drug-resistant epilepsy patients with implanted stereotactic electroencephalography (SEEG) electrodes. Each personalized brain model was derived from structural data of magnetic resonance imaging (MRI) and diffusion tensor weighted imaging (DTI), comprising 88 nodes equipped with region specific neural mass models capable of demonstrating a range of epileptiform discharges. Each patients virtual brain was further personalized through the integration of the clinically hypothesized EZ. Subsequent simulations and connectivity modulations were performed and uncovered a finite repertoire of seizure propagation patterns. Across patients, we found that (i) patient-specific network connectivity is predictive for the subsequent seizure propagation pattern; (ii) seizure propagation is characterized by a systematic sequence of brain

states; (iii) propagation can be controlled by an optimal intervention on the connectivity matrix; (iv) the degree of invasiveness can be significantly reduced via the here proposed seizure control as compared to traditional resective surgery. To stop seizures, neurosurgeons typically resect the EZ completely. We showed that stability analysis of the network dynamics using graph theoretical metrics estimates reliably the spatiotemporal properties of seizure propagation. This suggests novel less invasive paradigms of surgical interventions to treat and manage partial epilepsy.

This work has been submitted for publication and is available as [28].

4.1.6. Effect of disorder and noise in shaping the dynamics of power grids

Participants: Liudmila Tumash [Technical University of Berlin, Germany], Simona Olmi, Eckehard Schöll [Technical University of Berlin, Germany].

The aim of this paper is to investigate complex dynamic networks which can model high-voltage power grids with renewable, fluctuating energy sources. For this purpose we use the Kuramoto model with inertia to model the network of power plants and consumers. In particular, we analyse the synchronization transition of networks of N phase oscillators with inertia (rotators) whose natural frequencies are bimodally distributed, corresponding to the distribution of generator and consumer power. First, we start from globally coupled networks whose links are successively diluted, resulting in a random Erdös-Renyi network. We focus on the changes in the hysteretic loop while varying inertial mass and dilution. Second, we implement Gaussian white noise describing the randomly fluctuating input power, and investigate its role in shaping the dynamics. Finally, we briefly discuss power grid networks under the impact of both topological disorder and external noise sources.

This work has been published in Europhysics Letters and is available as [20].

4.2. Mean field theory and stochastic processes

4.2.1. Emergence of collective phenomena in a population of neurons

Participants: Benjamin Aymard, Fabien Campillo, Romain Veltz.

In this work, we propose a new model of biological neural network, combining a two-dimensional integrate-and-fire neuron model with a deterministic model of electrical synapse, and a stochastic model of chemical synapse. We describe the dynamics of a population of neurons in interaction as a piecewise deterministic Markov process. We prove the weak convergence of the associated empirical process, as the population size tends to infinity, towards a McKean-Vlasov type process and we describe the associated PDE. We are also interested in the simulation of these dynamics, in particular by comparing "detailed" simulations of a finite population of neurons with a simulation of the system with infinite population. Benjamin Aymard has the adapted toolkit to attack these questions numerically. The mean field equations studied by Benjamin are of transport type for which numerical methods are technical. However, they are the domain of expertise of Benjamin. His postdoc is funded by the Flagship Human Brain Project.
Latest results are as follows. A first manuscript concerning the numerical simulation of the mean field model is in preparation. We managed to find a new numerical scheme which is positive, semi-implicit and adaptive in time (of order 2). A second manuscript concerning the existence of stationary distribution for the mean field limit is also in preparation.

4.2.2. Long time behavior of a mean-field model of interacting neurons

Participants: Quentin Cormier [Inria TOSCA], Étienne Tanré [Inria TOSCA], Romain Veltz.

We study the long time behavior of the solution to some McKean-Vlasov stochastic differential equation (SDE) driven by a Poisson process. In neuroscience, this SDE models the asymptotic dynamic of the membrane potential of a spiking neuron in a large network. We prove that for a small enough interaction parameter, any solution converges to the unique (in this case) invariant measure. To this aim, we first obtain global bounds on the jump rate and derive a Volterra type integral equation satisfied by this rate. We then replace temporary the interaction part of the equation by a deterministic external quantity (we call it the external current). For constant current, we obtain the convergence to the invariant measure. Using a perturbation method, we extend this result to more general external currents. Finally, we prove the result for the non-linear McKean-Vlasov equation.

This work has been submitted for publication and is available as [24].

4.2.3. *Exponential stability of the stationary distribution of a mean field of spiking neural network*

Participants: Audric Drogoul [Thales, France], Romain Veltz.

In this work, we study the exponential stability of the stationary distribution of a McKean-Vlasov equation, of nonlinear hyperbolic type which was recently derived in [33], [40]. We complement the convergence result proved in [40] using tools from dynamical systems theory. Our proof relies on two principal arguments in addition to a Picard-like iteration method. First, the linearized semigroup is positive which allows to precisely pinpoint the spectrum of the infinitesimal generator. Second, we use a time rescaling argument to transform the original quasilinear equation into another one for which the nonlinear flow is differentiable. Interestingly, this convergence result can be interpreted as the existence of a locally exponentially attracting center manifold for a hyperbolic equation.

This work has been submitted for publication and is available as [23].

4.2.4. *On a toy network of neurons interacting through their dendrites*

Participants: Nicolas Fournier Cormier [LPSM, Sorbonne Université, Paris], Étienne Tanré [Inria TOSCA], Romain Veltz.

Consider a large number n of neurons, each being connected to approximately N other ones, chosen at random. When a neuron spikes, which occurs randomly at some rate depending on its electric potential, its potential is set to a minimum value vmin, and this initiates, after a small delay, two fronts on the (linear) dendrites of all the neurons to which it is connected. Fronts move at constant speed. When two fronts (on the dendrite of the same neuron) collide, they annihilate. When a front hits the soma of a neuron, its potential is increased by a small value w_n. Between jumps, the potentials of the neurons are assumed to drift in $[v_{min}, \infty)$, according to some well-posed ODE. We prove the existence and uniqueness of a heuristically derived mean-field limit of the system when $n, N \to \infty$ with $w_n \simeq N1/2$. We make use of some recent versions of the results of Deuschel and Zeitouni [37] concerning the size of the longest increasing subsequence of an i.i.d. collection of points in the plan. We also study, in a very particular case, a slightly different model where the neurons spike when their potential reach some maximum value v_{max}, and find an explicit formula for the (heuristic) mean-field limit.

This work has been submitted for publication and is available as [26].

4.2.5. *Mathematical statistical physics applied to neural populations*

Participants: Émilie Soret, Olivier Faugeras, Étienne Tanré [Inria, project-team TOSCA, Sophia-Antipolis].

This project focuses on Mean-Field descriptions or thermodynamics limits of large populations of neurons. They study a system of Stochastic Differential Equations (SDEs) which describes the evolution of membrane potential of each neuron over the time when the synaptic weights are random variables (not assumed to be independent). This setup is well suited to Émilie, who has worked during her PhD and first postdoc on mathematical statistical physics and stochastic processes. Her postdoc is funded by the Flagship Human Brain Project. A manuscript is in preparation.

4.3. Neural fields theory

4.3.1. A neural field model for color perception unifying assimilation and contrast

Participants: Anna Song [ENS Paris, France], Olivier Faugeras, Romain Veltz.

We propose a neural field model of color perception in context, for the visual area V1 in the cortex. This model reconciles into a common framework two opposing perceptual phenomena, simultaneous contrast and chromatic assimilation. Previous works showed that they act simultaneously, and can produce larger shifts in color matching when acting in synergy with a spatial pattern. At some point in an image,the color perceptually seems more similar to that of the adjacent locations, while being more dissimilar from that of remote neighbors. The influence of neighbors hence reverses its nature above some characteristic scale. Our model fully exploits the balance between attraction and repulsion in color space, combined at small or large scales in physical space. For that purpose we rely on the opponent color theory introduced by Hering, and suppose a hypercolumnar structure coding for colors. At some neural mass, the pointwise influence of neighbors is spatially integrated to obtain the final effect that we call a color sensation. Alongside this neural field model, we describe the search for a color match in asymmetric matching experiments as a mathematical projector. We validate it by fitting the parameters of the model to data from [45] and [46] and our own data. All the results show that we are able to explain the nonlinear behavior of the observed shifts along one or two dimensions in color space, which cannot be done using a simple linear model.

This work has been submitted for publication and is available as [29].

4.4. Slow-fast dynamics in Neuroscience

4.4.1. Spike-adding in a canonical three time scale model: superslow explosion & folded-saddle canards

Participants: Mathieu Desroches, Vivien Kirk [University of Auckland, New-Zealand].

We examine the origin of complex bursting oscillations in a phenomenological ordinary differential equation model with three time scales. We show that bursting solutions in this model arise from a Hopf bifurcation followed by a sequence of spike-adding transitions, in a manner reminiscent of spike-adding transitions previously observed in systems with two time scales. However, the details of the process can be much more complex in this three-time-scale context than in two-time-scale systems. In particular, we find that spike-adding can involve canard explosions occurring on two different time scales and is associated with passage near a folded-saddle singularity. We show that the character of the bursting and the form of spike-adding transitions that occur depend on the geometry of certain singular limit systems, specifically the relative positions of the critical and superslow manifolds. We also show that, unlike the case of spike-adding in two-time-scale systems, the onset of a new spike in our model is not typically associated with a local maximum in the period of the bursting oscillation.

This work has been published in SIAM Journal on Applied Dynamical Systems and is available as [14].

4.4.2. Parabolic bursting, spike-adding, dips and slices in a minimal model

Participants: Mathieu Desroches, Jean-Pierre Françoise [LJLL, Sorbonne Université, Paris], Martin Krupa [Université de Nice - LJAD, UCA, Inria MathNeuro].

A minimal system for parabolic bursting, whose associated slow flow is integrable, is presented and studied both from the viewpoint of bifurcation theory of slow-fast systems, of the qualitative analysis of its phase portrait and of numerical simulations. We focus the analysis on the spike-adding phenomenon. After a reduction to a periodically forced 1-dimensional system, we uncover the link with the dips and slices first discussed by J. E. Littlewood in his famous articles on the periodically forced van der Pol system.

This work has been submitted for publication and is available as [25].

4.4.3. Piecewise-linear (PWL) canard dynamics : Simplifying singular perturbation theory in the canard regime using piecewise-linear systems

Participants: Mathieu Desroches, Soledad Fernández-García [University of Sevilla, Spain], Martin Krupa [Université de Nice - LJAD, UCA, Inria MathNeuro], Rafel Prohens [University of the Balearic Islands, Spain], Antonio E. Teruel [University of the Balearic Islands, Spain].

In this work we have gathered recent results on piecewise-linear (PWL) slow-fast dynamical systems in the canard regime. By focusing on minimal systems in $\mathbb{R}^2$ (one slow and one fast variables) and $\mathbb{R}^3$ (two slow and one fast variables), we prove the existence of (maximal) canard solutions and show that the main salient features from smooth systems is preserved. We also highlight how the PWL setup carries a level of simplification of singular perturbation theory in the canard regime, which makes it more amenable to present it to various audiences at an introductory level. Finally, we present a PWL version of Fenichel theorems about slow manifolds, which are valid in the normally hyperbolic regime and in any dimension, which also offers a simplified framework for such persistence results.

This work has been published as a chapter in the book "Nonlinear Systems, Vol. 1: Mathematical Theory and Computational Methods" published by Springer as part of the Understanding Complex Systems book series, and it is available as [22].

4.4.4. Anticipation via canards in excitable systems

Participants: Elif Köksal Ersöz, Mathieu Desroches, Claudio Mirasso [University of the Balearic Islands, Spain], Serafim Rodrigues [Ikerbasque & Basque Center for Applied Mathematics, Spain].

Neurons can anticipate incoming signals by exploiting a physiological mechanism not well understood. This article offers a novel explanation on how a receiver neuron can predict the sender's dynamics in a unidirectionally-coupled configuration, in which both sender-receiver follow the evolution of a multi-scale excitable system. We present a novel theoretical view point based on a mathematical object, called canard, to explain anticipation in excitable systems. We provide a numerical approach, which allows to determine the transient effects of canards. To demonstrate the general validity of canard-mediated anticipation in the context of excitable systems, we illustrate our framework in two examples, a multi-scale radio-wave circuit (the van der Pol model) that inspired a caricature neuronal model (the FitzHugh-Nagumo model) and a biophysical neuronal model (a 2-dimentional reduction of the Hodgkin-Huxley model), where canards act as messengers to the senders' prediction. We also propose an experimental paradigm that would enable experimental neuroscientists to validate our predictions. We conclude with an outlook to possible fascinating research avenues to further unfold the mechanisms underpinning anticipation. We envisage that our approach can be employed to a wider class of excitable systems with appropriate theoretical extensions. Anticipation appears as a counter-intuitive observation in a wide range of dynamical systems ranging from biology to engineering applications. It can occur in unidirectionally coupled systems when the receiver is subject to a self-delayed feedback in addition to a signal coming from the sender. This particular interaction permits the receiver to predict the future trajectory of the sender. Anticipation can occur transiently, thus straightforwardly denoted anticipation, or in long-term dynamics, in which case it is referred to as anticipated synchronization. In this study, we focus on both aspects of anticipatory dynamics in the context of excitable systems and explain it via a counter-intuitive phenomenon, namely canards. Canard trajectories structure the excitability and synchronization properties of multiple timescale systems exhibiting excitable dynamics. By developing a theoretical framework enhanced by numerical continuation, we show that the underlying canard structure in excitable systems is responsible for delaying sub-threshold solutions, but anticipating the spiking ones. We also propose an experimental set up that would enable experimentalists to observe anticipated behavior in neural systems, in particular in type-II neurons.

This work has been accepted for publication in Chaos and is available as [17].

4.4.5. Canard-induced complex oscillations in an excitatory network

Participants: Elif Köksal Ersöz, Mathieu Desroches, Antoni Guillamon [Polytechnic University of Catalunya, Spain], Joel Tabak [University of Exeter, UK].

In this work we have revisited a rate model that accounts for the spontaneous activity in the developing spinal cord of the chicken embryo [50]. The dynamics is that of a classical square-wave burster, with alternation of silent and active phases. Tabak et al. [50] have proposed two different three-dimensional (3D) models with variables representing average population activity, fast activity-dependent synaptic depression and slow activity-dependent depression of two forms. In [47], [48], [49] various 3D combinations of these four variables have been studied further to reproduce rough experimental observations of spontaneous rhythmic activity. In this work, we have first shown the spike-adding mechanism via canards in one of these 3D models from [50] where the fourth variable was treated as a control parameter. Then we discussed how a canard-mediated slow passage in the 4D model explains the sub-threshold oscillatory behavior which cannot be reproduced by any of the 3D models, giving rise to mixed-mode bursting oscillations (MMBOs); see [6]. Finally, we relateed the canard-mediated slow passage to the intervals of burst and silent phase which have been linked to the blockade of glutamatergic or GABAergic/glycinergic synapses over a wide range of developmental stages [49].

This work is in progress and is available as [27].

4.4.6. High-frequency forced oscillations in neuronlike elements

Participants: Denis Zakharov [Institute of Applied Physics RAS, Russia], Martin Krupa [Université de Nice - LJAD, UCA, Inria MathNeuro], Boris Gutkin [Group for Neural Theory, ENS Paris, France], Alexey Kuznetsov [Indiana University - Purdue University Indianapolis, USA].

We analyzed a generic relaxation oscillator under moderately strong forcing at a frequency much greater that the natural intrinsic frequency of the oscillator. Additionally, the forcing is of the same sign and, thus, has a nonzero average, matching neuroscience applications. We found that, first, the transition to high-frequency synchronous oscillations occurs mostly through periodic solutions with virtually no chaotic regimes present. Second, the amplitude of the high-frequency oscillations is large, suggesting an important role for these oscillations in applications. Third, the 1:1 synchronized solution may lose stability, and, contrary to other cases, this occurs at smaller, but not at higher frequency differences between intrinsic and forcing oscillations. We analytically built a map that gives an explanation of these properties. Thus, we found a way to substantially “overclock” the oscillator with only a moderately strong external force. Interestingly, in application to neuroscience, both excitatory and inhibitory inputs can force the high-frequency oscillations.

This work has been published in Physical Review E and is available as [21].

MIMESIS Team

7. New Results

7.1. A Unified Bayesian and Physics-Based Approach for Non-rigid 3D Shape Reconstruction from 2D Images

We developed a method to reconstruct the 3D shape of the interventional device, based on a constrained physics-based simulation combined with 2D monocular fluoroscopic images through a Bayesian filter (see Fig. 4). Whereas the physics-based model provides a prediction of the device shape within the blood vessel, taking into account non-linear interactions between the catheter and the surrounding anatomy adds further information on its current position. In addition, an Unscented Kalman Filter is used to combine the navigation model with the 2D external observations. We focused on a medical application as we believe the method could provide an actual solution to some of the current limitations of fluoroscopy-based procedures. The use of a Bayesian formalism allows for retrieving a good estimate in presence of ambiguous views (i.e. in presence of overlapping anatomies) and to take into account uncertainties on the prediction model (errors in constraint definition, as well as inaccuracies in the mechanical characterization of the catheter) and errors in the external measurements.

The method has been implemented through software developed within the team; for more details see sec. 6.4 and 6.3 . Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario GIMENEZ and Dr. Alain GARCIA from IHU-Strasbourg.

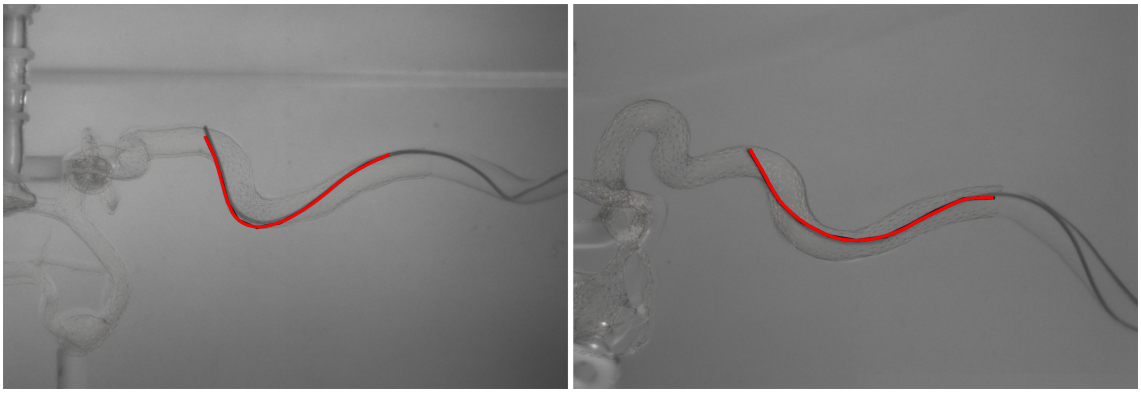


Figure 4. Catheter reconstruction in acquisition view (left) and validation view (right)

- *Authors:* Raffaella Trivisonne and Stéphane Cotin and Erwan Kerrien
- *Type:* PhD Thesis

7.2. Biomechanics-based graph matching for augmented CT-CBCT

Augmenting intraoperative cone beam computed tomography (CBCT) images with preoperative computed tomography (CT) data in the context of image-guided liver therapy is proposed (see Fig. 5). The expected benefit is an improved visualization of tumor(s), vascular system and other internal structures of interest. An automatic elastic registration based on matching of vascular trees extracted from both the preoperative and

intraoperative images is presented. Although methods dedicated to non-rigid graph matching exist, they are not efficient when large intraoperative deformations of tissues occur, as is the case during the liver surgery. First, an improved graph matching algorithm using Gaussian process is introduced by imposing additional constraints during the matching when the number of hypotheses is large; this extended version does not require a manual initialization of matching. Second, a fast biomechanical model is employed to make the method capable of handling large deformations. The proposed automatic intraoperative augmentation is evaluated on both synthetic and real data. It is demonstrated that the algorithm is capable of handling large deformations, thus being more robust and reliable than previous approaches. Moreover, the time required to perform the elastic registration is compatible with the intraoperative navigation scenario. The input data and result of the biomechanics-based graph matching method, which can handle large deformations and augment intraoperative CBCT, is shown in Fig. 5 .

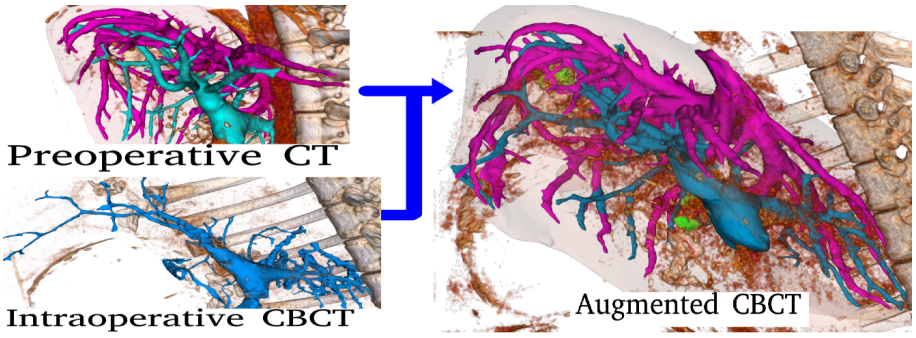


Figure 5. In the top left the preoperative CT image with complete portal and hepatic vessels clearly visible. In the bottom left the intraoperative deform, noisy CBCT image with partial only partial portal vessels visible. In the right the augmented CBCT view with the complete vessels added form preoperative image.

- *Authors:* Jaime Garcia Guevara and Igor Peterlik and Marie-Odile Berger and Stephane Cotin
- *Type:* Journal publication IJCARS 2018

7.3. A Combined Simulation and Machine Learning Approach for Force Classification during Robotized Intravitreal Injections

Intravitreal injection is one of the most common treatment strategies for chronic ophthalmic diseases. The last decade has seen the number of intravitreal injections dramatically increase, and with it, adverse effects and limitations. To overcome these issues, medical assistive devices for robotized injections have been proposed and are projected to improve delivery mechanisms for new generation of pharmacological solutions. In our work, we propose a method aimed at improving the safety features of such envisioned robotic systems. Our vision-based method uses a combination of 2D OCT data, numerical simulation and machine learning to estimate the range of the force applied by an injection needle on the sclera (see Fig. 6). We build a Neural Network (NN) to predict force ranges from Optical Coherence Tomography (OCT) images of the sclera directly. To avoid the need of large training data sets, the NN is trained on images of simulated deformed sclera. We validate our approach on real OCT images collected on five *ex vivo* porcine eyes using a robotically-controlled needle. Results show that the applied force range can be predicted with 94% accuracy. Being real-time, this solution can be integrated in the control loop of the system, allowing for in-time withdrawal of the needle.

- *Authors:* Andrea Mendizabal and Stephane Cotin
- *Type:* Conference publication MICCAI 2018

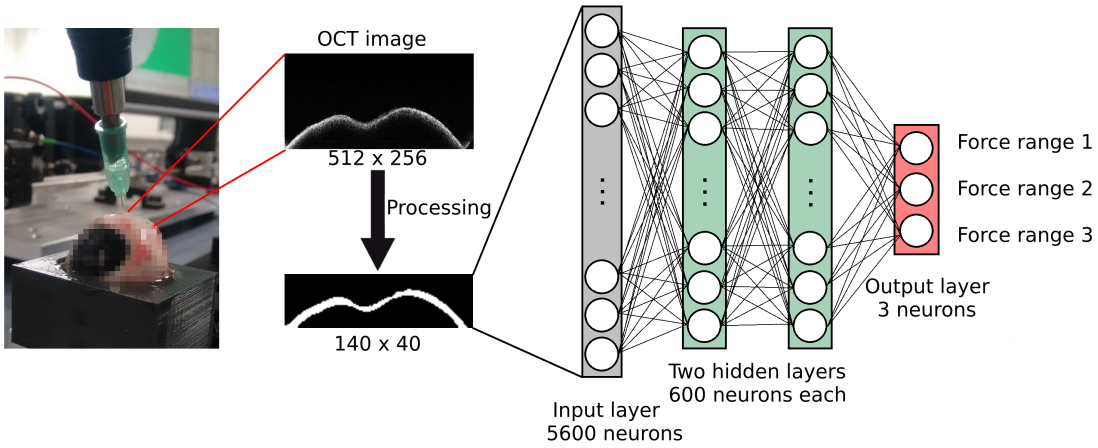


Figure 6. During the robotized intravitreal injection an OCT image of the deformed sclera is collected. The obtained image is processed and given a to a Neural Network that predicts the force range of the force applied by the robotically guided needle.

7.4. Inverse simulation for Robotic control of needle insertion

We recently published in a numerical method allowing for automatic control of a robot during needle insertion procedures (see Fig. 7). Our approach is to develop control models allowing for the correction and prediction of the deformation of structures (needle, tissues or the robot itself) and to adapt the behavior of the robot in order to reach an objective. We showed that inverse steps can be used to control an articulated robot while considering deformations of structures during needle insertion. The method has been used for a needle insertion inside a polyurethane foam using a Mitsubishi RV1A anthropomorphic robot arm. During the insertion vertical and lateral deformations were generated (see Fig. 7) leading to significant modification of the undeformed trajectory, important bending of the needle and even an off-plane shift between the base of the needle and the insertion point. Despite these strong modifications, the method was able to maintain the tip of the needle within the thickness of 1 cm of the foam and followed the desired curved path with accuracy lower than 1 mm without any human intervention.

7.5. Marker-Based Registration for Large Deformations

We proposed an Augmented Reality (AR) system for open liver surgery (see Fig. 8). Although open surgery remains the gold-standard for the treatment of complex tumors and central lesions, technological issues actually prevent using AR with sufficient accuracy for clinical use. We propose a markers-based method allowing for the tracking and the deformation of a preoperative model in real-time during the surgery. Markers are manually placed on the surface of the organ after opening the abdominal cavity, and tracked in real-time by a set of infrared cameras. Our framework is composed of both a non-rigid initial registration method, providing an estimation of the location of the markers in the preoperative model, and a real-time tracking algorithm to deform the model during the surgery (even for large deformation or partial occlusion of the organ). The method is validated on both synthetic and ex-vivo samples; in addition, we demonstrate its applicability in the operating room during a liver resection surgery on a human patient. Preliminary studies provided promising results to improve the location of tumors, and to help surgeons into planning the ideal resection intraoperatively.

- *Authors:* Yinoussa Adagolodjo, Nicolas Golse, Eric Vibert, Michel De Mathelin, Stéphane Cotin, Hadrien Courtecuisse

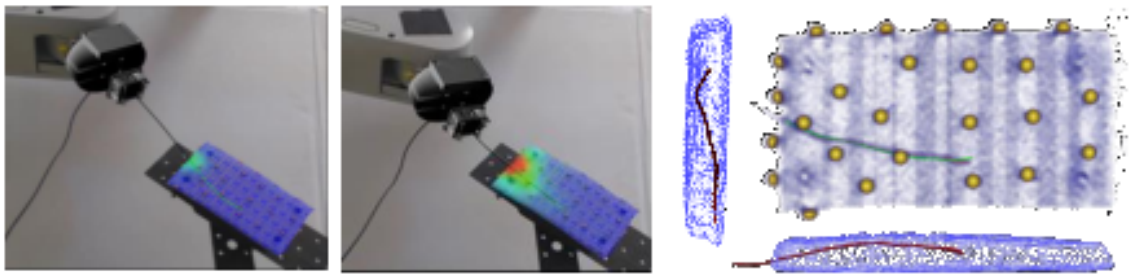


Figure 7. Robotic control based on inverse finite element simulations for needle insertion in deform-able structures. The models are registered in real-time using makers and infrared image-tracking system. We measured an average distance of 1.2 mm between needle's (red) and the desired trajectory (green).

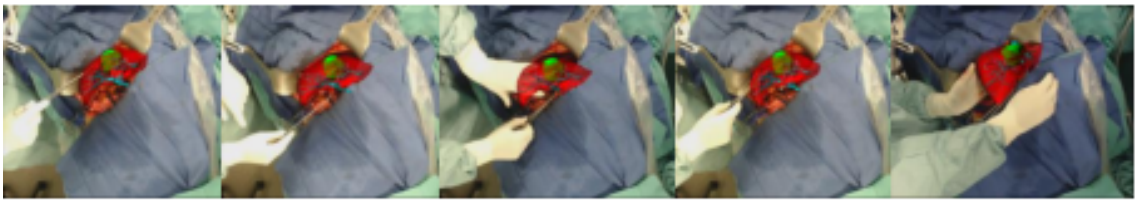


Figure 8. Figure 3: Research prototype for Augmented Reality during open surgery of the liver. We pro-posed a method based on markers for the registration of a preoperative model in real time during surgery. The markers are placed manually on the surface of the body after the opening of the abdominal cavity, and followed in real time by a set of infrared cameras.

- *Type:* PhD Thesis, publication to ICRA

7.6. Automatic and robust 2D/3D registration on fluoroscopic images

We introduce a unified solution to detect, register and track the liver in 2D live fluoroscopic X-ray images, in order to provide augmented reality and guidance during surgery (see Fig. 9). The solution can be decomposed into two phases, with an initial phase to globally estimate the rigid pose through template matching, and a second local rigid refinement step. A main contribution lies in the combination, for the pose refinement step, of intensity and contour based features over the contrasted vessels of the liver and surrounding organs, by integrating corresponding visual cues in a local optimization framework with respect to the pose. The method does not need any 2D segmentation of the contrasted vessels but relies on a synthetic X-ray rendering algorithm, and requires very few assumptions or priors. Our solution has been tested on synthetic and porcine data, showing its efficiency on realistic scenarios.

A non-rigid registration technique to account for local deformations of the target is also investigated. Once the model is rigidly aligned, local estimation of the deformations undergone by the vasculature and the parenchyma, given a linear or volumetric elastic deformation model of the vessels and the parenchyma, driven by local optical flow features.

The method has been implemented through software developed within the team; for more details see sec. 6.4 -6.3 . Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario GIMENEZ and Dr. Alain GARCIA, Pr. Federico Davrieux, and Pim Hendriks, Daan Kuppens and from IHU-Strasbourg.

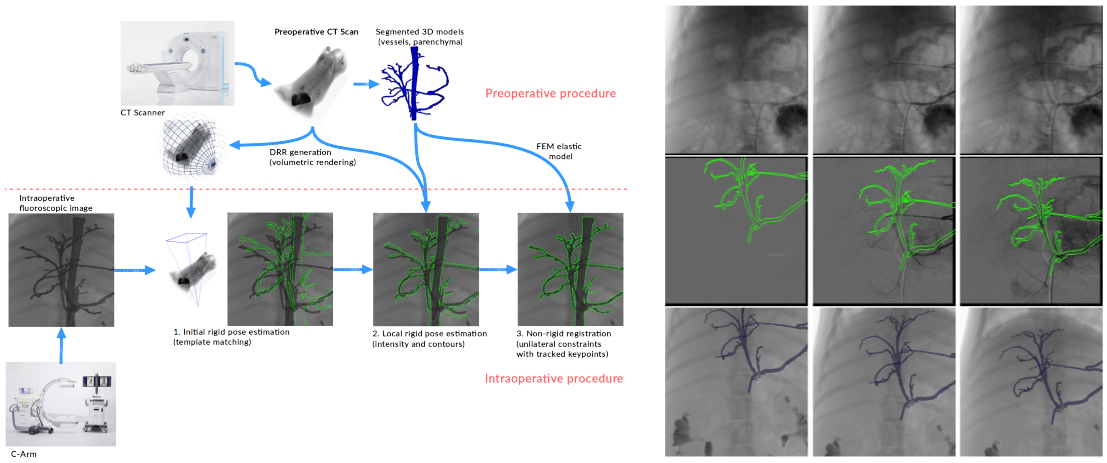


Figure 9. Pipeline of the system and results of the rigid registration system.

- *Authors:* Antoine Petit, Bruno Marques, Stéphane Cotin
- *Type:* Submission to IPCAI 2019

7.7. Capturing Deformations of Interacting Soft Objects Using RGB-D Data

We present a method for tracking multiple interacting deformable objects undergoing rigid motions, elastic deformations and contacts, using images and point cloud data provided by an RGB-D sensor (see Fig. 10). A joint registration framework is proposed, based on physical Finite Element Method (FEM) elastic and interaction models. It first relies on a visual segmentation of the considered objects in the RGB images.

The different segmented point clouds are then processed to estimate rigid transformations with on an ICP algorithm, and to determine geometrical point-to-point correspondences with the meshes. External forces resulting from these correspondences and between the current and the rigidly transformed mesh can then be derived. It provides both non-rigid and rigid data cues. Classical collision detection and response model is also integrated, giving contact forces between the objects. The deformations of the objects are estimated by solving a dynamic system balancing these external and contact forces with the internal or regularization forces computed through the FEM elastic model. This approach has been here tested on different scenarios involving two or three interacting deformable objects of various shapes, with promising results.

A case study in open surgery on the liver has also been investigated. Yet in this case a major improvement in the accuracy of the registration is provided by the integration of anatomical shape constraints, which are naturally hidden from the RGB-D camera, and that we account for through a registration with the pre-operative CT data. With a comparative study, we demonstrate the relevance of our method in a real-world application mimicking an open surgery scenario where the liver has to be tracked to provide an augmented reality view.

The method has been implemented through software developed within the team, especially the RGBDTracking plugin. 6.4 and 6.3 . Validation has been performed on porcine in-vivo data, acquired in accordance with UE norms, in collaboration with Pr. Mario GIMENEZ and Dr. Alain GARCIA, Pr. Federico Davrieux, and Pim Hendriks, Daan Kuppens and from IHU-Strasbourg.

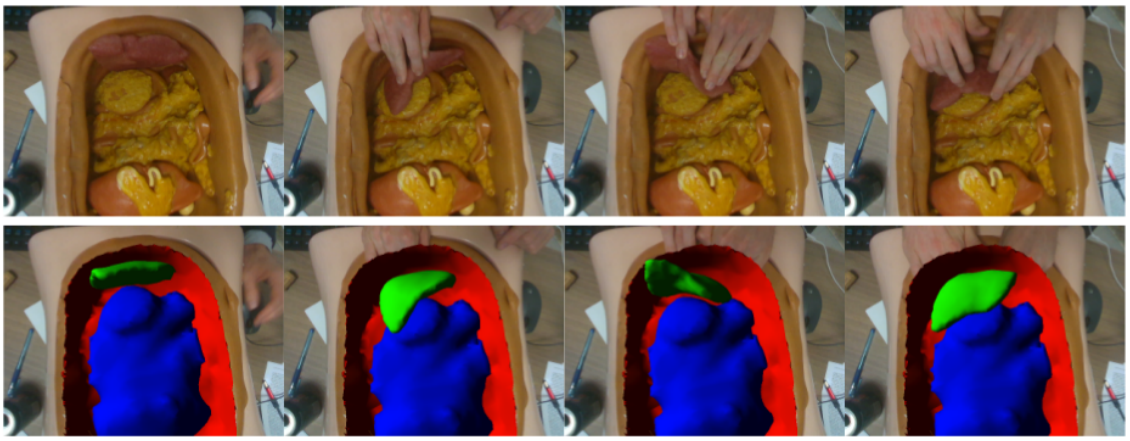


Figure 10. Pipeline of the system and results of the rigid registration system.

- *Authors:* Antoine Petit, Stéphane Cotin
- *Type:* Publications to IROS 2018 and ACCV Workshops 2018

MNEMOSYNE Project-Team

7. New Results

7.1. Overview

This year we have explored the main cortico-basal loops of the cerebral architecture 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 associative loop (*cf.* § 7.3) is about mechanisms of working memory. Concerning the motor loop (*cf.* § 7.4), we have studied mechanisms of song acquisition and production in birds.

We have also worked on the systemic integration of our models (*cf.* § 7.5), raising the question of the conditions of autonomous learning and certain global characteristics such as the transition from goal-directed to habitual behaviors.

Finally, we study the links between our bio-inspired modeling work and Machine Learning (*cf.* § 7.6), revisiting this latter domain in the light of the principles highlighted by our models.

7.2. The limbic loop

We have continued to explore the mnemonic characteristics of the main information flows and cerebral structures interacting with the limbic loop. This is the case with the main learning paradigms associated to the involvement of the amygdala in respondent conditioning [16]. We have also continued our work on the hippocampus and particularly its function of pattern separation [7].

7.3. The associative loop

The prefrontal cortex is known to be involved in many high-level cognitive functions, in particular working memory. In [23], we studied to what extent a group of randomly connected units (namely an Echo State Network, ESN) can store and maintain (as output) an arbitrary real value from a streamed input, i.e. can act as a sustained working memory unit. Furthermore, we explored to what extent such an architecture can take advantage of the stored value in order to produce non-linear computations. Comparison between different architectures (with and without feedback, with and without a working memory unit) shows that an explicit memory improves the performances. This property of the model can be considered as a gated memory: a value enters the memory at the moment of the (input) trigger and is kept constant in face of incoming distractors (the continuous streamed input).

7.4. The motor loop

Sensorimotor learning represents a challenging problem for artificial and natural systems. Several computational models try to explain the neural mechanisms at play in the brain to implement such learning. These models have several common components: a motor control model, a sensory system and a learning architecture. In S. Pagliarini's PhD, our challenge is to build a biologically plausible model for song learning in birds including neuro-anatomical and developmental constraints.

We made a review on a specific type of sensorimotor learning referred to as imitative vocal learning and exemplified by song learning in birds or human complex vocalizations. We aim to compare the various approaches used in existent sensorimotor models relevant for our purpose and to place them in a common framework. We propose a bio-inspired model for an imitative sensorimotor learning, which aims at building a map between the sensory representations of gestures (sensory targets) and their underlying motor pattern through random exploration of the motor space. An example of such learning process occurs during vocal learning in humans or birds, when young subjects babble and learn to copy previously heard adult vocalizations. Previous work (by Hahnloser and Ganguli) has suggested that a simple Hebbian learning rule allows perfect imitation when sensory feedback is a purely linear function of the motor pattern underlying movement production. We aim at generalizing this model to the more realistic case where sensory responses are sparse and non-linear. To this end, we explore the performance of various learning rules and normalizations. Importantly, the proposed model is robust whatever normalization is chosen. We showed that both the imitation quality and the convergence time are highly dependent on the sensory selectivity and dimension of the motor representation.

On this topic, X. Hinaut is also collaborating with Catherine del Negro's team (CNRS, University of Paris-Sud, Orsay) on the representation of syntax in songbird brains. In particular, the project aims at (1) linking the neural activity of a sensori-motor area (HVC) to syntax elements in the songs of domestic canaries ; (2) analysing the audio files and transcripts of canary songs in order to find syntax cues and higher order representations (graph properties of songs, evaluate Markovian forward and backward transition probabilities of various orders). The song analyses part has been performed with the help of a L3 intern, Pierre Marcus.

7.5. Systemic integration

This year, we have considered characteristics of interactions of cortico-basal loops [32], firstly to continue the development of a software environment based on the Minecraft videogame allowing for the survival behavior of an autonomous agent [25], [24] and secondly to revisit the principles of habits formation.

The dorsal pallium (a.k.a. the cortex in the mammals) makes a large loop circuit with the basal ganglia and the thalamus known to control and adapt behavior but the who's who of the functional roles of these structures is still debated. Influenced by the Triune brain theory that was proposed in the early sixties, many current theories propose a hierarchical organization on the top of which stands the cortex to which the subcortical structures are subordinated. In particular, habits formation has been proposed to reflect a switch from conscious on-line control of behavior by the cortex, to a fully automated subcortical control. We have proposed in [3] instead to revalue the function of the network in light of the current experimental evidence concerning the anatomy and physiology of the basal ganglia-cortical circuits in vertebrates.

This theory is supported by a model [11] that includes interactions between the cortex, the basal ganglia and the thalamus based on a dual competition. We hypothesize that the striatum, the subthalamic nucleus, the internal pallidum (GPi), the thalamus, and the cortex are involved in closed feedback loops through the hyperdirect and direct pathways. These loops support a competition process that results in the ability of basal ganglia to make a cognitive decision followed by a motor decision. Considering lateral cortical interactions, another competition takes place inside the cortex allowing the latter to make a cognitive and a motor decision. We show how this dual competition endows the model with two regimes. One is driven by reinforcement learning, the other by Hebbian learning. The final decision is made according to a combination of these two mechanisms with a gradual transfer from the former to the latter. We confirmed these theoretical results on primates (Macaca mulata) using a novel paradigm predicted by the model.

7.6. Machine Learning

In this section, we report on some neuronal adaptive mechanisms, that we develop at the frontier between Machine Learning and Computational Neuroscience. Our goal is to consider and adapt models in Machine Learning for their integration in a bio-inspired framework.

Concerning the manipulation of temporal sequences, we have proposed an original algorithm for the extraction of sequences from LSTM, a major class of recurrent neural networks [1]. These sequences are then interpreted as rules representing implicit knowledge within electrical diagrams (*cf.* § 8.1).

Concerning our work on reservoir computing, X. Hinaut is collaborating with Michael Spranger (Sony Lab, Tokyo, Japan) on grounding of language, adapting Hinaut's previous Reservoir Language Model (RLM) with the representational system of Spranger: IRL (Incremental Recruitment Language). He is also collaborating with Hamburg on the use of reservoir models for robotic tasks (*cf.* § 9.3). In this work, we have shown that the RLM can successfully learn to parse sentences related to home scenarios in fifteen languages [6]. This demonstrates that (1) the learning principle of our model is not limited to a particular language (or particular sentence structures), and (2) it can deal with various kinds of representations (not only predicates), which enable users to adapt it to their own needs.

Regarding the extraction of characteristics from and the use of hierarchical networks, as in the case of deep networks, we have been able to consider how to deal with not-so-big data sets, and target the notion of interpretability of the obtained results which is a key issue: since deep learning applications are increasingly present in the society, it is important that the underlying processes be accessible and understandable to every one. In order to target these challenges, we have analyzed how considering prototypes in a rather generalized sense (with respect to the state of the art) allows to reasonably work with small data sets while providing an interpretable view of the obtained results. Some mathematical interpretation of this proposal have also been discussed. Sensitivity to hyperparameters is a key issue for reproducible deep learning results, and has been carefully considered in our methodology. Performances and (even more interesting, in a sense) limitations of the proposed setup have been explored in details, under different hyperparameters sets, in an analogous way as biological experiments are conducted. We obtain a rather simple architecture, easy to explain, and which allows, combined with a standard method, to target both performances and interpretability [4].

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Axel Hutt Amélie Aussel, Nathalie Azevedo Carvalho.
In collaboration with Radu Ranta (univ. Lorraine), Dominique Martinez (CNRS), Abderrahman Iggidr (Inria), Patrick Hénaff (univ. Lorraine), Beate Knauer and Motoharu Yoshida (Ruhr university) and LieJune Shiau (university of Houston)

7.1.1. Memory & sleep

We proposed a detailed anatomical and mathematical model of the hippocampal formation for the generation of sharp-wave ripples and theta-nested gamma oscillations [1], [7]. Indeed, the mechanisms underlying the broad variety of oscillatory rhythms measured in the hippocampus during the sleep-wake cycle are not yet fully understood. We proposed a computational model of the hippocampal formation based on a realistic topology and synaptic connectivity, and we analyzed the effect of different changes on the network, namely the variation of synaptic conductances, the variations of the CAN channel conductance and the variation of inputs. By using a detailed simulation of intracerebral recordings, we showed that this model is able to reproduce both the theta-nested gamma oscillations that are seen in awake brains and the sharp-wave ripple complexes measured during slow-wave sleep. The results of our simulations support the idea that the functional connectivity of the hippocampus, modulated by the sleep-wake variations in Acetylcholine concentration, is a key factor in controlling its rhythms. A presentation of this work received a best poster award at the 27th annual Computational Neuroscience Meeting, CNS'2018.

7.1.2. Parkinson's network

Using a Hodgkin and Huxley's model, we modeled pathological oscillations of Parkinson's disease in basal ganglia. Our hypothesis was that the pathological oscillations are generated by a MSN-GPeA-FSN circuit and then transferred to the STN by the GPeP. The normal state is represented by neurons of the MSN emitting at a frequency of $\sim 1Hz$ and the parkinsonian state is represented by MSN neurons that emit at a frequency of $\sim 15Hz$. Our results correspond to the experimental results of the rat. In the normal state, there is no visible synchronization, whereas in the parkinsonian state pathological synchronizations are formed at the level of the circuit. There is even a rhythm that is created, that is to say, the neurons of MSN emit first then those of the FSN and then those of the GPeA and so on. We performed large-scale simulations of 1.5 million neurons in the basal ganglia in rats using the Grid5000 (parallel computation platform).

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Sébastien Rimbert, Oleksii Avilov, Rahaf Al-Chwa.
In collaboration with Stéphanie Fleck (Univ. Lorraine) and Patrick Hénaff (univ. Lorraine)

7.2.1. Motor system

In collaboration with Stéphanie Fleck (Univ. Lorraine)
Kinesthetic motor imagery (KMI) tasks induce brain oscillations over specific regions of the primary motor cortex within the contralateral hemisphere of the body part involved in the process. This activity can be measured through the analysis of electroencephalographic (EEG) recordings and is particularly interesting for Brain-Computer Interface (BCI) applications.

7.2.1.1. Electroencephalographic modulations during an open- or closed-eyes motor task

There is fundamental knowledge that during the resting state cerebral activity recorded by electroencephalography (EEG) is strongly modulated by the eyes-closed condition compared to the eyes-open condition, especially in the occipital lobe. However, little research has demonstrated the influence of the eyes-closed condition on the motor cortex, particularly during a self-paced movement. This prompted the question: How does the motor cortex activity change between the eyes-closed and eyes-open conditions? To answer this question, we recorded EEG signals from 15 voluntary healthy subjects who performed a simple motor task (i.e., a voluntary isometric flexion of the right-hand index) under two conditions: eyes-closed and eyes-open. Our results confirmed strong modulation in the mu rhythm (7–13 Hz) with a large event-related desynchronisation. However, no significant differences have been observed in the beta band (15–30 Hz). Furthermore, evidence suggests that the eyes-closed condition influences the behaviour of subjects [5]. Our study gives greater insight into the motor cortex and could also be useful to improve brain-computer interface (BCI) based on motor imagery.

7.2.1.2. Can a Subjective Questionnaire be used as a Brain-Computer Interface performance Predictor?

Predicting a subject's ability to use a Brain Computer Interface (BCI) is one of the major issues in the BCI domain. Relevant applications of forecasting BCI performance include: the ability to adapt the BCI to the needs and expectations of the user; assessing the efficiency of BCI use in stroke rehabilitation; and finally, homogenizing a research population. A limited number of recent studies have proposed the use of subjective questionnaires, such as, the Motor Imagery Questionnaire Revised-Second Edition (MIQ-RS). However, further research is necessary to confirm the effectiveness of this type of subjective questionnaire as a BCI performance estimation tool. In this study we aim to answer the following questions: can the MIQ-RS be used to estimate the performance of an MI-based BCI? If not, can we identify different markers that could be used as performance estimators? To answer these questions, we recorded EEG signals from 35 voluntary healthy subjects during BCI use. The subjects previously had completed the MIQ-RS questionnaire. We conducted an offline analysis to assess the correlation between the questionnaire scores related to Kinesthetic and Motor imagery tasks and the performances of four classification methods. Our results show no significant correlation between BCI performance and the MIQ-RS scores. However, we reveal that BCI performance is correlated to habits and frequency of practicing manual activities [15] (accepted in Front. Hum. Neurosci. | doi: 10.3389/fnhum.2018.00529).

7.2.1.3. Median nerve stimulation based BCI: a new approach to detect intraoperative awareness during general anesthesia

Hundreds of millions of general anesthesia are performed each year on patients all over the world. Among these patients, 0.1-0.2% are victims of Accidental Awareness during General Anesthesia (AAGA), i.e. an unexpected awakening of the patient during a surgical procedure under general anesthesia. This terrifying experience may be very traumatic for the patient and should be avoided by the anesthesiologists. Out of all the techniques used to prevent these awakenings, there is currently no solution based on the EEG signal to detect this phenomenon efficiently. Since the first reflex for a patient during an AAGA is to move, a passive BCI based on the intention of movement is conceivable. However, the challenge of using such BCI is that the intention to move from the waking patient is not initiated by a trigger that could be used to guide a classifier. We proposed a solution based on Median Nerve Stimulation (MNS), which causes specific modulations in the motor cortex and can be altered by an intention of movement. We showed that MNS may provide a foundation for an innovative BCI that would allow the detection of an AAGA [17].

PARIETAL Project-Team

7. New Results

7.1. Reducing the number of samples in spatiotemporal dMRI acquisition design

Acquisition time is a major limitation in recovering brain white matter microstructure with diffusion magnetic resonance imaging. The aim of this work is to bridge the gap between growing demands on spatio-temporal resolution of diffusion signal and the real-world time limitations. We introduce an acquisition scheme that reduces the number of samples under adjustable quality loss. Finding a sampling scheme that maximizes signal quality and satisfies given time constraints is NP-hard. Therefore, a heuristic method based on genetic algorithm is proposed in order to find sub-optimal solutions in acceptable time. The analyzed diffusion signal representation is defined in the qτ space, so that it captures both spacial and temporal phenomena. The experiments on synthetic data and in vivo diffusion images of the C57Bl6 wild-type mouse corpus callosum reveal the superiority of the proposed approach over random sampling and even distribution in the qτ space. The use of genetic algorithm allows to find acquisition parameters that guarantee high signal reconstruction accuracy under given time constraints. In practice, the proposed approach helps to accelerate the acquisition for the use of q-dMRI signal representation.

More information can be found in [12]

7.2. Robust EEG-based cross-site and cross-protocol classification of states of consciousness

Determining the state-of-consciousness in patients with disorders-of-consciousness (DOC) is a challenging practical and theoretical problem. Recent findings suggest that multiple markers of brain activity extracted from the electroencephalogram (EEG) may index the state of consciousness in the human brain. Furthermore, machine learning has been found to optimize their capacity to discriminate different states of consciousness in clinical practice. However, it is unknown how dependable these EEG-markers are in the face of signal variability due to different EEG-configurations, EEG-protocols and subpopulations from different centers encountered in practice. In our recent paper [11] we addressed the following questions: What is the impact of the EEG configuration (selection of sensors, duration of EEG used)? Do models based on current EEG-markers achieve prospective generalization on independent data from other EEG protocols and other hospitals? Are single markers sufficiently powerful and when does multivariate classification provide the clearest advantage? For summary of methods and approahc see Figure4 . Our results highlight the effectiveness of classical well-studied EEG-signatures such as alpha [8-12Hz] and theta [5-7Hz] frequency band oscillations for detecting consciousness when combined with machine learning. While univariate predictive models achieved good performance, multivariate models showed better generalization capacity and increased robustness to different types of noise while mitigating the impact of the EEG-configuration. Our findings suggest that pooling data over multiple centers for predictive modeling of DOC is a concrete possibility and can become a promising alley for the field of cognitive neurology.

7.3. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series

Sleep stage classification constitutes an important preliminary exam in the diagnosis of sleep disorders. It is traditionally performed by a sleep expert who assigns to each 30 s of signal a sleep stage, based on the visual inspection of signals such as electroencephalograms (EEG), electrooculograms (EOG), electrocardiograms (ECG) and electromyograms (EMG). We introduce here the first deep learning approach for sleep stage

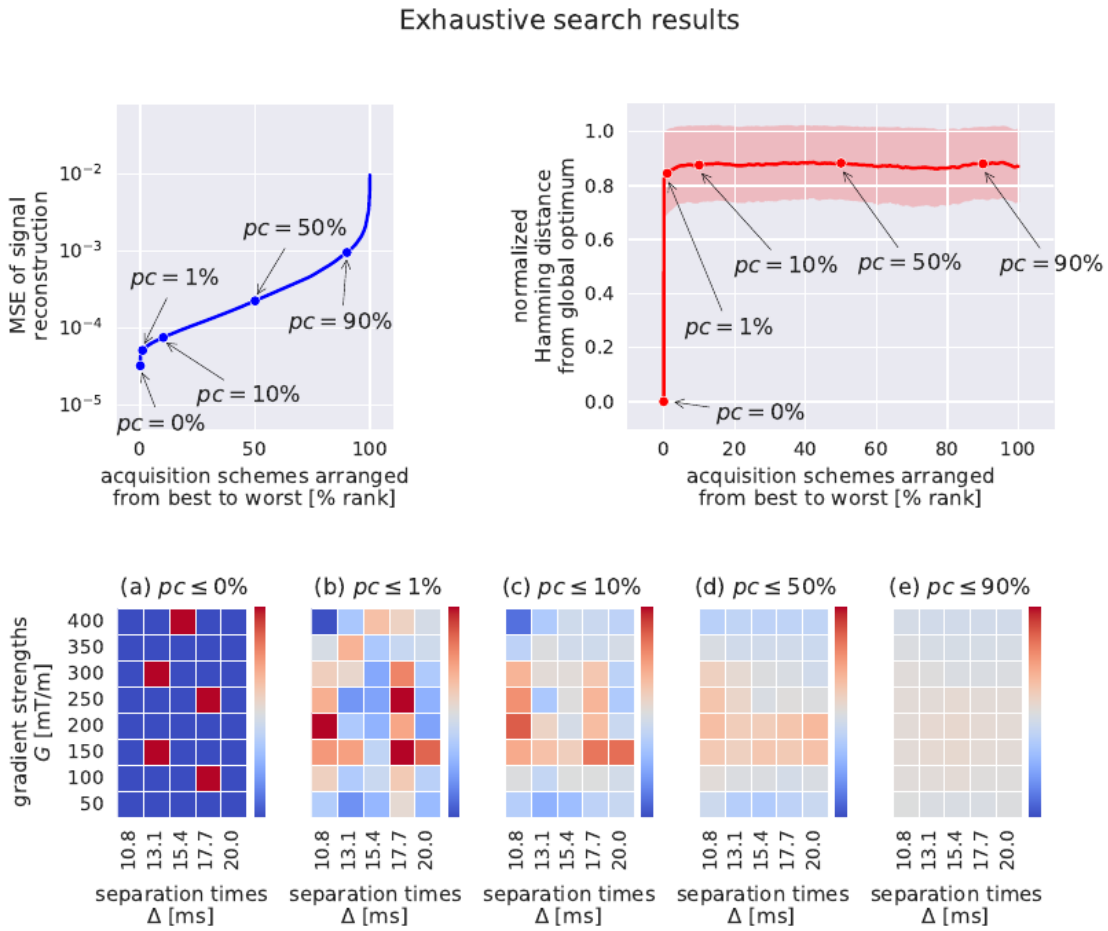


Figure 3. Exhaustive search results of the optimization by shells for the in silico experiment with nmax = 100. The plots at the top present all the 658,008 feasible acquisition schemes arranged from best to worst, illustrating the mean squared errors (MSEs) of signal reconstruction (top-left plot) and the normalized Hamming distances from the global optimum ± 1 standard deviation (top-right). In order to visualize the analyzed (G, Δ) parameter space, the percentiles pc = 0%, 1%, 10%, 50%, 90% are annotated on both plots, showing respectively the global optimum, the top 1% solutions, the top 10% solutions, etc. The corresponding cumulative averages of acquisition schemes are depicted in the heat maps at the bottom. The colors reflect the likelihood of a given (G, Δ) pair in the scheme. The heat maps for pc ≤ 0% and pc ≤ 1% represent, respectively, the global optimum and its proximity. The interval between pc = 10% and pc = 90% contains a huge spectrum of schemes with similar MSEs and almost equally large distances from the global optimum.

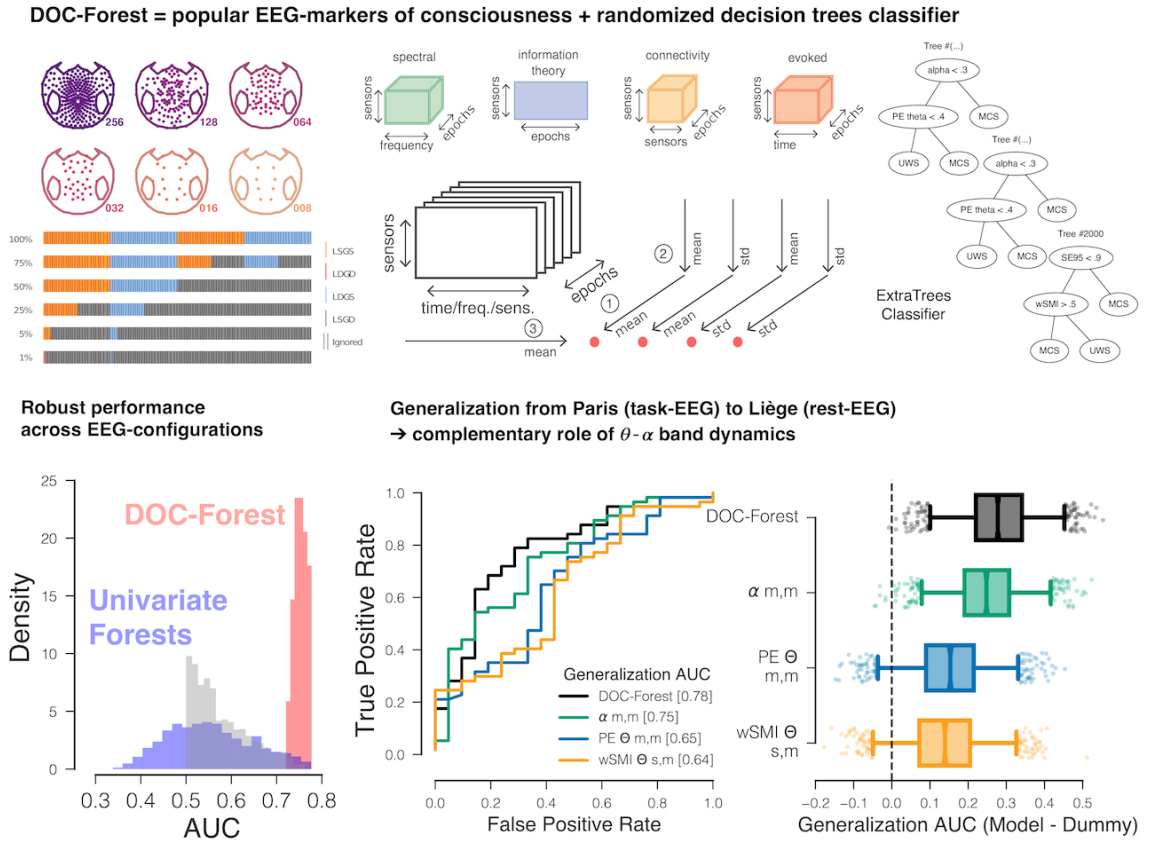


Figure 4. We probed the robustness and validity of EEG-markers of consciousness. Using the robust Extra-Trees algorithm (Geurts, Ernst, & Wehenkel, 2006) we developed a classifier trained to differentiate UWS from MCS patients. This classifier (named "DOC-forest") was trained and tested using 28 potential EEG-markers of consciousness (112 features) from 249 patients recorded at the Paris Pitié-Salpêtrière and 78 patients from the University Hospital of Liège. We used the MNE-Python software for EEG processing and the scikit-learn package for machine learning. Our results show that optimally combining multiple EEG-markers of states of consciousness using machine learning enables robust generalization across EEG-configurations, EEG-protocols and sites. Our recipe for extracting biomarkers is available on Github: https://nice-tools.github.io/nice. For a neuroscientific discussion of our work see the accompanying commentary article by Sokoliuk and Cruse (https://doi.org/10.1093/brain/awy267).

classification that learns end-to-end without computing spectrograms or extracting hand-crafted features, that exploits all multivariate and multimodal Polysomnography (PSG) signals (EEG, EMG and EOG), and that can exploit the temporal context of each 30 s window of data. For each modality the first layer learns linear spatial filters that exploit the array of sensors to increase the signal-to-noise ratio, and the last layer feeds the learnt representation to a softmax classifier. Our model is compared to alternative automatic approaches based on convolutional networks or decisions trees. Results obtained on 61 publicly available PSG records with up to 20 EEG channels demonstrate that our network architecture yields state-of-the-art performance. Our study reveals a number of insights on the spatio-temporal distribution of the signal of interest: a good trade-off for optimal classification performance measured with balanced accuracy is to use 6 EEG with 2 EOG (left and right) and 3 EMG chin channels. Also exploiting one minute of data before and after each data segment offers the strongest improvement when a limited number of channels is available. As sleep experts, our system exploits the multivariate and multimodal nature of PSG signals in order to deliver state-of-the-art classification performance with a small computational cost.

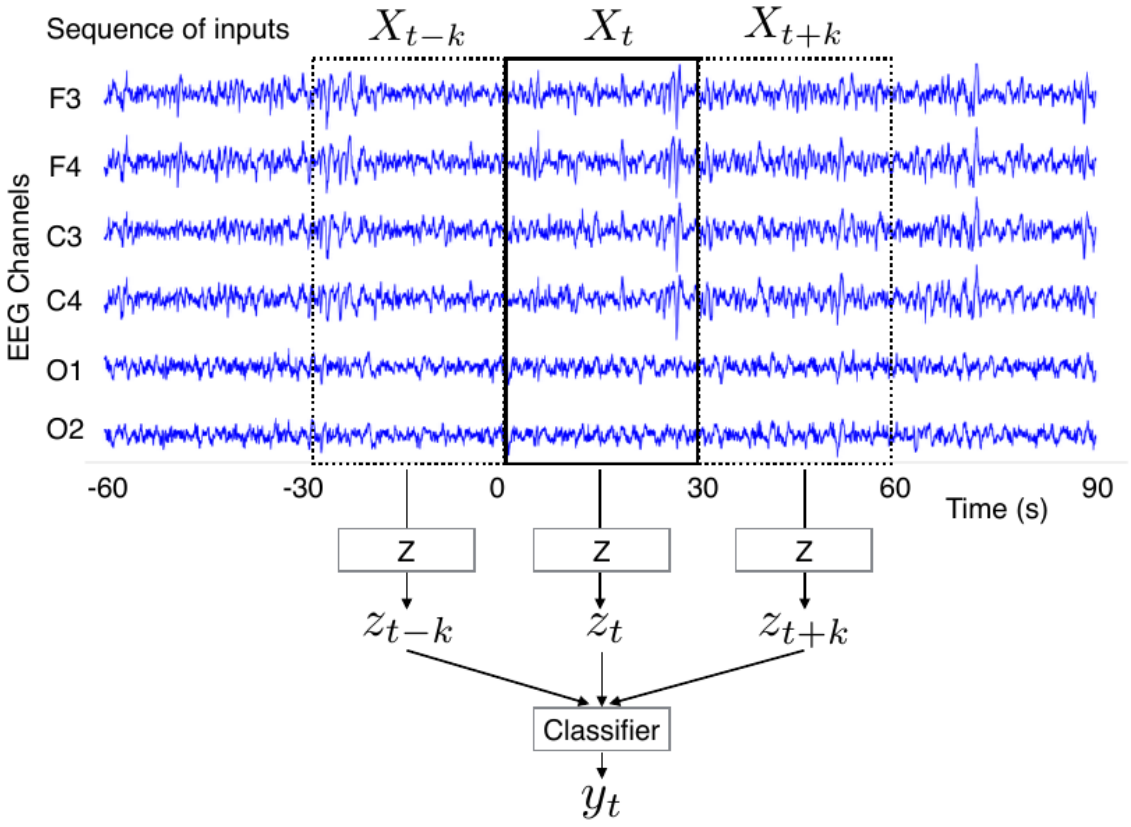


Figure 5. Time distributed architecture to process a sequence of inputs $S_t^k = \{X_{t-k}, \cdots, X_t, \cdots, X_{t+k}\}$ with k = 1. X_k stands for the multivariate input data over 30 s that is fed into the feature extractor Z. Features are extracted from consecutive 30 s samples: X_{t-k}, ..., X_t, ..., X_{t+k}. Then the obtained features are aggregated $[z_{t-k}, \cdots, z_t, \cdots, z_{t+k}]$. The resulting aggregation of features is finally fed into a classifier to predict the label y_t associated with the sample X_t.

More information can be found in [8].

7.4. Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping

Functional Magnetic Resonance Imaging (fMRI) has furthered brain mapping on perceptual, motor, as well as higher-level cognitive functions. However, to date, no data collection has systematically addressed the functional mapping of cognitive mechanisms at a fine spatial scale. The Individual Brain Charting (IBC) project stands for a high-resolution multi-task fMRI dataset that intends to provide the objective basis toward a comprehensive functional atlas of the human brain. The data refer to a cohort of 12 participants performing many different tasks. The large amount of task-fMRI data on the same subjects yields a precise mapping of the underlying functions, free from both inter-subject and inter-site variability. The present article gives a detailed description of the first release of the IBC dataset. It comprises a dozen of tasks, addressing both low- and high-level cognitive functions. This openly available dataset is thus intended to become a reference for cognitive brain mapping.

More information can be found in [25]

7.5. Atlases of cognition with large-scale brain mapping

To map the neural substrate of mental function, cognitive neuroimaging relies on controlled psychological manipulations that engage brain systems associated with specific cognitive processes. In order to build comprehensive atlases of cognitive function in the brain, it must assemble maps for many different cognitive processes, which often evoke overlapping patterns of activation. Such data aggregation faces contrasting goals: on the one hand finding correspondences across vastly different cognitive experiments, while on the other hand precisely describing the function of any given brain region. Here we introduce a new analysis framework that tackles these difficulties and thereby enables the generation of brain atlases for cognitive function. The approach leverages ontologies of cognitive concepts and multi-label brain decoding to map the neural substrate of these concepts. We demonstrate the approach by building an atlas of functional brain organization based on 30 diverse functional neuroimaging studies, totaling 196 different experimental conditions. Unlike conventional brain mapping, this functional atlas supports robust reverse inference: predicting the mental processes from brain activity in the regions delineated by the atlas. To establish that this reverse inference is indeed governed by the corresponding concepts, and not idiosyncrasies of experimental designs, we show that it can accurately decode the cognitive concepts recruited in new tasks. These results demonstrate that aggregating independent task-fMRI studies can provide a more precise global atlas of selective associations between brain and cognition.

More information can be found in [28].

7.6. Celer: a Fast Solver for the Lasso with Dual Extrapolation

Convex sparsity-inducing regularizations are ubiquitous in high-dimensional machine learning, but solving the resulting optimization problems can be slow. To accelerate solvers, state-of-the-art approaches consist in reducing the size of the optimization problem at hand. In the context of regression, this can be achieved either by discarding irrelevant features (screening techniques) or by prioritizing features likely to be included in the support of the solution (working set techniques). Convex duality comes into play at several steps in these techniques. Here, we propose an extrapolation technique starting from a sequence of iterates in the dual that leads to the construction of improved dual points. This enables a tighter control of optimality as used in stopping criterion, as well as better screening performance of Gap Safe rules. Finally, we propose a working set strategy based on an aggressive use of Gap Safe screening rules. Thanks to our new dual point construction, we show significant computational speedups on multiple real-world problems compared to alternative state-of-the-art coordinate descent solvers.

More information can be found in [54]. Code can be found at https://mathurinm.github.io/celer/.

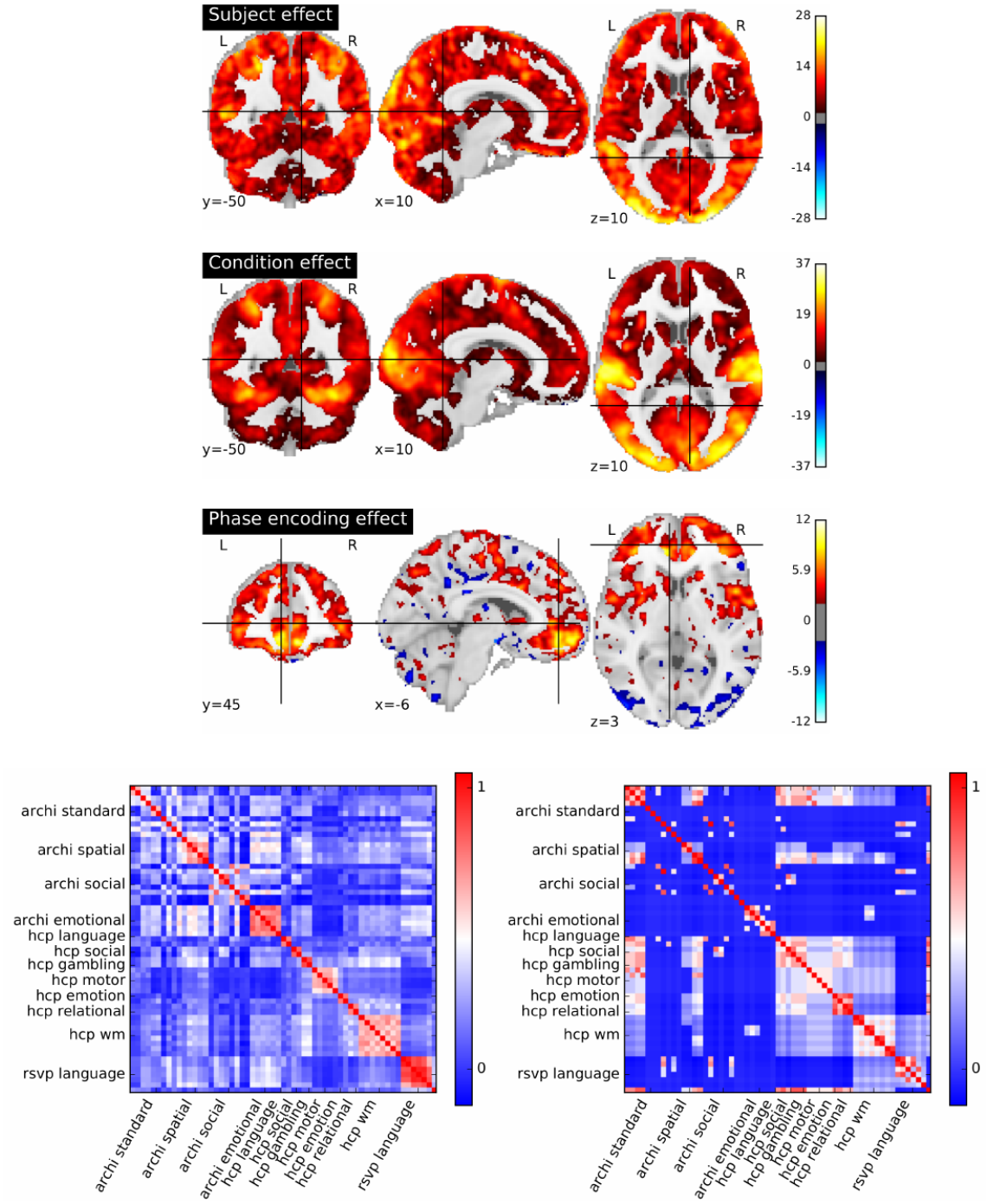


Figure 6. Overview of information conveyed by activation maps resulting from a first-level analysis. (top) Global effects of experimental subject condition, and phase-encoding direction. A per-voxel ANOVA breaks the variance of the set of brain maps into subject, experimental condition, and phase-encoding direction values. All maps are given in z-scale and thresholded at an FDR level of 0.05. (Bottom) Focusing on condition effect, the similarity between condition-related maps, averaged across subjects (left) is clearly related to the dissimilarity of the conditions, when these are characterized in terms of the Cognitive Atlas (right).

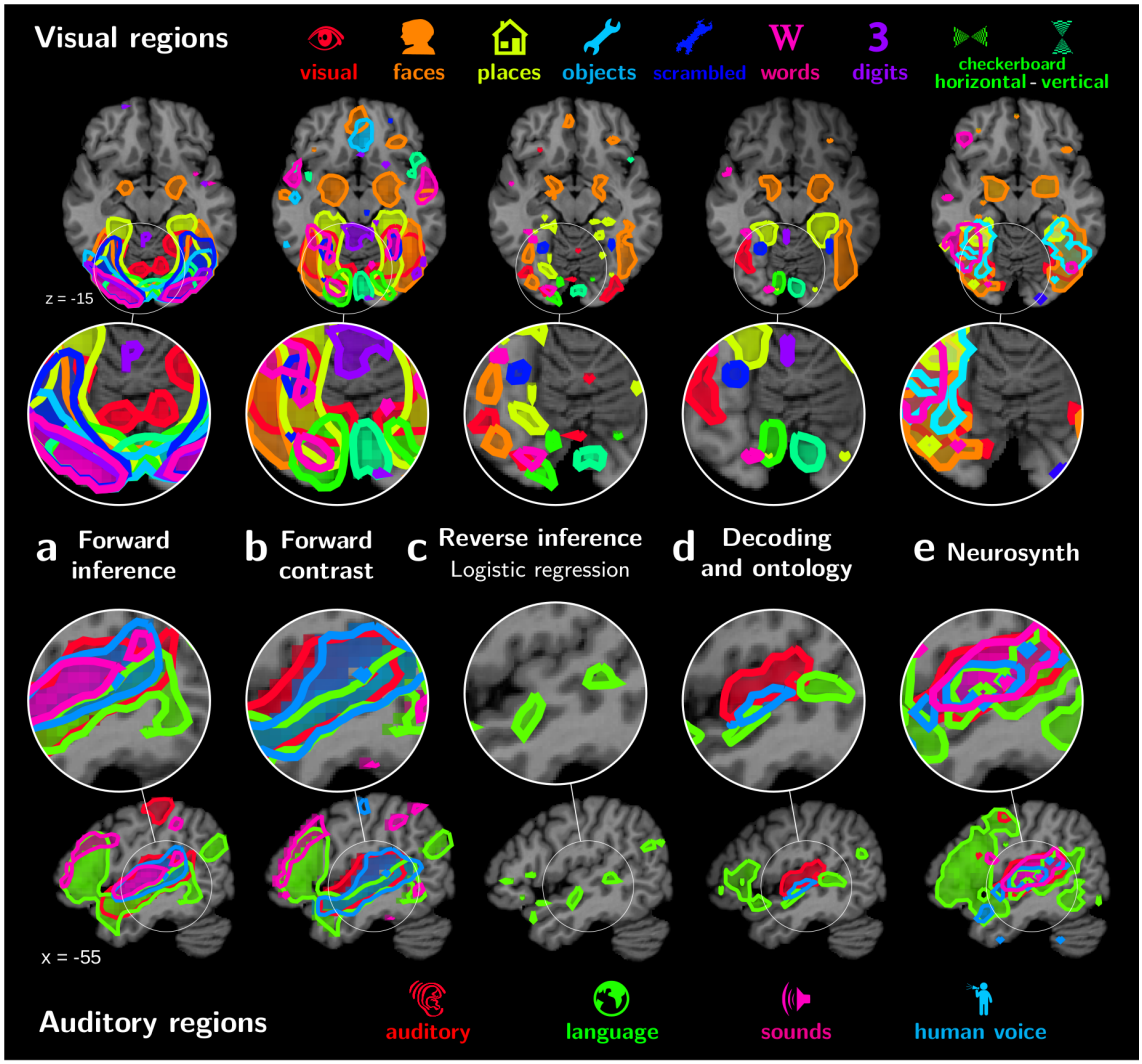


Figure 7. Different functional atlases – Regions outlined using different functional mapping approaches, from left to right: a. forward term mapping; b. forward inference with ontology contrasts (standard analysis); c. reverse inference with logistic regression; d. NeuroSynth reverse inference; and e. our approach, mapping with decoding and an ontology. The top part shows visual regions, and the lower one auditory regions in the left hemisphere. Forward term mapping outlines overlapping regions, as brain responses capture side effects such as the stimulus modality: for visual and auditory regions every cognitive term is represented in the corresponding primary cortex. Forward mapping using contrasts removes the overlap in primary regions, but a large overlap persists in mid-level regions, as control conditions are not well matched across studies. Standard reverse inference, specific to a term, creates overly sparse regions though with little overlap. Reverse inference with Neurosynth also displays large overlap in mid-level regions. Finally, ontology-based decoding maps recover known functional areas the visual and auditory cortices.

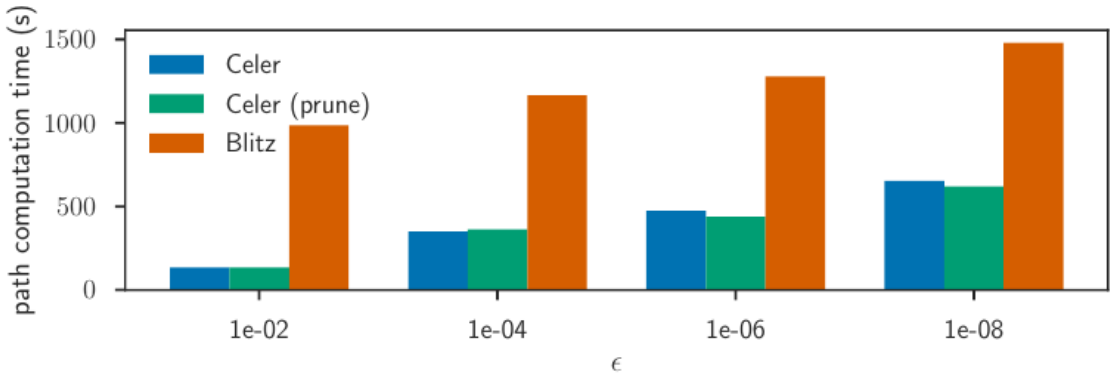


Figure 8. Times to solve the Lasso path to precision ϵ for 100 values of λ, from λ_{max} to $\lambda_{max}/100$, on the Finance data. CELER outperforms BLITZ . Both safe and prune versions behave similarly.

7.7. Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals

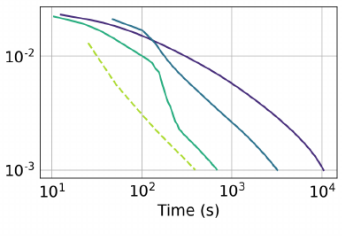
Frequency-specific patterns of neural activity are traditionally interpreted as sustained rhythmic oscillations, and related to cognitive mechanisms such as attention, high level visual processing or motor control. While alpha waves (8–12 Hz) are known to closely resemble short sinusoids, and thus are revealed by Fourier analysis or wavelet transforms, there is an evolving debate that electromagnetic neural signals are composed of more complex waveforms that cannot be analyzed by linear filters and traditional signal representations. In this work, we propose to learn dedicated representations of such recordings using a multivariate convolutional sparse coding (CSC) algorithm. Applied to electroencephalography (EEG) or magnetoencephalography (MEG) data, this method is able to learn not only prototypical temporal waveforms, but also associated spatial patterns so their origin can be localized in the brain. Our algorithm is based on alternated minimization and a greedy coordinate descent solver that leads to state-of-the-art running time on long time series. To demonstrate the implications of this method, we apply it to MEG data and show that it is able to recover biological artifacts. More remarkably, our approach also reveals the presence of non-sinusoidal mu-shaped patterns, along with their topographic maps related to the somatosensory cortex.

More information can be found in [52]. Code can be found at https://alphacsc.github.io/.

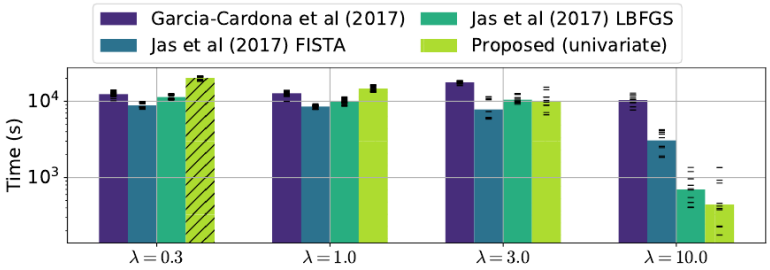
7.8. Stochastic Subsampling for Factorizing Huge Matrices

We present a matrix-factorization algorithm that scales to input matrices with both huge number of rows and columns. Learned factors may be sparse or dense and/or non-negative, which makes our algorithm suitable for dictionary learning, sparse component analysis, and non-negative matrix factorization. Our algorithm streams matrix columns while subsampling them to iteratively learn the matrix factors. At each iteration, the row dimension of a new sample is reduced by subsampling, resulting in lower time complexity compared to a simple streaming algorithm. Our method comes with convergence guarantees to reach a stationary point of the matrix-factorization problem. We demonstrate its efficiency on massive functional Magnetic Resonance Imaging data (2 TB), and on patches extracted from hyperspectral images (103 GB). For both problems, which involve different penalties on rows and columns, we obtain significant speed-ups compared to state-of-the-art algorithms.

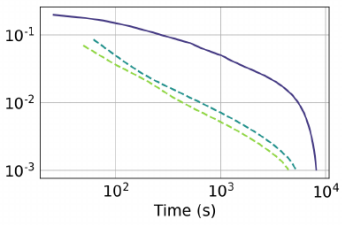
More information can be found in [24].



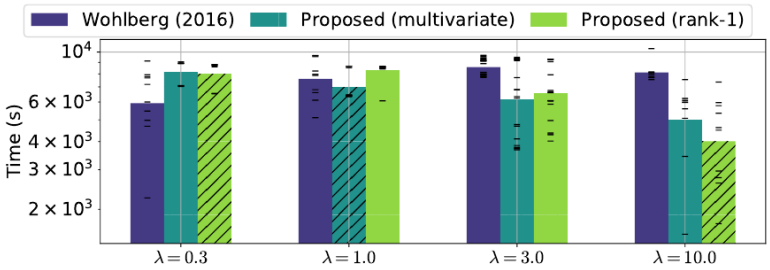
(a) $\lambda = 10$ (univariate).



(b) Time to reach a precision of 0.001 (univariate).



(c) $\lambda = 10$ (multivariate).



(d) Time to reach a precision of 0.001 (multivariate).

Figure 9. Comparison of state-of-the-art univariate (a, b) and multivariate (c, d) methods with our approach. (a) Convergence plot with the objective function relative to the obtained minimum, as a function of computational time. (b) Time taken to reach a relative precision of 10^{-3}, for different regularization parameters λ. (c, d) Same as (a, b) in the multivariate setting P=5.

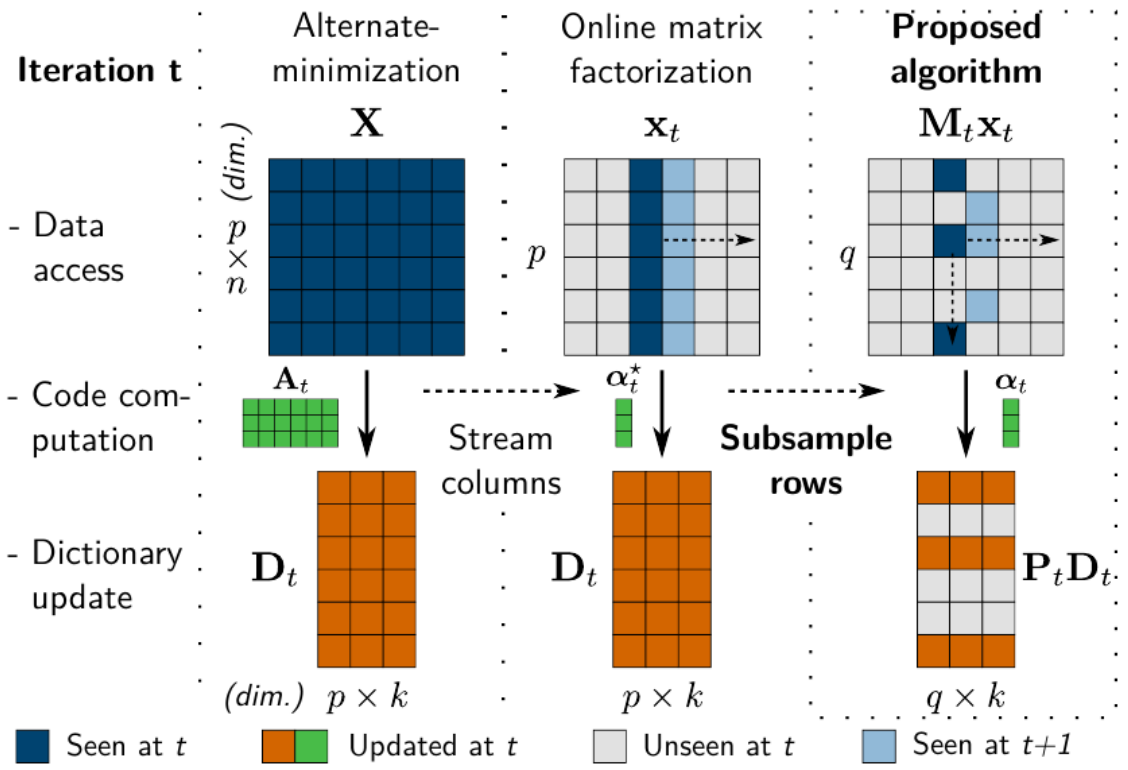


Figure 10. Stochastic subsampling further improves online matrix factorizationhandle datasets with large number of columns and rows. X is the input $p \times n$ matrix, D_t and A_t are respectively the dictionary and code at time t.

7.9. Text to brain: predicting the spatial distribution of neuroimaging observations from text reports

Despite the digital nature of magnetic resonance imaging, the resulting observations are most frequently reported and stored in text documents. There is a trove of information untapped in medical health records, case reports, and medical publications. In this paper, we propose to mine brain medical publications to learn the spatial distribution associated with anatomical terms. The problem is formulated in terms of minimization of a risk on distributions which leads to a least-deviation cost function. An efficient algorithm in the dual then learns the mapping from documents to brain structures. Empirical results using coordinates extracted from the brain-imaging literature show that i) models must adapt to semantic variation in the terms used to describe a given anatomical structure, ii) voxel-wise parameterization leads to higher likelihood of locations reported in unseen documents, iii) least-deviation cost outperforms least-square. As a proof of concept for our method, we use our model of spatial distributions to predict the distribution of specific neurological conditions from text-only reports.

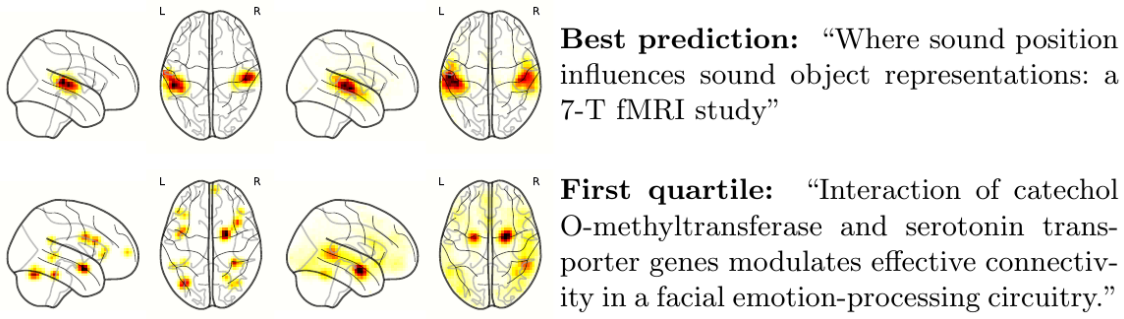


Figure 11. True probability density function (estimated with kernel density estimator) and the prediction for the articles which obtained respectively the best and the first- quartile scores.

More information can be found in [37].

7.10. Similarity encoding for learning with dirty categorical variables

For statistical learning, categorical variables in a table are usually considered as discrete entities and encoded separately to feature vectors, e.g., with one-hot encoding. "Dirty" non-curated data gives rise to categorical variables with a very high cardinality but redundancy: several categories reflect the same entity. In databases, this issue is typically solved with a deduplication step. We show that a simple approach that exposes the redundancy to the learning algorithm brings significant gains. We study a generalization of one-hot encoding, similarity encoding, that builds feature vectors from similarities across categories. We perform a thorough empirical validation on non-curated tables, a problem seldom studied in machine learning. Results on seven real-world datasets show that similarity encoding brings significant gains in prediction in comparison with known encoding methods for categories or strings, notably one-hot encoding and bag of character n-grams. We draw practical recommendations for encoding dirty categories: 3-gram similarity appears to be a good choice to capture morphological resemblance. For very high-cardinality, dimensionality reduction significantly reduces the computational cost with little loss in performance: random projections or choosing a subset of prototype categories still outperforms classic encoding approaches.

More information can be found in [7].

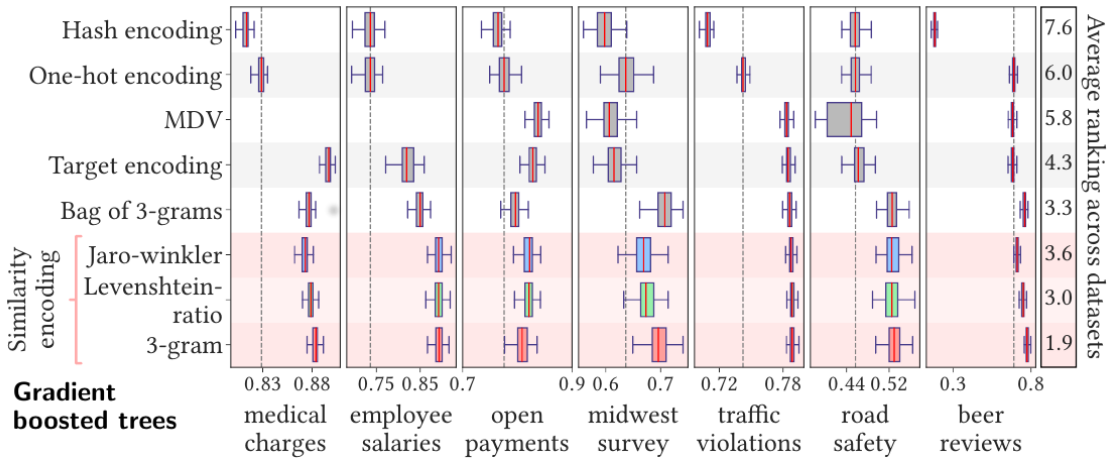


Figure 12. Performance of different encoding methods in a gradient boosting classification task. Each box-plot summarizes the prediction scores of 100 random splits (with 80% of the samples for training and 20% for testing). For all datasets, the prediction score is upper bounded by 1 (a higher score means a better prediction). The right side of the figure indicates the average ranking across datasets for each method. The vertical dashed line indicates the median value of the one-hot encoding method.

VISAGES Project-Team

7. New Results

7.1. Research axis 1: Medical Image Computing in Neuroimaging

Extraction and exploitation of complex imaging biomarkers involve an imaging processing workflow that can be quite complex. This goes from image physics and image acquisition, image processing for quality control and enhancement, image analysis for features extraction and image fusion up to the final application which intends to demonstrate the capability of the image processing workflow to issue sensitive and specific markers of a given pathology. In this context, our objectives in the recent period were directed toward 4 major methodological topics:

7.1.1. Diffusion imaging

7.1.1.1. Optimal selection of diffusion-weighting gradient waveforms using compressed sensing and dictionary learning

Participants: Raphaël Truffet, Emmanuel Caruyer.

Acquisition sequences in diffusion MRI rely on the use of time-dependent magnetic field gradients. Each gradient waveform encodes a diffusion weighted measure; a large number of such measurements are necessary for the in vivo reconstruction of microstructure parameters. We proposed here a method to select only a subset of the measurements while being able to predict the unseen data using compressed sensing. We learnt a dictionary using a training dataset generated with Monte-Carlo simulations; we then compare two different heuristics to select the measures to use for the prediction. We found that an undersampling strategy limiting the redundancy of the measures allows for a more accurate reconstruction when compared with random undersampling with similar sampling rate [57].

7.1.1.2. A Bayes Hilbert Space for Compartment Model Computing in Diffusion MRI

Participant: O. Commowick.

The single diffusion tensor model for mapping the brain white matter microstructure has long been criticized as providing sensitive yet non-specific clinical biomarkers for neurodegenerative diseases because (i) voxels in diffusion images actually contain more than one homogeneous tissue population and (ii) diffusion in a single homogeneous tissue can be non-Gaussian. Analytic models for compartmental diffusion signals have thus naturally emerged but there is surprisingly little for processing such images (estimation, smoothing, registration, atlas-ing, statistical analysis). We propose to embed these signals into a Bayes Hilbert space that we properly define and motivate. This provides a unified framework for compartment diffusion image computing. Experiments show that (i) interpolation in Bayes space features improved robustness to noise compared to the widely used log-Euclidean space for tensors and (ii) it is possible to trace complex key pathways such as the pyramidal tract using basic deterministic tractography thanks to the combined use of Bayes interpolation and multi-compartment diffusion models [26]

This work was done in collaboration with A. Stamm, A. Menafoglio and S.K. Warfield.

7.1.2. Arterial Spin Labeling

7.1.2.1. Patch-Based Super-Resolution of Arterial Spin Labeling Magnetic Resonance Images

Participants: Cédric Meurée, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Arterial spin labeling is a magnetic resonance perfusion imaging technique that, while providing results comparable to methods currently considered as more standard concerning the quantification of the cerebral blood flow, is subject to limitations related to its low signal-to-noise ratio and low resolution. In this work, we investigated the relevance of using a non-local patch-based super-resolution method driven by a high-resolution structural image to increase the level of details in arterial spin labeling images. This method was evaluated by comparison with other image dimension increasing techniques on a simulated dataset, on images of healthy subjects and on images of subjects diagnosed with brain tumors, who had a dynamic susceptibility contrast acquisition. The influence of an increase of ASL images resolution on partial volume effects was also investigated in this work. [56]

The development of this super-resolution algorithm in the context of a thesis financed by Siemens Healthineers conducted to a stay of one month of the PhD candidate in Erlangen, during summer 2018. This immersion into the neuro-development team allowed to integrate the proposed solution with tools in use within this team. Part of the work also consisted in reducing the calculation time, a factor of 5 being achieved at the end of these four weeks.

7.1.2.2. Resting-state ASL : Toward an optimal sequence duration

Participants: Corentin Vallée, Pierre Maurel, Isabelle Corouge, Christian Barillot.

Resting-state functional Arterial Spin Labeling (rs-fASL) in clinical daily practice and academic research stay discreet compared to resting-state BOLD. However, by giving direct access to cerebral blood flow maps, rs-fASL leads to significant clinical subject scaled application as CBF can be considered as a biomarker in common neuropathology. Our work here focused on the link between overall quality of rs-fASL and duration of acquisition. To this end, we consider subject self-Default Mode Network (DMN), and assess DMN quality depletion compared to a gold standard DMN depending on the duration of acquisition [46].

7.1.3. Quantitative imaging

7.1.3.1. Identification of Gadolinium contrast enhanced regions in MS lesions using brain tissue microstructure information obtained from diffusion and T2 relaxometry MRI

Participants: S. Chatterjee, O. Commowick, C. Barillot.

A multiple sclerosis (MS) lesion at an early stage undergoes active blood brain barrier (BBB) breakdown. Identifying MS lesions in a patient which are undergoing active BBB breakdown is of critical importance for MS burden evaluation and treatment planning. However in non-contrast enhanced structural magnetic resonance imaging (MRI) the regions of the lesion undergoing active BBB breakdown cannot be distinguished from the other parts of the lesion. Hence gadolinium (Gd) contrast enhanced T1-weighted MR images are used for this task. However some side effects of Gd injection into patients have been increasingly reported recently. The BBB breakdown is reflected by the condition of tissue microstructure such as increased inflammation, presence of higher extra-cellular matter and debris. We have thus proposed a framework to predict enhancing regions in MS lesions using tissue microstructure information derived from T2 relaxometry and diffusion MRI (dMRI) multicompartment models. We showed that combination of the dMRI and T2 relaxometry microstructure information can distinguish the Gd enhancing lesion regions from the other regions in MS lesions [23].

7.1.3.2. A three year follow-up study of gadolinium enhanced and non-enhanced regions in multiple sclerosis lesions using a multi-compartment T_2 relaxometry model

Participants: S. Chatterjee, O. Commowick, B. Combes, C. Barillot.

Demyelination, axonal damage and inflammation are critical indicators of the onset and progress of neurodegenerative diseases such as multiple sclerosis (MS) in patients. Due to physical limitations of imaging such as acquisition time and imaging resolution, a voxel in a MR image is heterogeneous in terms of tissue microstructure such as myelin, axons, intra and extra cellular fluids and free water. We present a multi-compartment tissue model which estimates the water fraction (WF) of tissues with short, medium and high T_2 relaxation times in a T_2 relaxometry MRI voxel. The proposed method is validated on test-retest data of healthy controls. This model was then used to study longitudinal trends of the tissue microstructures for two sub-regions of the

lesions: gadolinium enhanced (E+) and non-enhanced (L-) regions of MS lesions in 10 MS patients over a period of three years. The water fraction values in E+ and L- regions were found to be significantly different ($p < 0.05$) over the period of first three months. The results of this study also showed that the estimates of the proposed T_2 relaxometry model on brain tissue microstructures have potential to distinguish between regions undergoing active blood brain barrier breakdown from the other regions of the lesion [49].

This work was done in collaboration with Onur Afacan and Simon K. Warfield from Harvard Medical School.

7.1.3.3. Multi-Compartment Model of Brain Tissues from T2 Relaxometry MRI Using Gamma Distribution
Participants: S. Chatterjee, O. Commowick, C. Barillot.

The brain microstructure, especially myelinated axons and free fluids, may provide useful insight into brain neurodegenerative diseases such as multiple sclerosis (MS). These may be distinguished based on their transverse relaxation times which can be measured using T_2 relaxometry MRI. However, due to physical limitations on achievable resolution, each voxel contains a combination of these tissues, rendering the estimation complex. We presented a novel multi-compartment T_2 (MCT2) estimation based on variable projection, applicable to any MCT2 microstructure model. We derived this estimation for a three-gamma distribution model. We validated our framework on synthetic data and illustrated its potential on healthy volunteer and MS patient data [32].

This work was done in collaboration with Onur Afacn and Simon K. Warfield.

7.1.3.4. A 3-year follow-up study of enhancing and non-enhancing multiple sclerosis (MS) lesions in MS patients demonstrating clinically isolated syndrome (CIS) using a multi-compartment T_2 relaxometry (MCT2) model
Participants: S. Chatterjee, O. Commowick, B. Combes, A. Kerbrat, C. Barillot.

Obtaining information on condition of tissue microstructures (such as myelin, intra/extra cellular cells, free water) can provide important insights into MS lesion. However, MRI voxels are heterogeneous in terms of tissue microstructure due to the limited imaging resolution owing to existing physical limitations of MRI scanners. Here we evaluated a multi-compartment T_2 relaxometry model and then used it to study the evolution of enhancing (USPIO and gadolinium positive) and non-enhancing lesions in 6 MS patients with CIS characteristics over a period 3 years with 7 follow-up scans post baseline [31].

This work was done in collaboration with Onur Afacn and Simon K. Warfield.

7.1.4. Atlases

7.1.4.1. Anisotropic similarity, a constrained affine transformation: Application to brain development analysis
Participants: A. Legouhy, O. Commowick, C. Barillot.

The study of brain development provides insights in the normal trend of brain evolution and enables early detection of abnormalities. We proposed a method to quantify brain growth in three arbitrary orthogonal directions of the brain through linear registration. We introduced a 9 degrees of freedom transformation that gives the opportunity to extract scaling factors describing brain growth along those directions by registering a database of subjects in a common basis. We applied this framework to create longitudinal curves of scaling ratios along fixed orthogonal directions from 0 to 16 years highlighting anisotropic brain development [39].

This work was done in collaboration with François Rousseau under the ANR MAIA project.

7.1.5. Simultaneous EEG/fMRI

7.1.5.1. Automated Electrodes Detection during simultaneous EEG/fMRI
Participants: M. Fleury, C. Barillot, E. Bannier, P. Maurel.

The coupling of Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) enables the measure of brain activity at high spatial and temporal resolution. The localisation of EEG sources depends on several parameters including the knowledge of the position of the electrodes on the scalp. An accurate knowledge about this information is important for source reconstruction. Currently, when acquiring EEG and fMRI together, the position of the electrodes is generally estimated according to fiducial points by using a template. In the context of simultaneous EEG/fMRI acquisition, a natural idea is to use magnetic resonance (MR) images to localise EEG electrodes. However, most MR compatible electrodes are built to be almost invisible on MR Images. Taking advantage of a recently proposed Ultra short Echo Time (UTE) sequence, we introduce a fully automatic method to detect and label those electrodes in MR images. Our method was tested on 8 subjects wearing a 64-channel EEG cap. This automated method showed an average detection accuracy of 94% and the average position error was 3.1 mm. These results suggest that the proposed method has potential for determining the position of the electrodes during simultaneous EEG/fMRI acquisition with a very light cost procedure" [11], [35].

This work was done in collaboration with Marsel Mano from Biotrial.

7.1.6. Inference in neuroimaging

7.1.6.1. Validity of summary statistics-based mixed-effects group fMRI

Participant: Camille Maumet.

Statistical analysis of multi-subject functional Magnetic Resonance Imaging (fMRI) data is traditionally done using either: 1) a mixed-effects GLM (MFX GLM) where within-subject variance estimates are used and incorporated into per-subject weights or 2) a random-effects General linear model (GLM) (RFX GLM) where within-subject variance estimates are not used. Both approaches are implemented and available in major neuroimaging software packages including: SPM (MFX analysis; 2nd-Level statistics), FSL (FLAME; OLS) and AFNI (3dMEMA; 3dttest++). While MFX GLM provides the most efficient statistical estimate, its properties are only guaranteed in large samples, and it has been shown that RFX GLM is a valid alternative for one-sample group analyses in fMRI [1]. We recently showed that MFX GLM for image-based meta-analysis could lead to invalid results in small-samples. We investigated whether this issue also affects group fMRI [42].

This work was done in collaboration with Prof. Thomas Nichols from the Oxford Big Data Institute, UK.

7.1.6.2. Choosing a practical and valid Image-Based Meta-Analysis

Participant: Camille Maumet.

Meta-analysis provides a quantitative approach to summarise the rich functional Magnetic Resonance Imaging (fMRI) literature. When image data is available for each study, the optimal approach is to perform an Image-Based Meta-Analysis (IBMA) [1]. A number of IBMA approaches have been proposed including combination of standardised statistics (Z's), just effect estimates (E's) or both effect estimates and their standard errors (SE's). While using both E's & SE's and estimating between-study variance should be optimal, the methods are not guaranteed to work for small number of studies. Also, often only standardised estimates are shared, reducing the possible meta-analytic approaches. Finally, because the BOLD signal is non-quantitative care has to be taken in order to insure that E's are expressed in the same units [2,3], especially when combining data from different software packages. Given the growing interest in data sharing in the neuroimaging community there is a need to identify what is the minimal data to be shared in order to allow for future IBMAs. We investigated the validity of 8 IBMA approaches [41].

This work was done in collaboration with Prof. Thomas Nichols from the Oxford Big Data Institute, UK.

7.1.7. Machine learning

7.1.7.1. Learning sparse predictor from hybrid EEG-fMRI neurofeedback

Participants: C. Cury, C. Barillot, P. Maurel.

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) both allow measurement of brain activity for neuro-feedback (NF), respectively with high temporal resolution for EEG and high spatial resolution for fMRI. Using simultaneously fMRI and EEG for NF training is very promising to devise brain rehabilitation protocols, however performing NF-fMRI is costly, exhausting and time consuming, and cannot be repeated too many times for the same subject. The original contribution of this work concerns the prediction of NF scores from EEG recordings only, using a training phase where both EEG and fMRI NF are available. We have proposed a model able to predict NF scores from coupling EEG-fMRI (NF-EEG-fMRI) of 17 subjects in motor imagery task. The prediction of NF-EEG-fMRI scores was found as satisfactory with a significant improved performance with respect to what EEG can provide alone, when adding to NF-EEG, the prediction of NF-fMRI from EEG signals. The prediction of NF-fMRI significantly adds information and increases the quality of the estimated NF-EEG-fMRI scores.

This work was done in collaboration with Remi Gribonval from the Inria/IRISA PANAMA team.

7.2. Research axis 2: Applications in Neuroradiology and Neurological Disorders

7.2.1. Bimodal EEG-fMRI Neurofeedback for Stroke Rehabilitation: a Case Report

Participants: Giulia Lioi, Mathis Fleury, Simon Butet, Christian Barillot, Isabelle Bonan.

Neurofeedback (NF) consists on training self-regulation of brain activity by providing real-time information about the participant brain function. Few works have shown the potential of NF for stroke rehabilitation however its effectiveness has not been investigated yet. NF approaches are usually based on real-time monitoring of brain activity using a single imaging technique. Recent studies have revealed the potential of combining EEG and fMRI to achieve a more efficient and specific self-regulation. In a case report, we tested the feasibility of applying bimodal EEG-MRI NF on two stroke patients. [54]

This work was done in collaboration with Anatole Lecuyer from the Inria/IRISA HYBRID team.

7.2.2. Refining understanding of working memory buffers through the construct of binding: Evidence from a single case informs theory and clinical practise

Participants: Pierre-Yves Jonin, Quentin Duché.

Binding operations carried out in working memory enable the integration of information from different sources during online performance. While available evidence suggests that working memory may involve distinct binding functions, whether or not they all involve the episodic buffer as a cognitive substrate remains unclear. Similarly, knowledge about the neural underpinnings of working memory buffers is limited, more specifically regarding the involvement of medial temporal lobe structures. In the present study, we report on the case of patient KA, with developmental amnesia and selective damage to the whole hippocampal system. We found that KA was unable to hold shape-colours associations (relational binding) in working memory. In contrast, he could hold integrated coloured shapes (conjunctive binding) in two different tasks. Otherwise, and as expected, KA was impaired on three relational memory tasks thought to depend on the hippocampus that are widely used in the early detection of Alzheimer's disease. Our results emphasized a dissociation between two binding processes within working memory, suggesting that the visuo-spatial sketchpad could support conjunctive binding, and may rely upon a large cortical network including sub-hippocampal structures. By contrast, we found evidence for a selective impairment of relational binding in working memory when the hippocampal system is compromised, suggesting that the long-term memory deficit observed in amnesic patients may be related to impaired short-term relational binding at encoding. Finally, these findings may inform research on the early detection of Alzheimer's disease as the preservation of conjunctive binding in KA is in sharp contrast with the impaired performance demonstrated very early in this disease [15].

This work was done in collaboration with Mario Alfredo Parra and Clara Calia, from Herriot Wyatt University, Edinburgh, UK; with Emmanuel Barbeau and Sophie Muratot from the CNRS 5549 CerCo unit in Toulouse, France; and with Serge Belliard, from CHU de Rennes, Service de Neurologie, Rennes, France.

7.2.3. Superior explicit memory despite severe developmental amnesia: In-depth case study and neural correlates

Participants: Pierre-Yves Jonin, Christian Barillot.

The acquisition of new semantic memories is sometimes preserved in patients with hippocampal amnesia. Robust evidence for this comes from case reports of developmental amnesia suggesting that low-to-normal levels of semantic knowledge can be achieved despite compromised epi-sodic learning. However, it is unclear whether this relative preservation of semantic memory results from normal acquisition and retrieval or from residual episodic memory, combined with effortful repetition. Furthermore, lesion studies have mainly focused on the hippocampus itself, and have seldom reported the state of structures in the extended hippocampal system. Preserved components of this system may therefore mediate residual episodic abilities, contributing to the apparent semantic preservation. We recently reported an in-depth study of Patient KA, a 27-year-old man who had severe hypoxia at birth, in which we carefully explored his residual episodic learning abilities. We used novel speeded recognition paradigms to assess whether KA could explicitly acquire and retrieve new context-free memories. Despite a pattern of very severe amnesia, with a 44-point discrepancy between his intelligence and memory quotients, KA exhibited normal-to-superior levels of knowledge, even under strict time constraints. He also exhibited normal-to-superior recognition memory for new material, again under strict time constraints. Multimodal neuroimaging revealed an unusual pattern of selective atrophy within each component of the extended hippocampal system, contrasting with the preservation of anterior subhippocampal cortices. A cortical thickness analysis yielded a pattern of thinner but also thicker regional cortices, pointing toward specific temporal lobe reorganization following early injury. We thus report the first case of superior explicit learning and memory in a severe case of amnesia, raising important questions about how such knowledge can be acquired [14].

This work was done in collaboration with Emmanuel Barbeau and Gabriel Besson from the CNRS 5549 CerCo unit in Toulouse, France; with Renaud La Joie; with Jérémie Pariente from the Inserm UMR 1214 Tonic unit in Toulouse, France; and with Serge Belliard, from CHU de Rennes, Service de Neurologie, Rennes, France.

7.2.4. Retrieval practice based on recognition memory: testing the retrieval effort hypothesis

Participants: Pierre-Yves Jonin, Christian Barillot.

We tested a core prediction of the retrieval effort hypothesis as an account for the testing effect (TE). Retrieval effort predicts that automatic, effortful retrieval should not lead to TE. Experiment 1 (N=76) showed that despite an encoding duration of three times less, object pictures retention is better after repeated testing under Old/New recognition conditions, than following repeated study. Experiment 2 (N=30) used a speeded and accuracy boosting procedure to rule out the contribution of recollection to retrieval. Retention of object pictures at 25 minutes and 6 months was similar after repeated testing or after repeated studying. These results call for a revision of the retrieval effort hypothesis as a mechanistic account of TE [53].

This work was done in collaboration with Audrey Noël, from Université de Rennes 2, Rennes, France; with Gabriel Besson from the CNRS 5549 CerCo unit in Toulouse, France; with Emmanuel Barbeau and Sophie Muratot from the CNRS 5549 CerCo unit in Toulouse, France; and with Serge Belliard, from CHU de Rennes, Service de Neurologie, Rennes, France.

7.2.5. Voxel-wise Comparison with a-contrario Analysis for Automated Segmentation of Multiple Sclerosis Lesions from Multimodal MRI

Participants: Francesca Galassi, Olivier Commowick, Christian Barillot.

We have introduced a new framework for the automated and un-supervised segmentation of Multiple Sclerosis lesions from multimodal Magnetic Resonance images. It relies on a voxel-wise approach to detect local white matter abnormalities, with an a-contrario analysis, which takes into account local information. First, a voxel-wise comparison of multimodal patient images to a set of controls is performed. Then, region-based probabilities are estimated using an a-contrario approach. Finally, correction for multiple testing is performed. Validation was undertaken on a multi-site clinical dataset of 53 MS patients with various number and volume

of lesions. We showed that the proposed framework outperforms the widely used FDR-correction for this type of analysis, particularly for low lesion loads [37].

This work was done in collaboration with Emmanuel Vallée from Orange Labs, Lannion, France.

7.2.6. Integration of Probabilistic Atlas and Graph Cuts for Automated Segmentation of Multiple Sclerosis lesions

Participants: Francesca Galassi, Olivier Commowick, Christian Barillot.

We have proposed a framework for automated segmentation of Multiple Sclerosis (MS) lesions from MR brain images. It integrates a priori tissues and MS lesions information into a Graph-Cuts algorithm for improved segmentation results. [36].

7.2.7. Multiple sclerosis

7.2.7.1. Spinal Cord

Participants: Anne Kerbrat, Gilles Edan, Jean-Christophe Ferré, Benoit Combès, Olivier Commowick, Élise Bannier, Sudhanya Chatterjee, Haykel Snoussi, Emmanuel Caruyer, Christian Barillot.

The VisAGeS research team has a strong focus on applying the developed methodologies (illustrated in research axis 1) to multiple sclerosis (MS) understanding and the prediction of its evolution. Related to the EMISEP project on spinal cord injury evolution in MS, a first work investigated the magnetization transfer reproducibility across centers in the spinal cord and was accepted for publication [6]. Based on this work, a second work investigated the sensitivity of magnetization transfer to assess diffuse and focal burden in MS patients and was published in Multiple Sclerosis [5].

7.2.7.2. Reproducibility and Evolution of Diffusion MRI measurements within the Cervical Spinal Cord in Multiple Scleros

Participants: Haykel Snoussi, Anne Kerbrat, Benoit Combès, Olivier Commowick, Élise Bannier, Emmanuel Caruyer, Christian Barillot.

In Multiple Sclerosis (MS), there is a large discrepancy between the clinical observations and how the pathology is exhibited on brain images, this is known as the clinical-radiological paradox (CRP). One of the hypotheses is that the clinical deficit may be more related to the spinal cord damage than the number or location of lesions in the brain. Therefore, investigating how the spinal cord is damaged becomes an acute challenge to better understand and overcome the CRP. Diffusion MRI is known to provide quantitative

figures of neuronal degeneration and axonal loss, in the brain as well as in the spinal cord. In this work, we have proposed to investigate how diffusion MRI metrics vary in the different cervical regions with the progression of the disease. We first study the reproducibility of diffusion MRI on healthy volunteers with a test-retest procedure using both standard diffusion tensor imaging (DTI) and multi-compartment Ball-and-Stick models. Then, based on the test re-test quantitative calibration, we provided quantitative figures of pathology evolution between M0 and M12 in the cervical spine on a set of 31 MS patients, exhibiting how the pathology damage spans in the cervical spinal cord.

7.2.8. Epilepsy

Participants: Élise Bannier, Jean-Christophe Ferré.

Accurate localization of the thalamic subregions is of paramount importance for Deep Brain Stimulation (DBS) planning. Current MRI protocols use T2 and Gadolinium-enhanced T1 images, to visualize both the basal ganglia and the vessels, in order to define the electrode trajectory and target. A study showing the usefulness of Fluid and White Matter Suppression, i.e. FLAWS imaging, in eleven drug-resistant epileptic patients for preoperative Deep Brain Stimulation planning and anterior thalamic nucleus targeting was presented as a Power Pitch at the ISMRM Meeeting in Paris [27].

This work was done in collaboration with Giulio Gambarota, Anca Nica and Claire Haegelen from the LTSI and Tobias Kober from Lausanne.

7.2.9. Arterial Spin Labeling in pediatric populations

Participants: Élise Bannier, Christian Barillot, Olivier Commowick, Isabelle Corouge, Jean-Christophe Ferré, Antoine Legouhy, Maia Proisy.

Arterial Spin Labeling is an attractive perfusion MRI technique due to its complete non-invasiveness. However it still remains confidential in clinical practice. Over the years, we have developed several applications to evaluate its potential in different contexts. As part of the PhD of Maia Proisy, we have been working on processing and analysing MR perfusion images using arterial spin labeling in neonates and children for several purposes:

- Investigation of brain perfusion evolution between 6 month and 15 years using ASL sequence in order to provide reference values in this age range [4],
- Evaluation of the evolution of the cerebral blood flow changes betwren day-of-life 3 and day-of-life 10 in a population of neonates with hypoxic-ischemic encephalopathy [45], [*"Changes in brain perfusion in successive arterial spin labelling MRI scans in neonates with hypoxic-ischaemic encephalopathy"*, article in revision in Neuroimage: Clinical].

7.2.10. Diffusion MRI in depression

7.2.10.1. Diffusion MRI as an imaging marker of depression from a large and homogenous population study

Participants: Julie Coloigner, Jean-Marie Batail, Isabelle Corouge, Jean-Christophe Ferre, Christian Barillot.

Despite the extensive therapy options available for depression, up to 80% of patients will suffer from a relapse. Consequently, understanding the neural correlates underlying the depression will optimize the diagnosis and treatment of individual depressed patients. In an experimental study, we investigated alterations of white matter integrity in a large cohort of patients suffering from depression using diffusion tensor imaging. Our findings provide robust evidence that the reduction of white-matter integrity in the interhemispheric connections and fronto-limbic neuronal circuits may play an important role in depression pathogenesis. [34].

This work was done in collaboration with Dominique Drapier from Academic Psychiatry Department, Centre Hospitalier Guillaume Régnier, Rennes, France, EA 4712 Behavior and Basal Ganglia, CHU Rennes, Rennes 1 University, Rennes, France.

7.2.10.2. Diffusion MRI as a descriptive imaging marker of the pathogenesis of treatment-resistant depression

Participants: Julie Coloigner, Jean-Marie Batail, Isabelle Corouge, Jean-Christophe Ferre, Christian Barillot.

Despite the extensive therapy options available for depression, treatment-resistant depression (TRD) occurs in 20-30% of depressed patients. Consequently, identification of neural changes in TRD could support to better understand the mechanism of resistance and to improve the treatment of individual depressed patients. We aimed to investigate the white-matter microstructure in a sample of depressed patients in which response to treatment was subsequently evaluated 6 months after. Our findings suggest the abnormalities of the white-matter integrity in multiple white matter tracts, such as anterior limb of internal capsule and genu of corpus may play a role in the pathogenesis of treatment-resistant depression [33].

Depressive disorder is characterized by a profound dysregulation of affect and mood as well as additional abnormalities including cognitive dysfunction, insomnia, fatigue and appetite disturbance. This disease is the most prevalent mental illness, with an estimated lifetime prevalence reported to range from 10% to 15% worldwide. Despite the extensive therapy options available for depression, up to 80% of patients will suffer from a relapse. Consequently, understanding the neural correlates underlying the depression is critical for improving the specificity and efficacy of diagnostic and treatment strategies.Previous studies of structural and functional magnetic resonance imaging have reported several microstructural abnormalities in the prefrontal cortex, anterior cingulate cortex, hippocampus and thalamus. These observations suggest a dysfunction of the circuits connecting frontal and subcortical brain regions, leading to a "disconnection syndrome". Using graph theory-based analysis, we examined white matter changes in the organization of networks in patients suffering from depression. Our diffusion imaging data showed white matter alteration in patients suffering from depression is occurring in the anterior thalamic radiation and in the cingulate bundle. Our findings

suggest decreased fiber density in circuits connecting subcortical brain regions with the frontal and parietal cortex, supporting the theory of limbic-frontal circuit dysfunction [50]. *We were awarded for this workby the French Institute of Psychiatry for our communication at its annual Forum.*

This work was done in collaboration with Dominique Drapier from Academic Psychiatry Department, Centre Hospitalier Guillaume Régnier, Rennes, France, EA 4712 Behavior and Basal Ganglia, CHU Rennes, Rennes 1 University, Rennes, France.

7.3. Research axis 3: Management of Information in Neuroimaging

7.3.1. Large-scale data analyses

7.3.1.1. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure

Participants: O. Commowick, M. Kain, F. Leray, M. Simon, J.-C. Ferré, A. Kerbrat, G. Edan, C. Barillot.

In collaboration with OFSEP and France Life Imaging, we have proposed a study of multiple sclerosis segmentation algorithms conducted at the international MICCAI 2016 challenge. This challenge was operated using France Life Imaging (FLI-IAM), a new open-science computing infrastructure. This allowed for the automatic and independent evaluation of a large range of algorithms in a fair and completely automatic manner. This computing infrastructure was used to evaluate thirteen methods of MS lesions segmentation, exploring a broad range of state-of-the-art algorithms, against a high-quality database of 53 MS cases coming from four centers following a common definition of the acquisition protocol. Each case was annotated manually by an unprecedented number of seven different experts. Results of the challenge highlighted that automatic algorithms, including the recent machine learning methods (random forests, deep learning,...), are still trailing human expertise on both detection and delineation criteria. In addition, we demonstrated that computing a statistically robust consensus of the algorithms performs closer to human expertise on one score (segmentation) although still trailing on detection scores [7]

This work was done in collaboration with A. Istace, B. Laurent, S. C. Pop, P. Girard, R. Ameli, T. Tourdias, F. Cervenansky, T. Glatard, J. Beaumont, S. Doyle, F. Forbes, J. Knight, A. Khademi, A. Mahbod, C. Wang, R. Mckinley, F. Wagner, J. Muschelli, E. Sweeney, E. Roura, X. Lladó, M. M. Santos, W. P. Santos, A. G. Silva-Filho, X. Tomas-Fernandez, H. Urien, I. Bloch, S. Valverde, M. Cabezas, F. J. Vera-Olmos, N. Malpica, C. R. G. Guttmann, S. Vukusic, M. Dojat, M. Styner, S. K. Warfield and F. Cotton.

7.3.1.2. Same Data - Different Software - Different Results? Analytic Variability of Group fMRI Results

Participant: Camille Maumet.

A wealth of analysis tools are available to fMRI researchers in order to extract patterns of task variation and, ultimately, understand cognitive function. However, this 'methodological plurality' comes with a drawback. While conceptually similar, two different analysis pipelines applied on the same dataset may not produce the same scientific results. Differences in methods, implementations across software packages, and even operating systems or software versions all contribute to this variability. Consequently, attention in the field has recently been directed to reproducibility and data sharing. Neuroimaging is currently experiencing a surge in initiatives to improve research practices and ensure that all conclusions inferred from an fMRI study are replicable. In this work, our goal was to understand how choice of software package impacts on analysis results. We used publically shared data from three published task fMRI neuroimaging studies, reanalyzing each study using the three main neuroimaging software packages, AFNI, FSL and SPM, using parametric and nonparametric inference. We obtained all information on how to process, analyze, and model each dataset from the publications. We made quantitative and qualitative comparisons between our replications to gauge the scale of variability in our results and assess the fundamental differences between each software package. While qualitatively we found broad similarities between packages, we also discovered marked differences, such as Dice similarity coefficients ranging from 0.000-0.743 in comparisons of thresholded statistic maps between software [28], [48].

This work was done in collaboration with Alexander Bowring and Prof. Thomas Nichols from the Oxford Big Data Institute in the UK.

7.3.1.3. Detecting and Interpreting Heterogeneity and Publication Bias in Image-Based Meta-Analyses

Participant: Camille Maumet.

With the increase of data sharing, meta-analyses are becoming increasingly important in the neuroimaging community. They provide a quantitative summary of published results and heightened confidence due to higher statistical power. The gold standard approach to combine results from neuroimaging studies is an Image-Based Meta-Analysis (IBMA) [1] in which group-level maps from different studies are combined. Recently, we have introduced the IBMA toolbox, an extension for SPM that provides methods for combining image maps from multiple studies [2]. However, the current toolbox lacks diagnostic tools used to assess critical assumptions of meta-analysis, in particular whether there is inter-study variation requiring random-effects IBMA, and whether publication bias is present. We have proposed two new tools added to the IBMA toolbox to detect heterogeneity and to assess evidence of publication bias [40].

This work was done in collaboration with Thomas Maullin-Saper and Prof. Thomas Nichols from the Oxford Big Data Institute in the UK.

7.3.2. Infrastructures

7.3.2.1. Open Science for the Neuroinformatics community

Participants: Camille Maumet, Xavier Rolland, Michael Kain, Christian Barillot.

The Neuroinformatics community in OpenAire-Connect is represented by members of the France Life Imaging (FLI) collaboration. In this context, we aim at leveraging OpenAire-Connect services and give our community members the possibility to easily publish and exchange research artefacts from FLI platforms, such as VIP and Shanoir. This will enable open and reproducible science, since literature, data, and methods can be linked, retrieved, and replayed by all the members of the community [30].

This work was done in collaboration with Sorina Pop, Axel Bonnet and Tristan Glatard.

7.3.3. Standardisation and interoperability

7.3.3.1. Interoperability with Boutiques and CARMIN

Participants: Camille Maumet, Michael Kain, Christian Barillot.

A growing number of platforms and tools have lately been developed to meet the needs of various scientific communities. Most of these solutions are optimized to specific requirements from different user groups, leading to technological fragmentation and lack of interoperability. In our quest of open and reproducible science, we proposed two complementary tools, Boutiques and CARMIN, providing cross-platform interoperability for scientific applications, data sharing and processing [29].

This work was done in collaboration with Sorina Pop, Axel Bonnet and Tristan Glatard.

7.3.3.2. A standardised representation for non-parametric fMRI results

Participant: Camille Maumet.

Reuse of data collected and analysed at another site is becoming more prevalent in the neuroimaging community but this process usually relies on intensive data and metadata curation. Given the ever-increasing number of research datasets produced and shared, it is desirable to rely on standards that will enable automatic data and metadata retrieval for large-scale analyses. We recently introduced NIDM-Results, a data model to represent and publish data and metadata created as part of a mass univariate neuroimaging study (typically functional magnetic resonance imaging). In this work, we have proposed to extend this model to allow for the representation of non-parametric analyses and we introduce a JSON API that will facilitate export into NIDM-Results [25].

This work was done as part of an international collaboration with Guillaume Flandin, Martin Perez-Guevara, Jean-Baptiste Poline, Justin Rajendra, Richard Reynolds, Betrand Thirion, Thomas Maullin-Sapey and Thomas Nichols.

7.3.3.3. Development of an Ontology for the INCF Neuroimaging Data Model (NIDM)

Participant: C. Maumet.

The successful reuse of shared data relies on the existence of easily-available well-described metadata. The metadata, as a rich description of the data, must capture information on how the data was acquired, processed and analyzed. The terms used to describe the data should be chosen with a logical, consistent framework in mind and include definitions to avoid ambiguity. In addition, a lexicon or ontology should reuse terms from existing efforts as much as possible [38].

This work was done as part of an international collaboration with K.G. Helmer, K.B. Keator, T. Auer, S. Ghosh, T.E. Nichols, P. Smruti and J.B. Poline.

AIRSEA Project-Team

6. New Results

6.1. Modeling for Oceanic and Atmospheric flows

6.1.1. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Florian Lemarié, Sophie Thery.

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approximation of one [72]. For the last few years we have been actively working on the analysis of ocean-atmosphere coupling both in terms of its continuous and numerical formulation.Our activities over the last few years can be divided into four general topics

1. *Stability and consistency analysis of existing coupling methods*: in [72] we showed that the usual methods used in the context of ocean-atmosphere coupling are prone to splitting errors because they correspond to only one iteration of an iterative process without reaching convergence. Moreover, those methods have an additional condition for the coupling to be stable even if unconditionally stable time stepping algorithms are used. This last remark was further studied recently in [47] 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 and defended on Feb. 15, 2018, [2]) the scope was on including the formulation of physical parameterizations in the theoretical analysis of the coupling, in particular the parameterization schemes to compute air-sea fluxes [79]. 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. This metamodel is more adequate to conduct the mathematical analysis of the coupling while being physically satisfactory [80]. In parallel we have contributed to a general review gathering the main international specialists on the topic [64]. More recently we have started to work specifically on the discretization methods for the parameterization of planetary boundary layers in climate models [27] which takes the form of an nonstationary nonlinear parabolic equation. The objective is to derive a discretization for which we could prove nonlinear stability criteria and show robustness to large variations in parabolic Courant number while being consistent with our knowledge of the underlying physical principles (e.g. the Monin-Obukhov theory in the surface layer).
3. *Design of a coupled single column model*: in order to focus on specific problems of ocean-atmosphere coupling, a work on simplified equation sets has been started. The aim is to implement a one-dimensional (in the vertical direction) coupled model with physical parameterizations representative of those used in realistic models. Thanks to this simplified coupled model the objective is to develop a benchmark suite for coupled models evaluation. Last year the single column oceanic and atmospheric components have been developed and coupled during the PhD-thesis of Rémi Pellerej (defended on Mar. 26, 2018) and in the framework of the SIMBAD project [17]. A publication describing this model is currently in preparation for the Geoscientific Model Development journal.
4. *Analysis of air-sea-wave 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. [73]) in large-scale realistic ocean-atmosphere coupled simulations [14]. Moreover, within the ALBATROS project, we have contributed to the development of a 2-way coupling between an ocean global circulation model (NEMO) with a surface wave model (WW3). Such coupling is not straightforward to implement since it requires modifications of the governing equations, boundary conditions and subgrid scale closures in the oceanic model. A paper is currently under review in Geoscientific Model Development journal on that topic.

5. *Efficient coupling methods*: we have been developing coupling approaches for several years, based on so-called Schwarz algorithms. In particular, we addressed the development of efficient interface conditions for multi-physics problems representative of air-sea coupling [28] (paper in preparation). This work is done in the framework of S. Théry PhD (started in fall 2017).

These topics are addressed through strong collaborations between the applied mathematicians and the climate community (Meteo-France, Ifremer, LMD, and LOCEAN). Indeed, Our work on ocean-atmosphere coupling has steadily matured over the last few years and has reached a point where it triggered interest from the climate community. Through the funding of the COCOA ANR project (started in January 2017, PI: E. Blayo), Airsea team members play a major role in the structuration of a multi-disciplinary scientific community working on ocean-atmosphere coupling spanning a broad range from mathematical theory to practical implementations in climate models. An expected outcome of this project should be the design of a benchmark suite of idealized coupled test cases representative of known issues in coupled models. Such idealized test cases should motivate further collaborations at an international level.

6.1.2. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Matthieu Brachet, Laurent Debreu, Emilie Duval, Nicholas Kevlahan, Florian Lemarié, Christopher Eldred, Farshid Nazari.

The increase of model resolution naturally leads to the representation of a wider energy spectrum. As a result, in recent years, the understanding of oceanic submesoscale dynamics has significantly improved. However, dissipation in submesoscale models remains dominated by numerical constraints rather than physical ones. Effective resolution is limited by the numerical dissipation range, which is a function of the model numerical filters (assuming that dispersive numerical modes are efficiently removed). A review paper on coastal ocean models has been written with German colleagues and has been published in Ocean Modelling ([11]).
F. Lemarié and L. Debreu (with H. Burchard, K. Klingbeil and J. Sainte-Marie) have organized the international COMMODORE workshop on numerical methods for oceanic models (Paris, Sept. 17-19, 2018). https://commodore2018.sciencesconf.org/, see [12] for a summary of the scientific discussions

With the increase of resolution, the hydrostatic assumption becomes less valid and the AIRSEA group also works on the development of non-hydrostatic ocean models. The treatment of non-hydrostatic incompressible flows leads to a 3D elliptic system for pressure that can be ill conditioned in particular with non geopotential vertical coordinates. That is why we flavor the use of the non-hydrostatic compressible equations that removes the need for a 3D resolution at the price of reincluding acoustic waves [24].
In addition, Emilie Duval started her PhD in September 2018 on the coupling between the hydrostatic incompressible and non-hydrostatic compressible equations.
The team is involved in the HEAT (Highly Efficient ATmospheric Modelling) ANR project. This project aims at developing a new atmospheric dynamical core (DYNAMICO) discretized on an icosahedral grid. This project is in collaboration with Ecole Polytechnique, Meteo-France, LMD, LSCE and CERFACS. This year we worked on dispersion analysis of compatible Galerkin schemes for shallow water model both in 1D ([5]) and 2D ([39]). In addition, we worked on the discrete formulation of the thermal rotating shallow water equations. This formulation, based on quasi-Hamiltonian discretizations methods, allows for the first time total mass, buoyancy and energy conservation to machine precision ([4]).

6.1.3. Data assimilation for coupled models

In the context of operational meteorology and oceanography, forecast skills heavily rely on proper combination of model prediction and available observations via data assimilation techniques. Historically, numerical weather prediction is made separately for the ocean and the atmosphere in an uncoupled way. However, in recent years, fully coupled ocean-atmosphere models are increasingly used in operational centers to improve the reliability of seasonal forecasts and tropical cyclones predictions. For coupled problems, the use of separated data assimilation schemes in each medium is not satisfactory since the result of such assimilation process is generally inconsistent across the interface, thus leading to unacceptable artefacts. Hence, there is a strong need for adapting existing data assimilation techniques to the coupled framework. As part of our ERACLIM2 contribution, R. Pellerej started a PhD on that topic late 2014 and defended it early 2018

[1]. Three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a single column coupled model. The dynamical equations of the considered problem are coupled using an iterative Schwarz domain decomposition method. The aim is to properly take into account the coupling in the assimilation process in order to obtain a coupled solution close to the observations while satisfying the physical conditions across the air-sea interface. Results shows significant improvement compared to the usual approach on this simple system. The aforementioned system has been coded within the OOPS framework (Object Oriented Prediction System) in order to ease the transfer to more complex/realistic models.

Finally, CASIS, a new collaborative project with Mercator Océan has started late 2017 in order to extend developments to iterative Kalman smoother data assimilation scheme, in the framework of a coupled ocean-atmospheric boundary layer context.

6.1.4. Optimal control of grids and schemes for ocean model.

Participants: Laurent Debreu, Eugene Kazantsev.

In [33], variational data assimilation technique is applied to a simple bidimensional wave equation that simulates propagation of internal gravity waves in the ocean in order to control grids and numerical schemes. Grid steps of the vertical grid, Brunt-Vaisala frequency and approximation of the horizontal derivative were used as control parameters either separately or in the joint control. Obtained results show that optimized parameters may partially compensate errors committed by numerical scheme due to insufficient grid resolution.

Optimal vertical grid steps and coefficients in horizontal derivative approximation found in the variational control procedure allow us to get the model solution that is rather close to the solution of the reference model. The error in the wave velocity on the coarse grid is mostly compensated in experiments with joint control of parameters while the error in the wave amplitude occurs to be more difficult to correct.

However, optimal grid steps and discretization schemes may be in a disagreement with requirements of other model physics and additional analysis of obtained optimized parameters from the point of view of they agreement with the model is necessary.

6.1.5. Nonhydrostatic Modeling

Participants: Eric Blayo, Laurent Debreu, Emilie Duval.

In the context of the French initiative CROCO (Coastal and Regional Ocean COmmunity model, https://www.croco-ocean.org) for the development of a new oceanic modeling system, Emilie Duval started a PhD (Oct. 2018) focused on the design of methods to couple local nonhydrostatic models to larger scale hydrostatic ones. Such a coupling is quite delicate from a mathematical point of view, due to the different nature of hydrostatic and nonhydrostatic equations (where the vertical velocity is either a diagnostic or a prognostic variable).

6.2. Model reduction / multiscale algorithms

6.2.1. Model Order Reduction

Participants: Mohamed Reda El Amri, Youssef Marzouk, Maëlle Nodet, Clémentine Prieur, Alessio Spantini, Olivier Zahm.

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 [68], 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 [69]. In [66], 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 [70].

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 [67], 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 [70]. In [71], we propose a generalization of the probabilistic goal-oriented error estimation in [67] to parameter-dependent nonlinear problems. One aims at applying such results in the previous context of sensitivity of a controlled system.

More recently, in the context of the Inria associate team UNQUESTIONABLE, we have extended the focus of the axis on model order reduction. Our objectives are to understand the kinds of low-dimensional structure that may be present in important geophysical models; and to exploit this low-dimensional structure in order to extend Bayesian approaches to high-dimensional inverse problems, such as those encountered in geophysical applications. Our recent and future efforts are/will be concerned with parameter space dimension reduction techniques, low- rank structures in geophysical models and transport maps tools for probability measure approximation. At the moment, scientific progress has been achieved in different directions, as detailed below: A first paper [45] has been submitted on gradient- based dimension reduction of vector-valued functions. Multivariate functions encountered in high-dimensional uncertainty quantification problems often vary most strongly along a few dominant directions in the input parameter space. In this work, we propose a gradient-based method for detecting these directions and using them to construct ridge approximations of such functions, in the case where the functions are vector-valued. The methodology consists of minimizing an upper bound on the approximation error, obtained by subspace Poincaré inequalities. We have provided a thorough mathematical analysis in the case where the parameter space is equipped with a Gaussian probability measure. A second work [46] has been submitted, which proposes a dimension reduction technique for Bayesian inverse problems with nonlinear forward operators, non-Gaussian priors, and non-Gaussian observation noise. In this work, the likelihood function is approximated by a ridge function, i.e., a map which depends non-trivially only on a few linear combinations of the parameters. The ridge approximation is built by minimizing an upper bound on the Kullback-Leibler divergence between the posterior distribution and its approximation. This bound, obtained via logarithmic Sobolev inequalities, allows one to certify the error of the posterior approximation. A sample-based approximation of the upper bound is also proposed. In the framework of the PhD thesis of Reda El Amri, a work on data-driven stochastic inversion via functional quantization was submitted. In this paper [36], a new methodology is proposed for solving stochastic inversion problems through computer experiments, the stochasticity being driven by functional random variables. Main tools are a new greedy algorithm for functional quantization, and the adaptation of Stepwise Uncertainty Reduction techniques.

6.3. Dealing with uncertainties

6.3.1. Sensitivity Analysis

Participants: Elise Arnaud, Eric Blayo, Laurent Gilquin, Maria Belén Heredia, François-Xavier Le Dimet, Clémentine Prieur, Laurence Viry.

6.3.1.1. 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.

6.3.2. Extensions of the replication method for the estimation of Sobol' indices

Participants: Elise Arnaud, Eric Blayo, Laurent Gilquin, Alexandre Janon, Clémentine Prieur.

Sensitivity analysis studies how the uncertainty on an output of a mathematical model can be attributed to sources of uncertainty among the inputs. Global sensitivity analysis of complex and expensive mathematical models is a common practice to identify influent inputs and detect the potential interactions between them. Among the large number of available approaches, the variance-based method introduced by Sobol' allows to calculate sensitivity indices called Sobol' indices. Each index gives an estimation of the influence of an individual input or a group of inputs. These indices give an estimation of how the output uncertainty can be apportioned to the uncertainty in the inputs. One can distinguish first-order indices that estimate the main effect from each input or group of inputs from higher-order indices that estimate the corresponding order of interactions between inputs. This estimation procedure requires a significant number of model runs, number that has a polynomial growth rate with respect to the input space dimension. This cost can be prohibitive for time consuming models and only a few number of runs is not enough to retrieve accurate informations about the model inputs.

The use of replicated designs to estimate first-order Sobol' indices has the major advantage of reducing drastically the estimation cost as the number of runs n becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. However, if the input space is not properly explored, that is if n is too small, the Sobol' indices estimates may not be accurate enough. Gaining in efficiency and assessing the estimate precision still remains an issue, all the more important when one is dealing with limited computational budget.

We designed an approach to render the replication method iterative, enabling the required number of evaluations to be controlled. With this approach, more accurate Sobol' estimates are obtained while recycling previous sets of model evaluations. Its main characteristic is to rely on iterative construction of stratified designs, latin hypercubes and orthogonal arrays [61]

In [7] a new strategy to estimate the full set of first-order and second-order Sobol' indices with only two replicated designs based on orthogonal arrays of strength two. Such a procedure increases the precision of the estimation for a given computation budget. A bootstrap procedure for producing confidence intervals, that are compared to asymptotic ones in the case of first-order indices, is also proposed.

The replicated designs strategy for global sensitivity analysis was also implemented in the applied framework of marine biogeochemical modeling, making use of distributed computing environments [43].

6.3.3. 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 [65] in the case where input parameters are correlated. Clémentine

Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [53]. We obtained first results [54], 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 [52], [55]. We also considered the estimation of groups Sobol' indices, with a procedure based on replicated designs [63]. 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 [75]. The whole methodology proposed during the PhD is presented in [76].

More recently, the Shapley value, from econometrics, was proposed as an alternative to quantify the importance of random input variables to a function. Owen [77] derived Shapley value importance for independent inputs and showed that it is bracketed between two different Sobol' indices. Song et al. [82] recently advocated the use of Shapley value for the case of dependent inputs. In a very recent work [78], in collaboration with Art Owen (Standford's University), we show that Shapley value removes the conceptual problems of functional ANOVA for dependent inputs. We do this with some simple examples where Shapley value leads to intuitively reasonable nearly closed form values. We also investigated further the properties of Shapley effects in [41].

6.3.4. Global sensitivity analysis for parametrized stochastic differential equations

Participant: Clémentine Prieur.

Many models are stochastic in nature, and some of them may be driven by parametrized stochastic differential equations. It is important for applications to propose a strategy to perform global sensitivity analysis for such models, in presence of uncertainties on the parameters. In collaboration with Pierre Etoré (DATA department in Grenoble), Clémentine Prieur proposed an approach based on Feynman-Kac formulas [40].

6.3.5. Parameter control in presence of uncertainties: robust estimation of bottom friction

Participants: Victor Trappler, Elise Arnaud, Laurent Debreu, Arthur Vidard.

Many physical phenomena are modelled numerically in order to better understand and/or to predict their behaviour. However, some complex and small scale phenomena can not be fully represented in the models. The introduction of ad-hoc correcting terms, can represent these unresolved processes, but they need to be properly estimated.

A good example of this type of problem is the estimation of bottom friction parameters of the ocean floor. This is important because it affects the general circulation. This is particularly the case in coastal areas, especially for its influence on wave breaking. Because of its strong spatial disparity, it is impossible to estimate the bottom friction by direct observation, so it requires to do so indirectly by observing its effects on surface movement. This task is further complicated by the presence of uncertainty in certain other characteristics linking the bottom and the surface (eg boundary conditions). The techniques currently used to adjust these settings are very basic and do not take into account these uncertainties, thereby increasing the error in this estimate.

Classical methods of parameter estimation usually imply the minimisation of an objective function, that measures the error between some observations and the results obtained by a numerical model. In the presence of uncertainties, the minimisation is not straightforward, as the output of the model depends on those uncontrolled inputs and on the control parameter as well. That is why we will aim at minimising the objective function, to get an estimation of the control parameter that is robust to the uncertainties.

The definition of robustness differs depending of the context in which it is used. In this work, two different notions of robustness are considered: robustness by minimising the mean and variance, and robustness based on the distribution of the minimisers of the function. This information on the location of the minimisers is not a novel idea, as it had been applied as a criterion in sequential Bayesian optimisation. However, the constraint of

optimality is here relaxed to define a new estimate. To evaluate this estimation, a toy model of a coastal area has been implemented. The control parameter is the bottom friction, upon which classical methods of estimation are applied in a simulation-estimation experiment. The model is then modified to include uncertainties on the boundary conditions in order to apply robust control methods.

6.3.6. Development of a data assimilation method for the calibration and continuous update of wind turbines digital twins

Participants: Adrien Hirvoas, Elise Arnaud, Clémentine Prieur, Arthur Vidard.

In the context of the energy transition, wind power generation is developing rapidly in France and worldwide. Research and innovation on wind resource characterisation, turbin control, coupled mechanical modelling of wind systems or technological development of offshore wind turbines floaters are current research topics.

In particular, the monitoring and the maintenance of wind turbine is becoming a major issue. Current solutions do not take full advantage of the large amount of data provided by sensors placed on modern wind turbines in production. These data could be advantageously used in order to refine the predictions of production, the life of the structure, the control strategies and the planning of maintenance. In this context, it is interesting to optimally combine production data and numerical models in order to obtain highly reliable models of wind turbines. This process is of interest to many industrial and academic groups and is known in many fields of the industry, including the wind industry, as "digital twin".

The objective of Adrien Hirvoas's PhD work is to develop of data assimilation methodology to build the "digital twin" of an onshore wind turbine. Based on measurements, the data assimilation should allow to reduce the uncertainties of the physical parameters of the numerical model developed during the design phase to obtain a highly reliable model. Various ensemble data assimilation approches are currently under consideration to address the problem.

This work in done in collaboration with IFPEN.

6.3.7. Non-Parametric Estimation for Kinetic Diffusions

Participants: Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Chile and Uruguay. More precisely, we started working with Venezuela. Due to the crisis in Venezuela, our main collaborator on that topic moved to Uruguay.

We are focusing our attention on models derived from the linear Fokker-Planck equation. From a probabilistic viewpoint, these models have received particular attention in recent years, since they are a basic example for hypercoercivity. In fact, even though completely degenerated, these models are hypoelliptic and still verify some properties of coercivity, in a broad sense of the word. Such models often appear in the fields of mechanics, finance and even biology. For such models we believe it appropriate to build statistical non-parametric estimation tools. Initial results have been obtained for the estimation of invariant density, in conditions guaranteeing its existence and unicity [48] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [49] (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 [49], in collaboration with J.R. Leon (Montevideo, Uruguay) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [50], also recently accepted. In a recent collaboration with Adeline Samson from the statistics department in the Lab, we considered adaptive estimation, that is we proposed a data-driven procedure for the choice of the bandwidth parameters.

In [51], we focused on damping Hamiltonian systems under the so-called fluctuation-dissipation condition. Idea in that paper were re-used with applications to neuroscience in [74].

Note that Professor Jose R. Leon (Caracas, Venezuela, Montevideo, Uruguay) was funded by an international Inria Chair, allowing to collaborate further on parameter estimation.

We recently proposed a paper on the use of the Euler scheme for inference purposes, considering reflected diffusions.This paper could be extended to the hypoelliptic framework.

We also have a collaboration with Karine Bertin (Valparaiso, Chile), Nicolas Klutchnikoff (Université Rennes) and Jose R. León (Montevideo, Uruguay) funded by a MATHAMSUD project (2016-2017) and by the LIA/CNRS (2018). We are interested in new adaptive estimators for invariant densities on bounded domains [32], and would like to extend that results to hypo-elliptic diffusions.

6.3.8. 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 [56] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [81] 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 [57] 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 [58], [60], 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 [35]. This paper has now been accepted for publication.

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [59]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervised the PhD thesis of Patricia Tencaliec. We are working on risk assessment, concerning flood data for the Durance drainage basin (France). The PhD thesis started in October 2013 and was defended in February 2017. A first paper on data reconstruction has been accepted [83]. It was a necessary step as the initial series contained many missing data. A second paper is in revision, considering the modeling of precipitation amount with semi-parametric models, modeling both the bulk of the distribution and the tails, but avoiding the arbitrary choice of a threshold. We work in collaboration with Philippe Naveau (LSCE, Paris).

6.4. Assimilation of Images

Participants: Elise Arnaud, François-Xavier Le Dimet, Maëlle Nodet, Arthur Vidard, Long Li.

6.4.1. Direct assimilation of image sequences

At the present time the observation of Earth from space is done by more than thirty satellites. These platforms provide two kinds of observational information:

- Eulerian information as radiance measurements: the radiative properties of the earth and its fluid envelops. These data can be plugged into numerical models by solving some inverse problems.
- Lagrangian information: the movement of fronts and vortices give information on the dynamics of the fluid. Presently this information is scarcely used in meteorology by following small cumulus clouds and using them as Lagrangian tracers, but the selection of these clouds must be done by hand and the altitude of the selected clouds must be known. This is done by using the temperature of the top of the cloud.

Our current developments are targeted at the use of « Level Sets » methods to describe the evolution of the images. The advantage of this approach is that it permits, thanks to the level sets function, to consider the images as a state variable of the problem. We have derived an Optimality System including the level sets of the images. This approach is being applied to the tracking of oceanic oil spills in the framework of a Long Li's Phd in co-supervision with

A collaborative project started with C. Lauvernet (IRSTEA) in order to make use of our image assimilation strategies on the control of pesticide transfer.

6.4.2. Optimal transport for image assimilation

We investigate the use of optimal transport based distances for data assimilation, and in particular for assimilating dense data such as images. The PhD thesis of N. Feyeux studied the impact of using the Wasserstein distance in place of the classical Euclidean distance (pixel to pixel comparison). In a simplified one dimensional framework, we showed that the Wasserstein distance is indeed promising. Data assimilation experiments with the Shallow Water model have been performed and confirm the interest of the Wasserstein distance. Results have been presented at conferences and seminars and a paper has been published at NPG [6].

6.5. Land Use and Transport Models Calibration

Participants: Thomas Capelle, Laurent Gilquin, Clémentine Prieur, Arthur Vidard, Peter Sturm, Elise Arnaud.

Given the complexity of modern urban areas, designing sustainable policies calls for more than sheer expert knowledge. This is especially true of transport or land use policies, because of the strong interplay between the land use and the transportation systems. Land use and transport integrated (LUTI) modelling offers invaluable analysis tools for planners working on transportation and urban projects. Yet, very few local authorities in charge of planning make use of these strategic models. The explanation lies first in the difficulty to calibrate these models, second in the lack of confidence in their results, which itself stems from the absence of any well-defined validation procedure. Our expertise in such matters will probably be valuable for improving the reliability of these models. To that purpose we participated to the building up of the ANR project CITiES led by the STEEP EPI. This project started early 2013 and two PhD about sensitivity analysis and calibration were launched late 2013. Laurent Gilquin defended his PhD in October 2016 [62] and Thomas Capelle defended his in April 2017 and published his latest results in [3].

ANGE Project-Team

7. New Results

7.1. Numerical methods for fluid flows

Participants: Jacques Sainte-Marie, Virgile Dubos, Cindy Guichard, Martin Parisot, Marie-Odile Bristeau, Fabien Souillé, Edwige Godlewski, Yohan Penel.

7.1.1. Advancing dynamical cores of oceanic models across all scales

Oceanic numerical models are used to understand and predict a wide range of processes from global paleoclimate scales to short-term prediction in estuaries and shallow coastal areas. One of the overarching challenges, and the main topic of the COMMODORE workshop, is the appropriate design of the dynamical cores given the wide variety of scales of interest and their interactions with atmosphere, sea-ice, biogeochemistry, and even societal processes. The construction of a dynamical core is a very long effort which takes years and decades of research and development and which requires a collaborative mixture of scientific disciplines. In [14], we present a significant number of fundamental choices, such as which equations to solve, which horizontal and vertical grid arrangement is adequate, which discrete algorithms allows jointly computational efficiency and sufficient accuracy, etc.

7.1.2. A Well-balanced Finite Volume Scheme for Shallow Water Equations with Porosity

Our work [20] aims to study the ability of a single porosity-based shallow water model to modelize the impact of vegetation in open-channel flows. More attention on flux and source terms discretizations are required in order to archive the well-balancing and shock capturing. We present a new Godunov-type finite volume scheme based on a simple-wave approximation and compare it with some other methods in the literature. A first application with experimental data was performed.

7.1.3. The gradient discretisation method

The monograph [21] is dedicated to the presentation of the gradient discretisation method (GDM) and to some of its applications. It is intended for masters students, researchers and experts in the field of the numerical analysis of partial differential equations.The GDM is a framework which contains classical and recent discretisation schemes for diffusion problems of different kinds: linear or non-linear, steady-state or time-dependent.

7.1.4. Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow

This work [29] is devoted to the numerical resolution in multidimensional framework of a hierarchy of reduced models of the free surface Euler equations. In a first part, entropy-satisfying scheme is proposed fo the monolayer dispersive model [Green, Naghdi '76] and [Bristeau, Mangeney, Sainte-Marie, Seguin '15]. In a second part, the strategy is extended to the layerwise models proposed in [Fernandez-Nieto, Parisot, Penel, Sainte-Marie]. To illustrate the accuracy and the robustness of the strategy, several numerical experiments are performed. In particular, the strategy is able to deal with dry areas without particular treatment.

7.1.5. Numerical approximation of the 3d hydrostatic Navier-Stokes system with free surface

In this work [23], we propose a stable and robust strategy to approximate the 3d incompressible hydrostatic Euler and Navier-Stokes systems with free surface. Compared to shallow water approximation of the Navier-Stokes system, the idea is to use a Galerkin type approximation of the velocity field with piecewise constant basis functions in order to obtain an accurate description of the vertical profile of the horizontal velocity.

7.1.6. Congested shallow water model: floating object

In [27], we are interested in the floating body problem on a large space scale. We focus on objects floating freely in the water such as icebergs or wave energy converters. The formulation of the fluid-solid interaction using the congested shallow water model for the fluid and Newton's second law of motion for the solid is given and a strong coupling between the two systems is explained. The energy transfer between the solid and the water is focused on since it is of major interest for energy production. A numerical resolution based on the coupling of a finite volume scheme for the fluid and a Newmark scheme for the solid is presented. An entropy correction based on an adapted choice of discretization for the coupling terms is made in order to ensure a dissipation law at the discrete level. Simulations are presented to validate the method and to show the feasibility of more complex cases.

7.1.7. Numerical strategies for a dispersive layer-averaged model

A hierarchy of models has been derived in [12] to approximate the Euler equations by means of a layer-averaging procedure. This results in several dispersive models with one velocity field per layer. The structure of the equations induces issues of efficiency. The standard splitting between hydrostatic and non-hydrostatic components leads to a prohibitive computational costs. In a work in progress, we are investigating a new strategy to solve the projection step in a cheaper way. This is assessed by means of steady nontrivial solutions of the dispersive equations.

7.1.8. Methods of Reflections

The basic idea of the method of reflections appeared almost two hundred years ago; it is a method of successive approximations for the interaction of particles within a fluid, and it seems intuitively related to the Schwarz domain decomposition methods, the subdomains being the complements of the particle domains. We show in [25] that indeed there is a direct correspondence between the methods of reflections and Schwarz methods in the two particle/subdomain case. This allows us to give a new convergence analysis based on maximum principle techniques with precise convergence estimates that one could not obtain otherwise. We then show however also that in the case of more than two particles/subdomains, the methods of reflections and the Schwarz methods are really different methods, with different convergence properties. We finally also introduce for the first time coarse corrections for the methods of reflections to make them scalable in the case when the number of particles becomes large.

7.2. Modelling

Participants: Marie-Odile Bristeau, Jacques Sainte-Marie, Fabien Souillé, Emmanuel Audusse, Léa Boitin, Martin Parisot, Di Martino Bernard, Anne Mangeney.

7.2.1. How do microalgae perceive light in a high-rate pond? Towards more realistic Lagrangian experiments

In [10], we present a multidisciplinary downscaling study, where we first reconstructed single cell trajectories in an open raceway using an original hydrodynamical model offering a powerful discretization of the Navier–Stokes equations tailored to systems with free surfaces. The trajectory of a particular cell was selected and the associated high-frequency light pattern was computed. This light pattern was then experimentally reproduced in an Arduino-driven computer controlled cultivation system with a low density Dunaliella salina culture. The effect on growth and pigment content was recorded for various frequencies of the light pattern, by setting different paddle wheel velocities.

7.2.2. Modeling and simulation of sediment transport

A previous derivation of the sediment layer model has then been extended. Depending on the scaling chosen for the physical parameters, different models are obtained. The model we are interested in is the non-local model (with a viscosity term). Several numerical schemes are implemented and studied to simulate this model. Only one of these schemes is satisfactory. Simulations of the coupled water-sediment systems are made. The influence of the viscosity is emphasized. Turning on the non-local term allows to simulate dune growth and propagation.

Following the previous work, a numerical scheme for the sediment layer is proposed. The numerical scheme is tested. The influence of the viscosity on the behaviour of the sediment layer is studied. A numerical strategy for the resolution of the coupled model (water layer and sediment layer) is implemented. The behaviour of the coupled system is numerically assessed. Academic test cases are performed.

7.2.3. The Navier-Stokes system with temperature and salinity for free-surface flows

We model free surface flows where density variations coming e.g. from temperature or salinity differences play a significant role. Starting from the compressible Navier-Stokes system, a model is derived by performing the incompressible limit (the dependence of the density on the pressure is removed). A layer-averaged formulation of the model is proposed. The layer-averaged model satisfies a dissipative energy balance. A numerical scheme is proposed. It verifies several stability properties (positivity, well-balancing, maximum principle on the density). Numerical simulations are performed. The differences with models relying on the classical Boussinesq approximation are shown.

7.2.4. Various analytical solutions for the incompressible Euler and Navier-Stokes systems with free surface

In this paper [24], we propose several time dependent analytical solutions for the incompressible Euler and Navier-Stokes systems with free surface. The given analytical solutions concerns the hydrostatic and nonhydrostatic Euler and Navier-Stokes systems.

7.3. Functional analysis of PDE models in Fluid Mechanics

Participants: Bilal Al Taki, Boris Haspot.

7.3.1. New functional inequality and its application

In [22], we prove by simple arguments a new kind of Logarithmic Sobolev inequalities generalizing two known inequalities founded in some papers related to fluid dynamics models. As a by product, we show how our inequality can help in obtaining some important a priori estimates for the solution of the Navier-Stokes-Korteweg system.

7.3.2. Vortex solutions for the compressible Navier-Stokes equations with general viscosity coefficients in 1D: regularizing effects or not on the density

We consider Navier-Stokes equations for compressible viscous fluids in the one-dimensional case with general viscosity coefficients. We prove the existence of global weak solution when the initial momentum $\rho_0 u_0$ belongs to the set of the finite measure $\mathcal{M}(\mathbf{R})$ and when the initial density ρ_0 is in the set of bounded variation functions $BV(\mathbf{R})$. In particular it allows to deal with initial momentum which are Dirac masses and initial density which admit shocks. We can observe in particular that this type of initial data have infinite energy. Furthermore we show that if the coupling between the density and the velocity is sufficiently strong then the initial density which admits initially shocks is instantaneously regularized and becomes continuous. This coupling is expressed via the regularity of the so called effective velocity $v = u + \frac{\mu(\rho)}{\rho^2}\psi_x\rho$ with $\mu(\rho)$ the viscosity coefficient. Inversely if the coupling between the initial density and the initial velocity is too weak (typically $\rho_0 v_0 \in \mathcal{M}(\mathbf{R})$) then we prove the existence of weak energy in finite time but the density remains a priori discontinuous on the time interval of existence.

7.3.3. Strong solution for Korteweg system

In this paper we investigate the question of the local existence of strong solution for the Korteweg system in critical spaces when $N \geq 1$ provided that the initial data are small. More precisely the initial momentum $\rho_0 u_0$ belongs to $\mathrm{bmo}_T^{-1}(\mathbf{R}^N)$ for $T > 0$ and the initial density ρ_0 is in $L^\infty(\mathbf{R}^N)$ and far away from the vacuum. This result extends the so called Koch-Tataru theorem for the incompressible Navier-Stokes equations to the case of the Korteweg system. It is also interesting to observe that any initial shock on the density is instantaneously regularized inasmuch as the density becomes Lipschitz for any $\rho(t,\cdot)$ with $t > 0$. We also prove the existence of global strong solution for small initial data $(\rho_0 - 1, \rho_0 u_0)$ in the homogeneous Besov

spaces $(\dot{B}^{\mathbf{N}-1}_{2,\infty}(\mathbf{R}^{\mathbf{N}} \cap \dot{B}^{\mathbf{N}}_{2,\infty}(\mathbf{R}^{\mathbf{N}} \cap L^{\infty}(\mathbf{R}^{\mathbf{N}})) \times (\dot{B}^{\mathbf{N}-1}_{2,\infty}(\mathbf{R}^{\mathbf{N}}))^{\mathbf{N}}$. This result allows in particular to extend in dimension $N = 2$ the notion of Oseen solutions defined for incompressible Navier-Stokes equations to the case of the Korteweg system when the vorticity of the momentum $\rho_0 u_0$ is a Dirac mass $\alpha\delta_0$ with α sufficiently small. However unlike the Navier Stokes equations

7.4. Assessments of models by means of experimental data and assimilation

Participants: Vivien Mallet, Ngoc Bao Tran Le, Antoine Lesieur, Frédéric Allaire, Hammond Janelle.

7.4.1. Uncertainty quantification of on-road traffic emissions

Road traffic emissions of air pollutants depend on both traffic flow and vehicle emission factors. At metropolitan scale, traffic flow can be obtained by traffic assignment models, and emission factors can be computed from the traffic flow using COPERT IV formulas. Global sensitivity analyses, especially the computation of Sobol' indices, was carried out for the traffic model and the air pollutant emissions. In the process, the traffic model was replaced by a metamodel, or surrogate model, in order to reduce the high computational burden. The results identified the most important input parameters, e.g., the demand associated with small travel distances (for the traffic flow) or the gasoline car share (for the air pollutant emissions). Furthermore, the uncertainties in traffic flow and pollutant emissions was quantified by propagating into the model the uncertainties in the input parameters. Large ensembles of traffic flows were generated and evaluated with traffic flow measurements.

7.4.2. Uncertainty quantification in atmospheric dispersion of radionuclides

In collaboration with IRSN (Institute of Radiation Protection and Nuclear Safety), we investigated the uncertainties of the atmospheric-dispersion forecasts that are used during an accidental release of radionuclides such as the Fukushima disaster. These forecasts are subject to considerable uncertainties which originate from inaccurate weather forecasts, poorly known source term and modeling shortcomings. In order to quantify the uncertainties, we designed a metamodel and investigated the calibration of the probability distribution of the input variables like the source term or the meteorological variables.

7.4.3. Metamodeling corrected by observational data

An air quality model at urban scale computes the air pollutant concentrations at street resolution based on various emissions, meteorology, imported pollution and city geometry. Because of the computational cost of such model, we previously designed a metamodel using dimension reduction and statistical emulation. Novel work was dedicated to the correction of this metamodel using observational data. The proposed approach builds a corrected metamodel that is still much faster than the original model, but also performs better when compared to new observations.

7.4.4. Sensitivity analysis and metamodeling of an urban noise model

Urban noise mapping models simulate the propagation of noise, originating from emission sources (e.g., road traffic), in all street of a city, based on its geometry. They are subject to uncertainties due to incomplete and erroneous data. We carried out screening studies in order to evaluate the sensitivity of the computed noise to the uncertain data. Further work dealt with the development of a metamodel, which will open the way to uncertainty quantification. The work was carried out with the model NoiseModelling and applied to the noise mapping of Lorient (France).

7.4.5. Monte Carlo simulation and ensemble evaluation for wildland fire propagation

We worked on Monte Carlo simulations of wildland fires. The objective was to evaluate how the uncertainties lying in all the inputs of a fire propagation model can be propagated through the model. A careful review of the literature allowed us to define varying intervals for all the uncertain inputs. The Monte Carlo simulations were then evaluated with ensemble scores, using the observations of the final contours for a number of real cases. The ensemble scores were inspired by classical scores used in meteorology, but were adapted to the nature of the fire observations.

7.4.6. Metamodeling of a complete air quality simulation chain

With the objective of uncertainty quantification, we worked in [15] on the generation of a metamodel for the simulation of urban air quality, using a complete simulation chain including dynamic traffic assignment, the computation of air pollutant emissions and the dispersion of the pollutant in a city. The traffic model and the dispersion model are computationally costly and operate in high dimension. We employed dimension reduction, and coupled it with Kriging in order to build a metamodel for the complete simulation chain.

7.5. Software Devlopments

Participant: Cédric Doucet.

Improvements in the FRESHKISS3D code Several improvements have been achieved in FreshKiss3D :

1. installation step is simpler due to the usage of a YAML file listing third-party libraries;
2. Mac OS is now supported;
3. continuous integration is performed on Ubuntu 16 and OSX with different compilers (GCC, Clang) and different builds (debug, release);
4. a major bug in the computation of fluxes has been fixed;
5. the number of third-party libraries has been minimized (geomalgo, metis4py);
6. build automation is now based on CMake (instead of Waf);
7. documentation has been updated and it is now published during the continuous integration process by means of Gitlab pages;
8. continuous integration has been optimized (better slaves, parallelization)

CASTOR Project-Team

7. New Results

7.1. Block-structured meshes

Participants: Hervé Guillard, Alexis Loyer, Jalal Lakhlili [IPP Garching], Ahmed Ratnani [IPP Garching].

Due to the highly anisotropic character of strongly magnetized plasmas, a crucial point for numerical simulations is the construction of meshes that are aligned on the magnetic flux surfaces computed by Grad-Shafranov equilibrium solvers. This work has studied an original method for the construction of flux aligned grids that respect the magnetic equilibrium topology and that can be applied to block-structured meshes using C^1 finite element methods (Hermite-Bézier/Cubic spline). The method relies on the analysis of the singularities of the magnetic flux function and the construction of the Reeb graph that allows the segmentation of the physical domain into sub-domains that can be mapped to a reference square domain. Once this domain decomposition has been done, the mapping of the sub-domain to reference patches can be done using integration along the streamlines of the flux function [16]. This work was performed in the framework of the EoCoE European project (see section 8.2.1.1).

7.2. Unstructured triangular meshes for tokamaks

Participants: Hervé Guillard, Alexis Loyer, Adrien Loseille [Gamma3 team, Inria Saclay].

The construction of block-structured flux aligned grids that respect the magnetic equilibrium topology experiences difficulties in the SOL region of the tokamaks where the flux lines cross the material walls. As an alternative to the use of block structured meshes, we have studied the construction of unstructured triangular meshes using constrained anisotropic Delaunay mesh generation [16]. This work was also performed in the framework of the EoCoE European project (see section 8.2.1.1).

7.3. Simulations of hydraulic jumps with a turbulent Shallow Water model

Participants: Hervé Guillard, Argiris Delis [Technical University of Crete, Greece], Yih-Chin Tai [National Cheng Kung University, Taiwan].

We have pursued the work realized in 2017, on a new model designed for the computation of turbulent hydraulic jumps. This model is able to describe the oscillatory nature of turbulent hydraulic jumps and as such corrects the deficiency of the classical shallow water equations. The comparisons with experiments performed at Tainan University are very satisfactory given the simplicity of the model. A journal paper [3] on this subject have been published and these results have been presented at the ETAMM2018 (Emerging Trends in Applied Mathematics and Mechanics 2018) conference.

7.4. 2D C^1 triangular elements

Participants: Hervé Guillard, Ali Elarif, Boniface Nkonga.

In order to avoid some mesh singularities that arise when using quadrangular elements for complex geometries and flux aligned meshes, the use of triangular elements is a possible option that we have studied in the past years. In particular, we have developped the geometric tools necessary for the construction of Powell-Sabin splines and have applied these methods for the approximation of some simple hyperbolic PDE systems (namely the Euler equation of fluid dynamics [6]). The PhD thesis of Ali Elarif that has begun in october 2017 is devoted to the study of the applicability of these methods to more complex PDE models encountered in plasma physics and to an extension towards other triangular C^1 elements (Clough-Tocher elements). The work realized this year has allowed to apply these finite element spaces to the approximation of elliptic equations and to design penalization methods to enforce non-homogeneous Dirichlet boundary conditions. In particular, the use of reduced Clough-Tocher elements has been applied to obtain solution of the free-boundary non-linear Grad-Shafranov equation. The results show that the use of these C^1 elements produce results that are smoother than the ones obtained with low order P1 elements.

7.5. Equilibrium reconstruction at JET using Stokes model for polarimetry

Participant: Blaise Faugeras.

This paper presents the first application to real JET data of the new equilibrium code NICE which enables the consistent resolution of the inverse equilibrium reconstruction problem in the framework of non-linear free-boundary equilibrium coupled to the Stokes model equation for polarimetry. The conducted numerical experiments enable first of all to validate NICE by comparing it to the well-established EFIT code on 4 selected high performance shots. Secondly the results indicate that the fit to polarimetry measurements clearly benefits from the use of Stokes vector measurements compared to the classical case of Faraday measurements, and that the reconstructed p' and ff' profiles are better constrained with smaller error bars and are closer to the profiles reconstructed by EFTM, the EFIT JET code using internal MSE constraints.

7.6. Operational plasma boundary reconstruction with the NICE-VacTH code on WEST Tokamak

Participant: Blaise Faugeras.

A new regularization term has been proposed for the inverse problem of plasma boundary reconstruction using an expansion of the poloidal flux in toroidal harmonics. It has been implemented in the VacTH code and is used successfully on the WEST Tokamak.

7.7. Equilibrium reconstruction with NICE at WEST and within the framework of the European Integrated Tokamak Modelling WPCD project

Participant: Blaise Faugeras.

The adaptation of NICE to IMAS (the ITER standard using IDS as data type) has been carried on. Equilibrium reconstructions using IMAS have been performed on real JET measurements and are now performed routinely at WEST.

7.8. Equilibrium reconstruction with Equinox at JET

Participant: Blaise Faugeras.

The adaptation of NICE to IMAS the ITER standard using IDS as data type has been carried on. Equilibrium reconstructions using IMAS have been performed on real JET measurements and are now performed routineley at WEST.

7.9. Evolutive mode and iron model in NICE

Participants: Blaise Faugeras, Jacques Blum, Cédric Boulbe.

The capabilities of the equilbirum code NICE have been extended. The evolutive direct model and the iron model of the free boundary equilibrium code CEDRES++ have been ported in NICE.

7.10. Coupling CEDRES++ - WEST controller in IMAS

Participants: Cédric Boulbe, Jakub Urban [IPP Prague].

The free boundary equilibrium code has been fully adapted to IMAS and has been coupled to the magnetic controller of WEST. The code CEDRES++ simulate the plant and the controller provide the voltages applied to the PF supplies. This coupling has enabled to develop a tool in Python to interface easily Simulink controllers with IMAS. With that tool, it is possible to run a controller installed on a distant computer and to run it from IMAS. As a test case, the WEST controller has been interfaced with IMAS and coupled to CEDRES++ using an IMAS python workflow.

7.11. Spectral Element method for high order partial differential equations

Participants: Sebastian Minjeaud, Richard Pasquetti.

The Korteweg-de Vries equation has been addressed as an interesting model of high order partial differential equation. In [9] it is shown that it is possible to develop reliable and effective schemes, in terms of accuracy, computational efficiency, simplicity of implementation and, if required, conservation of the lower invariants, on the basis of a (only) H^1-conformal Galerkin approximation, namely the Spectral Element Method. The proposed approach is *a priori* easily extensible to other partial differential equations and to multidimensional problems.

7.12. Recent advances in Spectral element methods on simplicial meshes

Participants: Richard Pasquetti, Francesca Rapetti.

R. Pasquetti and F. Rapetti have investigated the cubature points based triangular spectral element method. Using cubature points, both for interpolations and quadratures, shows the advantage of yielding a diagonal mass matrix. Accuracy results are provided in [10], for elliptic problems in non polygonal domains, using various isoparametric mappings. The capabilities of the method are here again clearly confirmed.

7.13. Full-MHD with Jorek

Participants: Boniface Nkonga, Ashish Bhole.

In the context of A. Bohle PhD, we have developped a strategy to improve the formulation of finite element space in the context of iso-parametric finite elements with singular parametrization. This result in a set of constraints to be applied in the numerical formulation to fit in the well defined approximated space. Applied to interpolations, we recover the optimal order of convergence of the numerical approximation. Next step is applications to the resolution of reduced-MHD and then full-MHD.

7.14. A discontinuous Galerkin method for a two dimensional resistive MHD model

Participants: Ashish Bhole, Boniface Nkonga, Praveen Chandrashekar.

We consider the numerical approximation of two dimensional incompressible magnetohydrodynamics equations with vorticity and current as the dynamical variables. We construct a discontinuous Galerkin (DG) method for the MHD model written in symmetric form. The numerical flux is based on a Riemann solver and the scalar fluxes of velocity and magnetic field are computed using a Galerkin method. The performance of the method is demonstrated on some standard instability problems relevant to magnetically confined fusion reactors.

7.15. Fluctuation splitting Riemann solver for a non-conservative shear shallow water flow

Participants: Ashish Bhole, Boniface Nkonga, Sergey Gavrilyuk.

We propose a fluctuation splitting finite volume scheme for a non-conservative modeling of shear shallow water flow (SSWF). This model was originally proposed by Teshukov and was extended to include modeling of friction by Gavrilyuk (2018). We develop a cell-centered finite volume code to validate the proposed scheme with the help of some numerical tests. As expected, the scheme shows first order convergence. The numerical simulation of 1D roll waves shows a good agreement with the experimental results. The numerical simulations of 2D roll waves show similar transverse wave structures as observed by Gavrilyuk (Paper in revision at JCP).

7.16. Automating the design of Tokamak experiment scenarios

Participants: Jacques Blum, Holger Heumann, Xiao Song.

The real-time control of plasma position, shape and current in a tokamak has to be ensured by a number of electrical circuits consisting of voltage suppliers and axisymmetric coils. Finding good target voltages/currents for the control systems is a very laborious, non-trivial task due to non-linear effects of plasma evolution. We introduce here an optimal control formulation to tackle this task and present in detail the main ingredients for finding numerical solutions: the finite element discretization, accurate linearizations and Sequential Quadratic Programming. Case studies for the tokamaks WEST and HL2M highlight the exibility and broad scope of the proposed optimal control formulation.

7.17. Multiscales scheme for the MHD model in a tokamak

Participants: Hervé Guillard, Afeintou Sangam.

Recently, in [21], it is proven that the Reduced MHD equations are a singular limit of the Full MHD system when the inverse ratio parameter goes to zero. In this limit, the toroidal dynamics is almost entirely decoupled from the incompressible poloidal dynamics. From a numerical point of view, in this limit, the propagation of fast magnetosonic waves severely constraints the time step in explicit schemes. A possible remedy is therefore to design a semi-implicit time stepping strategy allowing an implicit handling of the fast waves but retaining an explicit treatment of the slow ones. In this work, we have derived a linear simplified model in two dimensions that retains the main characteristics of the formal passage from the Full MHD equations to the Reduced MHD system. A semi-implicit numerical scheme free of time step restrictions based on the fast wave velocity has been constructed for this model. The extension of this numerical scheme to the Full MHD model is under investigation.

7.18. Asymptotic Transport Models for heat and mass transport in reactive porious media

Participants: Bruno Dubroca, Afeintou Sangam.

Charrier and *Dubroca* in [20], have suggested an approach to derive rigously a family of models of mass and heat transfer in reactive porous media. At a microscopic level they proposed a model coupling the Boltzmann equation in the gas phase, the heat equation and appropriate interface conditions, including adsorption-deposition reactions. Then an asymptotic expansion mixing homogenization and fluid limit leads to a system of coupled diffusion equations where the effective diffusion tensors are defined from the microscopic geometry of the material. Open questions paved their work. We solve one of them, consisting in setting adequate conditions on interest models that ensure the uniqueness of solutions of the first order expansion. They are based on the concept of thermodynamically closed in average system.

COFFEE Project-Team

7. New Results

7.1. A few words on the results of the year

- Face based discretization of two-phase Darcy flows in fractured porous medium with matrix fracture interface local nonlinear solver. Application to the simulation of the desaturation by suction in nuclear waster storages [20], [17].
- Convergence analysis of the gradient discretization of a two-phase Darcy flow model in fractured porous media with nonlinear transmission conditions [10].
- Numerical method for non-isothermal compositional Darcy flows combining face based and nodal based discretizations on hybrid meshes [22].
- We introduced and analyzed a novel Hybrid High-Order method for the steady incompressible Navier-Stokes equations. We showed under general assumptions the existence of a discrete solution, we proved convergence of the sequence of discrete solutions to minimal regularity exact solutions for general data and we proved optimal convergence rates for the velocity and the pressure [9].
- We proposed a nonlinear Discrete Duality Finite Volume scheme to approximate the solutions of drift diffusion equations. The scheme is built to preserve at the discrete level even on severely distorted meshes the energy / energy dissipation relation [7].
- We studied a Discrete Duality Finite Volume scheme for the unsteady incompressible Navier-Stokes problem with outflow boundary conditions [24].
- We introduced a new non-overlapping optimized Schwarz method for anisotropic diffusion problems. We studied the new method at the continuous level, proved its convergence using energy estimates, and also derived convergence factors to determine the optimal choice of parameters in the transmission conditions, and presented a discretization of the algorithm using discrete duality finite volumes [23].
- We consider a non-local traffic model involving a convolution product. Unlike other studies, the considered kernel is discontinuous on $\mathbb{R}$. We prove Sobolev estimates and prove the convergence of approximate solutions solving a viscous and regularized non-local equation. It leads to weak, $C([0,T],L^2(\mathbb{R}))$ and $C([0,T],L^2(\mathbb{R}))$, and smooth, $W^{2,2N}([0,T]\times\mathbb{R})$ and $W^{2,2N}([0,T]\times\mathbb{R})$, solutions for the non-local traffic model [4].
- We proposed a new closure for Geometrical Shock Dynamics taking into account the effect of transverse Mach variation for the fast propagation of shocks. The model has been tested using a Lagrangian solver [28].
- We proposed a new explicit pseudo-energy conserving time-integration scheme for separated Hamiltonian systems. We proved the second-order accuracy and conditional stability of the scheme. In addition, the scheme can be adapted into an asynchronous version while retaining its properties, which is adapted to slow-fast splitting strategies [14].
- We proposed a well-balanced scheme for the modified Lifschitz-Slyozov-Wagner system with diffusion, which models Ostwald ripening. The scheme outperforms a standard advection-diffusion scheme for long time dynamics [25].
- We investigate several models describing interacting particles, either motivated form physics or population dynamics.

FLUMINANCE Project-Team

6. New Results

6.1. Fluid motion estimation

6.1.1. Stochastic uncertainty models for motion estimation

Participants: Musaab Khalid Osman Mohammed, Etienne Mémin.

This work is concerned with the design of motion estimation technique for image-based river velocimetry. The method proposed is based on an advection diffusion equation associated to the transport of large-scale quantity with a model of the unresolved small-scale contributions. Additionally, since there is no ground truth data for such type of image sequences, a new evaluation method to assess the results has been developed. It is based on trajectory reconstruction of few Lagrangian particles of interest and a direct comparison against their manually-reconstructed trajectories. The new motion estimation technique outperformed traditional optical flow and PIV-based methods used in hydrology [24]. This study has been performed within the PhD thesis of Musaab Khalid [18] and through a collaboration with the Irstea Lyon hydrology research group (HHLY).

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: (i) the development of precise models relating the large-scale flow observations to the velocity; (ii) appropriate large-scale regularization strategies; and (iii) adapted seeding and lighting systems, like Hellium Filled Soap Bubles (HFSB) and led ramp lighting. This work conducted within the PhD of Romain Schuster in collaboration with the compagny ITGA has started in february 2016. The first step has been to evaluate the performances of a stochastic uncertainty motion estimator when using large scale scalar images, like those obtained when seeding a flow with smoke. The PIV characterization of flows on large fields of view requires an adaptation of the motion estimation method from image sequences. The backward shift of the camera coupled to a dense scalar seeding involves a large scale observation of the flow, thereby producing uncertainty about the observed phenomena. By introducing a stochastic term related to this uncertainty into the observation term, we obtained a significant improvement of the estimated velocity field accuracy. The technique was validated on a mixing layer in a wind tunnel for HFSB and smoke tracers [40] and applied on a laboratory fume-hood [39]. This study demonstrated the feasibility of conducting on-site large-scale image-based measurements for indoor airflows characterization.

6.1.3. 3D flows reconstruction from image data

Participants: Dominique Heitz, Etienne Mémin.

Our work focuses on the design of new tools for the estimation of 3D turbulent flow motion in the experimental setup of Tomo-PIV. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms. This year, we continued our investigation on the problem of efficient volume reconstruction. We have developed an ensemble assimilation methods for the fast reconstruction of 3D tomo-PIV motion field. The approach relies on a simplified dynamics of the flow and is a generalization of one of the popular emerging flow reconstruction technique of the PIV community referred to as "Shake the box » (STB). The study developed within an Irstea post-doctoral funding introduced a novel approach originated from the data assimilation technique comprising a sampling-based optimal estimation algorithm, namely a group of ensemble-based filtering variational schemes. We found that employing such an ensemble-based optimal estimation method helped

tackling the problems associated with STB: the inaccurate predictor and/or the robustness of the optimization procedure. The proposed method (ENS) was quantitatively evaluated with synthetic particle image data built by transporting virtual particles in a turbulent cylinder wake-flow at Reynolds number equal to 3900 [43]. The study will be continued within an Irstea doctoral funding.

6.2. Tracking, Data assimilation and model-data coupling

6.2.1. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin, Gilles Tissot.

In this axis of work, we explore 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 planes enlightened with laser sheets. This approach is experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we aim to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models. A 4DVar assimilation technique based on the numerical code Incompact3D has been implemented for that purpose to control the inlet and initial conditions in order to reconstruct a turbulent wake flow behind an unknown obstacle [45]. We extended this first data assimilation technique to control the subgrid parameters. This study is performed in collaboration with Sylvain Laizet (Imperial College). In another axis of research, in collaboration with the CSTB Nantes centre and within the PhD of Yacine Ben Ali we will explore the definition of efficient data assimilation schemes for wind engineering. The goal is here to couple Reynolds average model to pressure data at the surface of buildings. The final purpose will consist in proposing improved data-driven simulation models for architects.

6.2.2. Ensemble variational data assimilation of large-scale dynamics with uncertainty

Participant: Etienne Mémin.

Estimating the parameters of geophysical dynamic models is an important task in Data Assimilation (DA) technique used for forecast initialization and reanalysis. In the past, most parameter estimation strategies were derived by state augmentation, yielding algorithms that are easy to implement but may exhibit convergence difficulties. The Expectation-Maximization (EM) algorithm is considered advantageous because it employs two iterative steps to estimate the model state and the model parameter separately. In this work, we propose a novel ensemble formulation of the Maximization step in EM that allows a direct optimal estimation of physical parameters using iterative methods for linear systems. This departs from current EM formulations that are only capable of dealing with additive model error structures. This contribution shows how the EM technique can be used for dynamics identification problem with a model error parameterized as arbitrary complex form. The proposed technique is used for the identification of stochastic subgrid terms that account for processes unresolved by a geophysical fluid model. This method, along with the augmented state technique, has been evaluated to estimate such subgrid terms through high resolution data. Compared to the augmented state technique, our method is shown to yield considerably more accurate parameters. In addition, in terms of prediction capacity, it leads to smaller generalization error as caused by the overfitting of the trained model on presented data and eventually better forecasts [27].

6.2.3. Reduced-order models for flows representation from image data

Participants: Dominique Heitz, Etienne Mémin, Gilles Tissot.

During the PhD thesis of Valentin Resseguier we have proposed a new decomposition of the fluid velocity in terms of a large-scale continuous component with respect to time and a small-scale non continuous random component. Within this general framework, an uncertainty based representation of the Reynolds transport theorem and Navier-Stokes equations can be derived, based on physical conservation laws. This physically relevant stochastic model has been applied in the context of POD-Galerkin methods. This uncertainty modeling methodology provides a theoretically grounded technique to define an appropriate subgrid tensor as well as drift correction terms. The pertinence of this stochastic reduced order model has been successfully assessed on several wake flows at different Reynold number. It has been shown to be much more stable than the usual reduced order model construction techniques. Beyond the definition of a stable reduced order model, the modeling under location uncertainty paradigm offers a unique way to analyse from the data of a turbulent flow the action of the small-scale velocity components on the large-scale flow. Regions of prominent turbulent kinetic energy, direction of preferential diffusion, as well as the small-scale induced drift can be identified and analyzed to decipher key players involved in the flow. This study has been published in the Journal of Fluid Mechanics [15]. Note that these reduced order models can be extended to a full system of stochastic differential equations driving all the temporal modes of the reduced system (and not only the small-scale modes). This full stochastic system has been evaluated on wake flow at moderate Reynolds number. For this flow the system has shown to provide very good uncertainty quantification properties as well as meaningful physical behavior with respect to the simulation of the neutral modes of the dynamics. This study described in the PhD of Valentin Resseguier will be soon submitted to a journal paper.

6.2.4. Estimation and control of amplifier flows

Participant: Gilles Tissot.

Estimation and control of fluid systems is an extremely hard problem. The use of models in combination with data is central to take advantage of all information we have on the system. Unfortunately all flows do not present the same physical and mathematical behaviour, thus using models and methodologies specialised to the flow physics is necessary to reach high performances.

A class of flows, denoted "oscillator flows", are characterised by unstable modes of the linearised operator. A consequence is the dominance of relatively regular oscillations associated with a nonlinear saturation. Despite the non-linear behaviour, associated structures and dynamical evolution are relatively easy to predict. Canonical configurations are the cylinder wake flow or the flow over an open cavity.

By opposition to that, "amplifier flows" are linearly stable with regard to the linearised operator. However, due to their convective nature, a wide range of perturbations are amplified in time and convected away such that it vanishes at long time. The consequence is the high sensitivity to perturbations and the broad band response that forbid a low rank representation. Jets and mixing layers show this behaviour and a wide range of industrial applications are affected by these broad band perturbations. It constitutes then a class of problems that are worth to treat separately since it is one of the scientific locks that render hard the transfer of methodologies existing in flow control and estimation to industrial applications.

There exists a type of models, that we will denote as "parabolised", that are able to efficiently represent amplifier flows. These models, such as parabolised stability equations and one-way Navier-Stokes propagate, in the frequency domain, hydrodynamic instability waves over a given turbulent mean flow. We can note that these models, by their structure, give access to a natural experimental implementation. They are an ingredient adapted to represent the system, but have a mathematical structure strongly different from the dynamical models classically used in control and data assimilation. It is then important to develop new methodologies of control, estimation and data assimilation with these models to reach our objectives. Moreover, inventing new models by introducing the modelling under location uncertainties in these parabolised models will be perfectly adapted to represent the evolution and the variability of an instability propagating within a turbulent flow. It will be consistent with actual postprocessing of experimental data performed in similar flow configurations.

6.3. Analysis and modeling of turbulent flows and geophysical flows

6.3.1. Geophysical flows modeling under location uncertainty

Participants: Werner Bauer, Long Li, Etienne Mémin.

In this research axis we have devised a principle to derive representation of flow dynamics under uncertainty. Such an uncertainty is formalized through the introduction of a random term that enables taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also partially known forcing. Rigorously derived from a stochastic version of the Reynolds transport theorem [9], this framework, referred to as modeling under location uncertainty, encompasses several meaningful mechanisms for turbulence modeling. It indeed introduces without any supplementary assumption the following pertinent mechanisms for turbulence modeling: (i) a dissipative operator related to the mixing effect of the large-scale components by the small-scale velocity; (ii) a multiplicative noise representing small-scale energy backscattering; and (iii) a modified advection term related to the so-called *turbophoresis* phenomena, attached to the migration of inertial particles in regions of lower turbulent diffusivity.

In a series of papers we have shown how such modeling can be applied to provide stochastic representations of a variety of classical geophysical flows dynamics [12], [13], [14]. Numerical simulations and uncertainty quantification have been performed on Quasi Geostophic approximation (QG) of oceanic models. It has been shown that such models lead to remarkable estimation of the unresolved errors at variance to classical eddy viscosity based models. The noise brings also an additional degree of freedom in the modeling step and pertinent diagnostic relations and variations of the model can be obtained with different scaling assumptions of the turbulent kinetic energy (i.e. of the noise amplitude). The performances of such systems have been assessed also on an original stochastic representation of the Lorenz 63 derived from the modeling under location uncertainty [22]. In this study it has been shown that the stochastic version enabled to explore in a much faster way the region of the deterministic attractor. This effort has been undertaken within a fruitful collaboration with Bertrand Chapron (LOPS/IFREMER). In the PhD of Long Li, we continue this effort. The goal is to propose relevant techniques to define or calibrate the noise term from data. In that prospect, we intend to explore statistical learning techniques.

6.3.2. *Large eddies simulation models under location uncertainty*

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

The models under location uncertainty recently introduced by Mémin (2014) [9] provide a new outlook on LES modeling for turbulence studies. These models are derived from a stochastic transport principle. The associated stochastic conservation equations are similar to the filtered Navier- Stokes equation wherein we observe a sub-grid scale dissipation term. However, in the stochastic version, an extra term appears, termed as "velocity bias", which can be treated as a biasing/modification of the large-scale advection by the small scales. This velocity bias, introduced artificially in the literature, appears here automatically through a decorrelation assumption of the small scales at the resolved scale. All sub-grid contributions for the stochastic models are defined by the small-scale velocity auto-correlation tensor. This large scale modeling has been assed and compared to several classical large-scale models on a flow over a circular cylinder at Re $\sim$ 3900 [21] and wall-bounded flows [45]. For all these flows the modeling under uncertainty has provided better results than classical large eddies simulation models. Within the PhD of Yacine Ben Ali we will explore with the CSTB Nantes centre the application of such models for the definition of Reynolds average simulation (RANS) models for wind engineering applications.

6.3.3. *Variational principles for structure-preserving discretizations in stochastic fluid dynamics*

Participants: Werner Bauer, Long Li, Etienne Mémin.

The overarching goal of this interdisciplinary project is to use variational principles to derive deterministic and stochastic models and corresponding accurate and efficient structure preserving discretizations and to use these schemes to obtain a deeper understanding of the principle conservation laws of stochastic fluid dynamics. The newly developed systematic discretization framework is based on discrete variational principles whose highly structured procedures shall be exploited to develop a general software framework that applies automatic code generation. This project, will first provide new stochastic fluid models and suitable approximations, with

potential future applications in climate science using the developed methods to perform accurate long term simulations while quantifying the solutions' uncertainties. The generality of our approach addresses also other research areas such as electrodynamics (EDyn), magnetohydrodynamics (MHD), and plasma physics.

6.3.4. Singular and regular solutions to the Navier-Stokes equations (NSE) and relative turbulent models

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The common thread of this work is the problem set by J. Leray in 1934 : does a regular solution of the Navier-Stokes equations (NSE) with a smooth initial data develop a singularity in finite time, what is the precise structure of a global weak solution to the Navier-Stokes equations, and are we able to prove any uniqueness result of such a solution. This is a very hard problem for which there is for the moment no answer. Nevertheless, this question leads us to reconsider the theory of Leray for the study of the Navier-Stokes equations in the whole space with an additional eddy viscosity term that models the Reynolds stress in the context of large-scale flow modelling. It appears that Leray's theory cannot be generalized turnkey for this problem, so that things must be reconsidered from the beginning. This problem is approached by a regularization process using mollifiers, and particular attention must be paid to the eddy viscosity term. For this regularized problem and when the eddy viscosity has enough regularity, we have been able to prove the existence of a global unique solution that is of class C^∞ in time and space and that satisfies the energy balance. Moreover, when the eddy viscosity is of compact support in space, uniformly in time, we recently shown that this solution converges to a turbulent solution to the corresponding Navier-Stokes equations,carried when the regularizing parameter goes to 0. These results are described in a paper that has been submitted to the journal Archive for Rational Mechanics and Analysis (ARMA).

Within a collaboration with L. Berselli (Univ. Pisa, Italy) we have achieved a study on the well known Bardina's turbulent model [20]. In this problem, we considered the Helmholtz filter usually used within the framework of Large Eddy Simulation. We carried out a similar analysis, by showing in particular that no singularity occurs for Bardina's model.

Within the same collaboration, we considered the three dimensional incompressible Navier-Stokes equations with non stationary source terms chosen in a suitable space. We proved the existence of Leray-Hopf weak solutions and that it is possible to characterize (up to sub-sequences) their long-time averages, which satisfy the Reynolds averaged equations, involving a Reynolds stress. Moreover, we showed that the turbulent dissipation is bounded by the sum of the Reynolds stress work and of the external turbulent fluxes, without any additional assumption, than that of dealing with Leray-Hopf weak solutions. Finally, we considered ensemble averages of solutions, associated with a set of different forces and we proved that the fluctuations continue to have a dissipative effect on the mean flow.

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 have tested the performances of an incompressible turbulence Reynolds-Averaged Navier-Stokes one-closure equation model in a boundary layer, which requires the determination of the mixing length ℓ. A series of direct numerical simulation have been performed, with flat and non trivial topographies, to obtain by interpolation a generic formula $\ell = \ell(Re_\star, z)$, $Re_\star$ being the frictional Reynolds number, and z the distance to the wall. Numerical simulations have been carried out at high Reynolds numbers with this turbulence model, in order to discuss its ability to properly reproduce the standard profiles observed in neutral boundary layers, and to assess its advantages, its disadvantages and its limits. We also proceeded to a mathematical analysis of the model.

6.3.5. Stochastic flow model to predict the mean velocity in wall bounded flows

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

To date no satisfying model exists to explain the mean velocity profile within the whole turbulent layer of canonical wall bounded flows. We propose a modification of the velocity profile expression that ensues from the stochastic representation of fluid flows dynamics proposed recently in the group and referred to as "modeling under location uncertainty". This framework introduces in a rigorous way a subgrid term

generalizing the eddy-viscosity assumption and an eddy-induced advection term resulting from turbulence inhomogeneity. This latter term gives rise to a theoretically well-grounded model for the transitional zone between the viscous sublayer and the turbulent sublayer. An expression of the small-scale velocity component is also provided in the viscous zone. Numerical assessment of the results have been performed for turbulent boundary layer flows, pipe flows and channel flows at various Reynolds numbers [49].

6.3.6. Numerical and experimental image and flow database

Participants: Pranav Chandramouli, Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The first challenge was to enlarge a database mainly based on two-dimensional flows, with three-dimensional turbulent flows. Synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with time resolved Particle Image Velocimetry measurements in wake and mixing layers flows. This database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images. The second challenge was to carry 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 have been be used within a 4DVar assimilation technique, to reconstruct three-dimensional turbulent flows from data [45]. The third challenge was to carry out a time resolved tomoPIV experiments in a turbulent wake flow.

6.3.7. Fast 3D flow reconstruction from 2D cross-plane observations

Participants: Pranav Chandramouli, Dominique Heitz, Etienne Mémin.

We proposed a computationally efficient flow reconstruction technique, exploiting homogeneity in a given direction, to recreate three dimensional instantaneous turbulent velocity fields from snapshots of two dimension planar fields. This methodology, termed as 'snapshot optimisation' or SO, can help provide 3D data-sets for studies which are currently restricted by the limitations of experimental measurement techniques. The SO method aims at optimising the error between an inlet plane with a homogeneous direction and snap-shots, obtained over a sufficient period of time, on the observation plane. The observations are carried out on a plane perpendicular to the inlet plane with a shared edge normal to the homogeneity direction. The method is applicable to all flows which display a direction of homogeneity such as cylinder wake flows, channel flow, mixing layer, and jet (axi-symmetric). The ability of the method is assessed with two synthetic data-sets, and three experimental PIV data-sets. A good reconstruction of the large-scale structures are observed for all cases. This study has been recently submitted to the journal "Experiments in Fluids".

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 controlling one of the prototypical flow configurations encountered in fluid mechanics: the spatially developing turbulent shear layer occuring between two parallel incident streams with different velocities. Our goal is to maintain the shear-layer in a desired state and thus to reject upstream perturbations. In our conference IFAC paper (https://hal.inria.fr/hal-01514361) we focused on perturbations belonging to the same space that the actuators, concretely that means that we were only able to face perturbations of the actuator itself, like failures of the actuator. This year we enlarged this result to purely exogenous perturbations. An optimal control law has been derived to minimize the influence of the perturbation on the flow. To do that, an on-line estimation of the perturbation has been used. This work will be submitted to the upcoming IEEE Conference on Decision and Control. We have also generalized the works initiated during the post-doctoral stay of Tudor-Bogdan Airimitoaie (https://hal.archives-ouvertes.fr/hal-01101089) concerning the benefits of increasing the controlled degrees of freedom in the particular case of the heat equation. This work has been validated on the shear flow.

6.5. Coupled models in hydrogeology

6.5.1. Characterizations of Solutions in Geochemistry

Participants: Jocelyne Erhel, Tangi Migot.

We study the properties of a geochemical model involving aqueous and precipitation-dissolution reactions at a local equilibrium in a diluted solution. This model can be derived from the minimization of the free Gibbs energy subject to linear constraints. By using logarithmic variables, we define another minimization problem subject to different linear constraints with reduced size. The new objective function is strictly convex, so that uniqueness is straightforward. Moreover, existence conditions are directly related to the totals, which are the parameters in the mass balance equation. These results allow us to define a partition of the totals into mineral states, where a given subset of minerals are present. This precipitation diagram is inspired from thermodynamic diagrams where a phase depends on physical parameters. Using the polynomial structure of the problem, we provide characterizations and an algorithm to compute the precipitation diagram. Numerical computations on some examples illustrate this approach. This work, supported by ANDRA, was presented at a workshop [29] and will be published in the journal Computational Geosciences [23].

6.5.2. Reactive transport in multiphase flow

Participants: Jocelyne Erhel, Bastien Hamlat.

Reactive transport modeling is a critical issue for energy industry, with application to carbon capture and storage (CCS) and geothermy. In this work, we focus on mathematical modeling and numerical solution of reactive transport in porous media. We study reaction kinetics leading to the appearance or disappearance of pure phases and present a mathematical model of PDEs with discontinuous right-hand sides. We propose a regularization process to obtain a PDE system with continuous right-hand sides which can be numerically solved. Using a simple academic case, we illustrate the benefits of our approach.

This work, supported by IFPEN, was presented at a workshop [30] and an international conference [35].

6.6. Sparse Linear solvers

6.6.1. Parallel GMRES

Participant: Jocelyne Erhel.

Sparse linear systems $Ax = b$ arise very often in computational science and engineering. Krylov methods are very efficient iterative methods, and restarted GMRES is a reference algorithm for non-symmetric systems. A first issue is to ensure a fast convergence, by preconditioning the system with a matrix M. Preconditioning must reduce the number of iterations, and be easy to solve. A second issue is to achieve high performance computing. The most time-consuming part in GMRES is to build an orthonormal basis V. With the Arnoldi process, many scalar products involve global communications. In order to avoid them, s-step methods have been designed to find a tradeoff between parallel performance and stability. Also, solving a system with the matrix M and multiplying a vector by the matrix A should be efficient. A domain decomposition approach involves mainly local communications and is frequently used. A coarse grid correction, based on deflation for example, improves convergence. These techniques can be combined to provide fast convergence and fast parallel algorithms. Numerical results illustrate various issues and achievements.

This work was presented at an international conference (invited talk) [28].

LEMON Team

7. New Results

7.1. Inland flow processes

7.1.1. Shallow water models with porosity

7.1.1.1. DDP model.

A new porosity model was published in 2018. The Depth-Dependent Porosity (DDP) model [4] was developed to account for subgrid-scale topographical features in shallow water models. The purpose is to allow flows to be modelled using coarse grids in the presence of strongly contrasted topography (e.g. ditches, narrow channels, submerged obstacles). Applications range from the modelling of lagoon/wetland dynamics to the submersion of urban areas by dambreak or tsunami waves. The development is the result of a team work in cooperation with the Tour du Valat research institute (O. Boutron). The developments are incorporated in the SW2D code.

7.1.1.2. Porosity model validation.

The first experimental results validating the Dual Integral Porosity (DIP) model were presented at the RiverFlow 2018 International conference [9]. This work was carried out in collaboration with S. Soares-Frazão at Université Catholique de Louvain (UCL). Two stays of Carole Delenne and Vincent Guinot at UCL to participate in the experimental campaign in 2017 had been supported financially by the LEMON budget. A journal article presenting these experiments in detail has been submitted to the Journal of Hydraulic Research and is currently awaiting the final decision.

7.1.2. Forcing

Stochastic approaches can be used to generate forcing scenarios randomly. To this end, an accurate characterization of the spatio-temporal variability and rainfall intensity distribution must be obtained from available data. So we have deeply studied a gridded hourly rainfall dataset in a region in Mediterranean France and proposed a semiparametric method to simulate spatio-temporal scenarios for extreme events. Our work was presented in the following international conferences: METMA 2018 - 9th Workshop on Spatio-temporal modelling (June 2018, Montpellier) [11] and SWGEN 2018 - Stochastic Weather Generators Conference (October 2018, Boulder, United States)[10]. Moreover we have invited P. Naveau (CNRS, LSCE) during two weeks and we have begun a collaboration concerning the construction of a unique temporal model permitting to deal with both ordinary and extreme events.

7.1.3. Inland hydrological systems

The PhD of Joseph Luis Kahn Casapia (co-advised by Antoine Rousseau and Céline Casenave from INRA) has just started (Oct. 2018). The objective of the thesis is the modelling of cyanobacteria blooms in shallow water lakes such as TaiHu, in China. This work is done in the framework of the ANSWER research project funded by ANR, with a co-funding by labex NUMEV in Montpellier.
A publication presenting the KarstMod modelling platform was accepted in Environmental Modelling & Software [6]. KarstMod incorporates a number of developments by Vincent Guinot, including the Hysteretic [90] and the infinite characteristic time [60] transfer functions.
A family of multi-region transport models in heterogeneous porous media have been derived and validated experimentally [5]. The publication with Carole Delenne and Vincent Guinot as co-authors presents not only experimental results, but also a theoretical analysis of the transport and dispersion properties of a variety of models, depending on the structure of the flow field.

7.1.4. Parametrization

With an objective of assessing flood hazard at large scale, the CASACADE project has been funded by the Luxembourg National Research Fund and provides for a phD (started in November 2018 by Vita Ayoub) concerning assimilation of satellite derived flood information for better parameterizing and controlling large scale hydraulic models over data scarce areas. One of the model used will be the DDP model [4]. The effective integration of remote sensing-derived flood information into this model will be investigated in this project for retrieving uncertain model parameters and boundary conditions. The PhD is co-directed by Carole Delenneand Renaud HOSTACHE from the Luxembourg Institute of Sciences and Technologies (LIST).

7.2. Marine and coastal systems

7.2.1. Multi-scale ocean modelling

We proposed in [2] a Schwarz-based domain decomposition method for solving a dispersion equation consisting on the linearized KdV equation without the advective term, using simple interface operators based on the exact transparent boundary conditions for this equation. An optimization process is performed for obtaining the approximation that provides the method with the fastest convergence to the solution of the monodomain problem.
We also moved towards more complex equations and derived in [13], [12] discrete transparent boundary conditions for a class of linearized Boussinesq equations. These conditions happen to be non-local in time and we test numerically their accuracy with a Crank-Nicolson time-discretization on a staggered grid. We used the derived transparent boundary conditions as interface conditions in a domain decomposition method, where they become local in time. We analyzed numerically their efficiency thanks to comparisons made with other interface conditions. A paper [19] is submitted for publication in addition to the aforementioned talks.

7.2.2. Data-model interactions

To go further with what have been explained in subsection Forcings, there are clear advantages of thresholding techniques in stochastic approaches aiming to simulate extreme events. They permit to exploit information from more data (compared to the block-maxima approach) and to explicitely model the original event. Pareto processes have been mostly used in a parametric framework, thereby using assumptions on the choce of the underlying dependence structure that may be strong. We have proposed a semi-parametric approach ([10], [11]) and we have shown the links between this semi-parametric method and the Pareto processes. A key benefit of the proposed method is to allow the generation of an unlimited number of realizations of these extreme fields. This work will be submitted for publication during the first trimester of 2019.

7.3. Methodological developments

7.3.1. Stochastic models for extreme events

In extreme value theory, there is two main approaches. The first one is based on block maxima and involve max-stable models. Indeed the use of extreme value copula for extreme events is justified by the theory of multivariate extreme. The most accessible models are too simplistic when they are used in a high-dimensional framework. That is why we have proposed in a spatial context to combine two Gumbel copulas. By doing that, we reduce the complexity considering the weight parameter as a function depending on covariates. Moreover interpolation becomes straightforward and enable the interpretation of the parameters with distances between sites. Properties of the proposed model such as the possible extremal dependencies varying in space are studied and inference relies on ABC techniques. This work will be submitted for publication during the first quarter of 2019 (see also [14]).

The second approach is based on high threshold exceedances. We have proposed a novel hierarchical model for this kind of data leading to asymptotic independence in space and time. Our approach is based on representing a generalized Pareto distribution as a Gamma mixture of an exponential distribution, enabling us to keep marginal distributions which are coherent with univariate extreme value theory. The key idea is to use a kernel convolution of a space-time Gamma random process based on influence zones defined as cylinders with an ellipsoidal basis to generate anisotropic spatio-temporal dependence in exceedances. Statistical inference is based on a composite likelihood for the observed censored excesses. The practical usefulness of our model is illustrated on the previously mentionned hourly precipitation data set from a region in Southern France. This work has been presented in two invited talks in 2018 ([16], [17]) and is under revision in JASA [27].

7.3.2. Integrating heterogeneous data

In the framework of the Cart'Eaux project, a stochastic algorithm has been set up to provide a set of probable wastewater networks geometries, obtained from manhole covers positions and cost functions based on general guidelines for such networks construction. The methodology and results are presented in a publication submitted to Computers, Environment and Urban Systems Journal.

Meanwhile, the MeDo project led by N. CHAHINIAN in collaboration with linguists, aims at identifying thematic entities related to wastewater networks in automatically collected documents on the web. This project has been presented in [8].

A PhD thesis has just been funded by ANRT and Berger-Levrault company concerning the fusion of the heterogeneous and uncertain data collected. This PhD (Yassine BEL-GHADDAR) will starts at the beginning of 2019 and will be co-directed by Carole Delenne and Ahlame BEGDOURI from the LSIA laboratory of FST Fes (Maroc).

MAGIQUE-3D Project-Team

6. New Results

6.1. Seismic Imaging and Inverse Problems

6.1.1. Shape-reconstruction and parameter identification of an elastic object immersed in a fluid

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Julien Diaz.

We have developed a procedure to reconstruct the shape and material parameters of an elastic obstacle immersed in a fluid medium from some external measurements given by the so called far-field pattern. It is a nonlinear and ill-posed problem which is solved by applying a Newton-like iterative method involving the Fréchet derivatives of the scattered field. These derivatives express the sensitivity of the scattered field with respect to the parameters of interest. They are defined as the solution of boundary value problems which differ from the direct one only at the right-hand sides level. We have been able to establish the well-posedness of each problem in the case of a regular obstacle and it would be interesting in the near future to extend those results to the case of scatterers with polygonal boundaries. It requires to work with less regular Sobolev spaces for which the definition of traces is not obvious. We have also provided an analytical representation of the Fréchet derivatives in the case of a circle.

Next, we have introduced a series of numerical experiments that have been performed by applying two algorithms which propose two strategies of full reconstruction regarding the material parameters are retrieved simultaneously with the shape or not. It turns out that both work similarly delivering the same level of accuracy but the simultaneous reconstruction requires less iterations. We have thus opted for retrieving all the parameters simultaneously. Since realistic configurations include noisy data, we have performed some simulations for the reconstruction of the shape along with the Lamé coefficients for different noise levels. Other interesting experiments have been carried out using a multistage procedure where the parameters of interest are the density of the solid interior, the shape of the obstacle and its position. We have considered the case of Limited Aperture Data in back-scattering configurations, using multiple incident plane waves, mimicing a physical disposal of non-destructive testing.

We extended the solution methodology to the case of anisotropic media. Since the impact of some of the anisotropic parameters on the FFP is even weaker than the Lamé coefficients, the reconstruction of these parameters together with the shape parameters requires several frequencies and carefully adapted regularization parameters. It is in particular difficult to retrieve the Thomsen parameters ϵ and δ because their reconstruction requires to have an accurate adjustment on the rest of material and shape parameters. The recovery process is thus computationally intensive and some efforts should be done in the near future to decrease the computational costs. We were able to recover all the anisotropic parameters when the shape were assumed to be known. However, when trying to recover both shape and material parameters, we could only recover the shape and some of the physical parameters (namely the three most important ones : the density and the two velocities V_p and V_s).

These results have been obtained in collaboration with Rabia Djellouli (California State University at Northridge, USA) and are presented in Izar Azpiroz Ph.D thesis [1]

6.1.2. Time-harmonic seismic inverse problem with dual-sensors data

Participants: Hélène Barucq, Florian Faucher.

We study the inverse problem for the time-harmonic acoustic wave equation. The seismic context implies restrictive set of measurements: it consists of reflection data (resulting from an artificial source) acquired from the near surface area only. The inverse problem aims at recovering the subsurface medium parameters and we use the Full Waveform Inversion (FWI) method, which defines an iterative minimization algorithm of the difference between the measurement and simulation.

We investigate the use of new devices that have been introduced in the acoustic setting. They are able to capture both the pressure field and the vertical velocity of the waves and are called *dual-sensors*. For solving the inverse problem of interest, we define a new cost function, adapted to these two-components data. We first note that the stability of the problem can be shown to be Lipschitz, assuming the parameters to be piecewise linear.

The usefulness of the cost function is to allow a separation between the observational and numerical sources. Therefore, the numerical sources do not have to coincide with the observational ones, offering new possibilities to create adapted computational acquisitions, and possibilities to reduce the numerical burden. We illustrate our approach with three-dimensional medium reconstructions, where we start with minimal information on the target models.
This work is a collaboration with Giovanni Alessandrini (Università di Trieste), Maarten V. de Hoop (Rice University), Romina Gaburro (University of Limerick) and Eva Sincich (Università di Trieste). It has been presented in the GDR-Meca Wave conference [31].

6.1.3. Stability and convergence analysis for seismic depth imaging using Full Waveform Inversion

Participants: Hélène Barucq, Florian Faucher.

We study the convergence of the inverse problem associated with the time-harmonic wave equations. In the context of seismic, the inverse problem uses reflection data which can only be obtained from the near surface area. We consider the propagation of waves in a domain Ω and the forward problem is defined from the displacement vector field u, solution to

$$-\rho\omega^2 u - \nabla \cdot \sigma = g, \quad \text{in } \Omega, \tag{1}$$

where g stands for the source, ρ is the density and σ the stress tensor. The inverse problem aims at the recovery of the medium parameters (contained in the stress tensor) and can be solved with an iterative minimization algorithm: this is the Full Waveform Inversion (FWI) method. We study the convergence of the minimization by introducing the framework of *Finite Curvature/Limited Deflection* (FC/LD) problems. The idea is to obtain the FC/LD properties by restricting the model space to guarantee *strictly quasiconvex* attainable set. It allows us to numerically estimate the size of the basin of attraction depending on characteristics of the inverse problem such as the frequency or the geometry of the target. In particular, it allows a quantitative comprehension of frequency progression during the iterative scheme, which is an aspect that appeared mostly intuitive. It also allows a comparison of methods from a convergence point of view. This analysis is to relate with stability estimates in order to provide a consistent scheme where frequency progression is justified from the quantitative estimates. Eventually, we illustrate our approach with elastic medium reconstructions, starting from minimal information on the initial models; this also serves to illustrate the numerical requirement of the large scale optimization seismic experiments.
This work is a collaboration with Guy Chavent (Inria Rocquencourt) and Henri Calandra (TOTAL). The results have been presented in the conference "Reconstruction Methods for Inverse Problems" [15].

6.1.4. Quantitative localization of small obstacles with single-layer potential fast solvers

Participants: Hélène Barucq, Florian Faucher, Ha Howard Faucher.

In this work, we numerically study the inverse problem of locating small circular obstacles in a homogeneous medium using noisy backscattered data collected at several frequencies. The main novelty of our work is the implementation of a single-layer potential based fast solver (called FSSL) in a Full-Waveform inversion procedure, to give high quality reconstruction with low-time cost. The efficiency of FSSL was studied in our previous works. We show reconstruction results with up to 12 obstacles in structured or random configurations with several initial guesses, all allowed to be far and different in nature from the target. This last assumption is not expected in results using nonlinear optimization schemes in general. For results with 6 obstacles, we also investigate several optimization methods, comparing between nonlinear gradient descent and quasi-Newton, as well as their convergence with different line search algorithms.

The work is published in Journal of Computational Physics [9]. This work has been presented at GDR-Meca Wave conference in Fréjus *cf.* [21].

6.1.5. Time Domain Full Waveform Inversion Adjoint Studies

Participants: Hélène Barucq, Julien Diaz, Pierre Jacquet.

Full Waveform Inversion (FWI) allows retrieving the physical parameters (e.g. the velocity, the density) from an iterative procedure underlying a global optimization technique. The recovering of the medium corresponds to the minimum of a cost function quantifying the difference between experimental and numerical data. In this study we have considered the adjoint state method to compute the gradient of this cost function. At each iteration the parameters are updated with the solution of an adjoint equation which can be defined either as the adjoint of the continuous equation or the discrete problem. Some studies have addressed the question of establishing what the best strategy is. The answer is still unclear and turns out to be strongly dependent on the problem under study.

The purpose of this study was to investigate several computations of the adjoint state as a preamble of a FWI method applied to the time-dependent acoustic wave approximated in a Discontinuous Galerkin framework involving Bernstein elements. We have considered different time schemes to feature the inherited properties of the computed adjoint state. By comparing the different discrete adjoint operators both from a mathematical and numerical point of view, we aim at defining the best option for computing the adjoint state with accuracy at least cost.

This work is a collaboration with Henri Calandra (TOTAL). It was presented at Total MATHIAS conference in Paris [28].

6.1.6. Seismic imaging of remote targets buried in an unknown media

Participant: Yder Masson.

Box Tomography:first application to the imaging of upper-mantle shear velocity and radial anisotropy structure beneath the North American continent: The EarthScope Transpotable Array (TA) deployment provides dense array coverage through- out the continental United States and with it, the opportunity for high-resolution 3-D seismic velocity imaging of the stable part of the North American (NA) upper mantle. Building upon our previous long-period waveform tomographic modeling, we present a higher resolution 3-D isotropic and radially anisotropic shear wave velocity model of the NA lithosphere and asthenosphere. The model is constructed using a combination of teleseismic and regional waveforms down to 40 s period and wavefield computations are performed using the spectral element method both for regional and teleseismic data. Our study is the first tomographic ap- plication of ‘Box Tomography’, which allows us to include teleseismic events in our inversion, while computing the teleseismic wavefield only once, thus significantly reducing the numerical computational cost of several iterations of the regional inversion. We confirm the presence of high-velocity roots beneath the Archean part of the continent, reaching 200–250 km in some areas, however the thickness of these roots is not everywhere correlated to the crustal age of the corresponding cratonic province. In particular, the lithosphere is thick (250 km) in the western part of the Superior craton, while it is much thinner (150 km) in its eastern part. This may be related to a thermomechanical erosion of the cratonic root due to the passage of the NA plate over the Great Meteor hotspot during the opening of the Atlantic ocean 200–110 Ma. Below the lithosphere, an upper-mantle low-velocity zone (LVZ) is present everywhere under the NA continent, even under the thickest parts of the craton, although it is less developed there. The depth of the minimum in shear velocity has strong lateral variations, whereas the bottom of the LVZ is everywhere relatively flat around 270–300 km depth, with minor undulations of maximum 30 km that show upwarping under the thickest lithosphere and downwarping under tectonic regions, likely reflecting residual temperature anomalies. The radial anisotropy structure is less well resolved, but shows distinct signatures in highly deformed regions of the lithosphere.

This is the first application to a real case study of a novel imaging method called "Box Tomography". These results were obtained through collaborations with Barbara Romanowicz (Berkeley Seimological Laboratory, UC Berkeley; Collège de France) and Pierre Clouzet (Institut de Physique du Globe de Paris). The results have been published in the Geophysical Journal International [11].

Additional developments are conducted in collaboration with Sevan Adourian and Barbara Romanowicz at the Berkeley Seismological Laboratory, UC Berkeley, in particular, to efficiently account for receivers located outside the imaged region. These new results have been presented in different international conferences [23], [19].

To strengthen existing collaborations, a proposal has been submitted to the France-Berkeley Fund (13000$ for travelling and living expenses). We propose a joint effort to further develop and apply a novel seismic tomographic approach, Box-Tomography, to image and characterize small scale structures of interest in the deep Earth, such as the roots of mantle plumes, ultra-low velocity zones, or the edges of large low shear velocity provinces. Our objective is to forge a long-term collaboration between applied mathematicians at Magique3D developing wave propagation modeling methods and the seismologists at the Berkeley Seismological Laboratory (UC Berkeley) using these methods to investigate the Earth's internal structure.

6.2. Mathematical modeling of multi-physics involving wave equations

6.2.1. Hybrid space discretization to solve elasto-acoustic coupling

Participants: Hélène Barucq, Julien Diaz, Aurélien Citrain.

Accurate wave propagation simulations require selecting numerical schemes capable of taking features of the medium into account. In case of complex topography, unstructured meshes are the most adapted and in that case, Discontinuous Galerkin Methods (DGM) have demonstrated great performance. Off-shore exploration involves propagation media which can be well represented by hybrid meshes combining unstructured meshes with structured grids that are best for representing homogeneous media like water layers. Then it has been shown that Spectral Element Methods (SEM) deliver very accurate simulations on structured grids with much lower computational costs than DGMs.

We have developed a SEM-DG numerical method for solving time-dependent elasto- acoustic wave problems. We consider the first-order coupled formulation for which we propose a SEM-DG formulation which turns out to be stable. In the 2D case, the coupling is quite straightforward due to the natural way of mixing triangles with quadrangles. 3D coupling is much more difficult and the interface between tetrahedra and hexahedra deserves a particular attention.

These results have been obtained in collaboration with Henri Calandra(TOTAL) and Christian Gout (INSA Rouen) and have been presented at the Fifth International congress on multiphysics, multiscale and optimization problems in Bilbao, the 13th World Congress on Computational Mecanics in New-York and MATHIAS conference in Paris [16], [17], [24].

6.2.2. Signal and noise in helioseismic holography

Participant: Hélène Barucq.

Helioseismic holography is an imaging technique used to study heterogeneities and flows in the solar interior from observations of solar oscillations at the surface. Holograms contain noise due to the stochastic nature of solar oscillations. Aims. We provide a theoretical framework for modeling signal and noise in Porter-Bojarski helioseismic holography. Methods. The wave equation may be recast into a Helmholtz-like equation, so as to connect with the acoustics literature and define the holography Green's function in a meaningful way. Sources of wave excitation are assumed to be stationary, horizontally homogeneous, and spatially uncorrelated. Using the first Born approximation we calculate holograms in the presence of perturbations in sound-speed, density, flows, and source covariance, as well as the noise level as a function of position. This work is a direct extension of the methods used in time-distance helioseismology to model signal and noise. Results. To illustrate the theory, we compute the hologram intensity numerically for a buried sound-speed perturbation at different depths in the solar interior. The reference Green's function is obtained for a spherically-symmetric solar model using a finite-element solver in the frequency domain. Below the pupil area on the surface, we find that the spatial resolution of the hologram intensity is very close to half the local wavelength. For a sound-speed perturbation of size comparable to the local spatial resolution, the signalto-noise ratio is approximately constant with depth. Averaging the hologram intensity over a number N of frequencies above 3 mHz increases

the signal-to-noise ratio by a factor nearly equal to the square root of N. This may not be the case at lower frequencies, where large variations in the holographic signal are due to the individual contributions of the long-lived modes of oscillation. This work has been done in collaboration with Laurent Gizon, Damien Fournier, Dan Yang and Aaron C. Birch of the Max-Planck-Institut für Sonnensystemforschung at Göttingen (Germany) and published in Astronomy and Astrophysics [14]

6.2.3. *Sensitivity kernels for time-distance helioseismology. Efficient computation for spherically symmetric solar models*

Participant: Hélène Barucq.

The interpretation of helioseismic measurements, such as wave travel-time, is based on the computation of kernels that give the sensitivity of the measurements to localized changes in the solar interior. These kernels are computed using the ray or the Born approximation. The Born approximation is preferable as it takes finite-wavelength effects into account, although it can be computationally expensive. Aims.We propose a fast algorithm to compute travel-time sensitivity kernels under the assumption that the background solar medium is spherically symmetric. Methods. Kernels are typically expressed as products of Green's functions that depend upon depth, latitude, and longitude. Here, we compute the spherical harmonic decomposition of the kernels and show that the integrals in latitude and longitude can be performed analytically. In particular, the integrals of the product of three associated Legendre polynomials can be computed. Results. The computations are fast and accurate and only require the knowledge of the Green's function where the source is at the pole. The computation time is reduced by two orders of magnitude compared to other recent computational frameworks. Conclusions. This new method allows flexible and computationally efficient calculations of a large number of kernels, required in addressing key helioseismic problems. For example, the computation of all the kernels required for meridional flow inversion takes less than two hours on 100 cores. This work has been done in collaboration with Damien Fournier, Chris S. Hanson and Laurent Gizon of the Max-Planck-Institut für Sonnensystemforschung at Göttingen (Germany) and published in Astronomy and Astrophysics [13]

6.2.4. *Characterization of partial derivatives with respect to material parameters in a fluid-solid interaction problem.*

Participants: Izar Azpiroz Iragorri, Hélène Barucq, Ha Howard Faucher.

For a fluid-solid interaction problem with Lipschitz interface, we investigate the partial Fréchet differentiability of the solutions and the approximate far-field-pattern with respect to solid material parameters. Differentiability is shown in standard Sobolev framework, and the derivatives are characterized as solutions to inhomogeneous fluid-solid transmission problems. To validate the accuracy of the characterization, we compare analytical values with numerical ones given by Interior Penalty Discontinuous Galerkin (IPDG) in a setting with circular obstacles. Our comparisons also show that IPDG gives results with high precision and incurs almost no effect of discretization error accumulation. This work has been done in colllaboration with Rabia Djellouli (california State University at Northridge, USA). It has been published in [3].

6.2.5. *Asymptotic modeling of multiple electromagnetic scattering problems by small obstacles*

Participants: Justine Labat, Victor Péron, Sébastien Tordeux.

The detection of small conductive heterogeneities in three dimensional domains by non-destructive electromagnetic imaging is a real challenge. Basic finite element-based methods are very expensive in terms of computation time and memory burden, since they involve a huge number of degrees of freedom when the obstacles are very small compared to the testing wavelength. Using the matched asymptotic expansions method, we have developed a meshless reduced model, which consists of replacing the scatterers by equivalent point sources. This method has been numerically implemented in Matlab and its accuracy validated with analytical solutions in spherical geometries. The details of the results are given in [35] and were presented at the fifth International Congress on Multiphysics, Multiscale and Optimization Problems in Bilbao [18] and at ECCOMAS conferences in Glasgow [22]. Following the Born and Foldy-Lax models, we can extend the results for one obstacle to the multiple scattering problem, thus provide meshless methods in this case. Numerical simulations with thousands of small scatterers, up to 10000, were presented at the seminar of RWTH Aachen University [40].

6.2.6. Discontinuous Galerkin Trefftz type method for solving the Maxwell equations

Participant: Sébastien Tordeux.

Trefftz type methods have been developed in Magique 3D to solve Helmholtz equation. Theese methods reduce the numerical dispersion and the condition number of the linear system. This work aims in pursuing this development for electromagnetic scattering. We have adapted and tested the method for an academical 2D configuration. This work has been achieved in the context of the Master trainee of Hakon Fure in collaboration with Sébastien Pernet of ONERA Toulouse.

6.2.7. Comparison between Galbrun and linearized Euler models in the context of helioseismology

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé, Nathan Rouxelin.

Helioseismology aims to probe the Sun's internal structure thanks to surface observations and our knowledge of acoustic wave propagation. In this work we focus on modeling and simulating the propagation of waves below the surface of the Sun.

In the first part, we establish the equations for acoustic wave propagation by linearizing the Euler equations describing the fluid flow. We then compare two linearization processes based on the eulerian and lagrangian description of fluid dynamics.

In the second part, we solve those equations in time-harmonic domain using high order Discontinuous Galerkin methods. Those numerical methods seem to lack consistency and stability when applied to our problems. Specifically, we notice the presence of spurious modes in our numerical solutions.

To fully understand those results further investigations are needed. In particular, two questions seem to stand out : Is the acoustic wave propagation problem in time-harmonic domain well posed for a recirculating background flow ? Is this approach valid ? Can we really assume that the solar plasma solves the Euler equations ?

6.2.8. Asymptotic Models for the Electric Potential across a Highly Conductive Casing

Participant: Victor Péron.

We analyze a configuration that involves a steel-cased borehole, where the casing that covers the borehole is considered as a highly conductive thin layer. We develop an asymptotic method for deriving reduced problems capable of efficiently dealing with the numerical difficulties caused by the casing when applying traditional numerical methods. We derive several reduced models by employing two different approaches, each of them leading to different classes of models. We prove stability and convergence results for these models. The theoretical orders of convergence are supported by numerical results obtained with the finite element method. These results have been obtained with D. Pardo (UPV/EHU, BCAM, Ikerbasque) and Aralar Erdozain. It was published in Computers and Mathematics with Applications [12].

6.2.9. Boundary Element Method for 3D Conductive Thin Layer in Eddy Current Problems

Participant: Victor Péron.

Thin conducting sheets are used in many electric and electronic devices. Solving numerically the eddy current problems in presence of these thin conductive sheets requires a very fine mesh which leads to a large system of equations, and becoming more problematic in case of high frequencies. In this work we show the numerical pertinence of asymptotic models for 3D eddy current problems with a conductive thin layer of small thickness based on the replacement of the thin layer by its mid-surface with impedance transmission conditions that satisfy the shielding purpose, and by using an efficient discretization with the Boundary Element Method in order to reduce the computational cost. These results have been obtained in collaboration with M. Issa, R. Perrussel and J-R. Poirier (LAPLACE, CNRS/INPT/UPS, Univ. de Toulouse) and O. Chadebec (G2Elab, CNRS/INPG/UJF, Institut Polytechnique de Grenoble). This work has been accepted for publication in COMPEL - The international journal for computation and mathematics in electrical and electronic engineering. This work has been presented in the symposium IABEM 2018.

6.2.10. Model-based digital pianos: from physics to sound synthesis

Participant: Juliette Chabassier.

A review article has been published in IEEE Signal Processing Magazine on model-based digital pianos in collaboration with Balasz Bank [4] .

6.2.11. The virtual workshop : towards versatile optimal design of musical wind instruments for the makers

Participants: Juliette Chabassier, Robin Tournemenne.

Our project aims at proposing optimization solutions for wind instrument making. Our approach is based on a strong interaction with makers and players, aiming at defining interesting criteria to optimize from their point of view. After having quantified those criteria under the form of a cost function and a design parameters space, we wish to implement state-of-the-art numerical methods (finite elements, full waveform inversion, neuronal networks, diverse optimization techniques...) that are versatile (in terms of models, formulations, couplings...) in order to solve the optimization problem. More precisely, we wish to take advantage of the fact that sound waves in musical instruments satisfy the laws of acoustics in pipes (PDE), which gives us access to the full waveform inversion technique, usable in harmonic or temporal regime. The methods that we want to use are attractive because the weekly depend on the chosen criterion, and they are easily adaptable to various physical situations (multimodal decomposition in the pipe, coupling with the embouchure, ...), which can therefore be modified a posteriori. The goal is to proceed iteratively between instrument making and optimal design (the virtual workshop) in order to get close to tone quality related and playability criteria. In 2018 we have implemented a python 3 toolbox named OpenWind that includes the first simulation module. Next modules will be implemented next year. This work has been presented to the Congrès Français d'Acoustique [26].

6.2.12. Collaboration with Augustin Humeau, wind instrument maker

Participants: Juliette Chabassier, Robin Tournemenne.

We have initiated a strong collaboration with Augustin Humeau, bassoon maker in Dordogne, France. The goal is to develop practical tools for instrument design, in the realistic context of an artisanal workshop. Until now, an input impedance measurement setup has been developed in collaboration with Samuel Rodriguez, I2M Univ. Bordeaux. It is based on the use of five microphones and the need of one calibration. It has been specifically adapted to the small entrances of some wind instruments (bassoon, oboes). We have attended the JFIS (journées facture instrumentale et science) in November 2018, Le Mans, where the approach has been presented and demonstrated. Given the great interest showed by other instrument makers attendanting the conference, the future of this collaboration is in discussion and may integrate an Inria startup process.

6.2.13. Optimization of brass wind instruments based on sound simulations

Participant: Robin Tournemenne.

We exploited a new optimization method of the inner shape of brass instruments using sound simulations to derive objective functions. The novelties are the obtention of optimal bores for objective functions representative of the intonation but also of the spectrum of the instrument, and the possibility to include constraints in the optimization problem. A complete physics-based model, taking into account the instrument and the musician's embouchure, is used, in order to simulate sounds' permanent regimes using the harmonic balance technique. The instrument is modeled by its input impedance computed with the transfer matrix method under plane wave propagation and visco-thermal losses. Some embouchure's parameters remain variable during the optimization procedure in order to get the average behavior of the instrument. The design variables are the geometrical dimensions of the resonator. Given the computationally expensive function evaluation and the unavailability of gradients, a surrogate-assisted optimization framework is implemented using the mesh adaptive direct search algorithm (MADS). Two optimization examples of a Bb trumpet's bore (with two and ten design optimization variables) demonstrate the effectiveness of the approach. Results show that solvers can deal flawlessly with high dimensional problems, under constraints, improving significantly the value of the objective functions.

6.2.14. Energy based model and simulation in the time domain of linear acoustic waves in a radiating pipe

Participants: Juliette Chabassier, Robin Tournemenne.

We model in the time domain linear acoustic waves in a radiating pipe without damping. The acoustic equations system in formulated in flow and pressure, which leads to a first order space time equations system. The radiation condition is also written as a first order in time equation, and is parametrized by two real coefficients. Moreover, an auxiliary variable is introduced at the radiating boundary. The choice of this variable is adapted to the considered source type in order to ensure the model stability by energy techniques, under some conditions on the radiating condition. We then propose a stable space time explicit discretization, which ensures the dissipation of a discrete energy. The novelty of the discretization lies, on the one hand, in the variational nature of the space approximation (which leads to arbitrary order finite elements with no required matrix inversion), and on the other hand, on the definition of the auxiliary variable for any acoustic source type (which leads to the decay of a well defined energy). Finally, we quantify the frequential domain of validity of the used radiation condition by comparison with theoretical and experimental models of the literature. This is a collaboration with Morgane Bergot (Université Claude Bernard, Lyon 1). An article has been written and will be submitted soon. This work has been presented to the Congrès Français d'Acoustique [25].

6.2.15. Computation of the entry impedance of a dissipative radiating pipe

Participants: Juliette Chabassier, Robin Tournemenne.

Modeling the entry impedance of wind instruments pipes is essential for sound synthesis or instrument qualification. We study this modeling with the finite elements method in one dimension (FEM1D) and with the more classically used transfer matrix method (TMM). The TMM gives an analytical formula of the entry impedance depending on the bore (intern geometry of the instrument) defined as a concatenation of simple elements (cylinders, cones, etc). The FEM1D gives the entry impedance for any instrument geometry. The main goals of this work are to assess the viability of the FEM1D and to study and analyse the approximations necessary for the TMM in dissipative pipes. First, lossless Webster's equation in one dimension is studied with arbitrary radiation conditions. In this context and for cylinders or cones, the TMM is exact. We verify that the error made with FEM1D for fine enough elements is as small as desired. When we consider viscothermal losses, the TMM does not solve the classical Kirchhoff model because two terms are supposed constant. In order to overcome this model approximation, simple segments, on which are based the TMM, are decomposed into much smaller segments. We show that using the TMM actually amounts to solving a different equation than the original one, on each small segment. The FEM1D does not necessitate any model approximation, and it is possible to show that it solves the dissipative equation with any arbitrarily small error. With this in hand, we can quantify the TMM model approximation error. The methods are compared in terms of accuracy and computational burden. On realistic cases as the case of a trumpet, the FEM show a better efficiency. Moreover, unusual phenomena as a non constant air temperature can easily be tackle with the FEM. An article has been written and will be submitted soon. This work has been presented to the Congrès Français d'Acoustique [27].

6.2.16. Seismic wave propagation in carbonate rocks at the core scale

Participants: Julien Diaz, Florian Faucher, Chengyi Shen.

Reproduction of large-scale seismic exploration at lab-scale with controllable sources is a promising approach that could not only be applied to study small-scale physical properties of the medium, but also contribute to significant progress in wave-propagation understanding and complex media imaging at exploration scale via upscaling methods. We propose to apply a laser-generated seismic source for lab-scale new geophysical experiments. This consists in generating seismic waves in various media by well-calibrated pulsed-laser impacts and measuring precisely the wavefield (displacement) by Laser Doppler Vibrometer. Parallel 2D/3D simulations featuring the Discontinuous Galerkin discretization method with Interior Penalties (IPDG) are done to match the experimental data. The IPDG method is of particular interest when it comes to solve wave propagation problems in highly heterogeneous media, such as the limestone cores that we are studying.

Current seismic data allowed us to retrieve V_p tomography slices. Further more, qualitative/quantitative comparisons between simulations and experimental data validated the experiment protocol and vice-versa the numerical schemes, opening the possibility of performing FWI on these high resolution data.

This work is in collaboration with Clarisse Bordes, Daniel Brito and Deyuan Zhang (LFCR, UPPA) and with Stéphane Garambois (ISTerre). It was presented at conference AGU [42].

6.3. Supercomputing for Helmholtz problems

6.3.1. Numerical libraries for hybrid meshes in a discontinuous Galerkin context

Participants: Hélène Barucq, Aurélien Citrain, Julien Diaz.

Elasticus team code has been designed for triangles and tetrahedra mesh cell types. The first part of this work was dedicated to add quadrangle libraries and then to extend them to hybrid triangles-quadrangles (so in 2D). This implied to work on polynomials to form functions basis for the (discontinuous) finite element method, to finally be able to construct reference matrices (mass, stiffness, ...).

A complementary work has been done on mesh generation. The goal was to encircle an unstructured triangle mesh, obtained by third-party softwares, with a quadrangle mesh layer. At first, we built scripts to generate structured triangle meshes, quadrangle meshes and hybrid meshes (triangles surrounded by quadrangles). In 2018, we have implemented the coupling between Discontinuous Galerkin methods (using the triangles/tetrahedra) and Spectral Element methods (using quadrangles/hexahedra). We have also implemented the PML in the SEM part, and we are now working on the local time-stepping feature.

6.4. Hybrid time discretizations of high-order

6.4.1. Space-Time Discretization of Elasto-Acoustic Wave Equation in Polynomial Trefftz-DG Bases

Participants: Elvira Shishenina, Hélène Barucq, Julien Diaz.

In the context of the strategic action "Depth Imaging Partnership" between Inria and Total we have investigated to the development of an explicit Trefftz-DG formulation for elasto-acoustic problem, solving the global sparse matrix by constructing an approximate inverse obtained from the decomposition of the global matrix into a block-diagonal one. The inversion is then justified under a CFL-type condition. This idea allows for reducing the computational costs but its accuracy is limited to small computational domains. According to the limitations of the method, we have investigated the potential of Tent Pitcher algorithms following the recent works of Gopalakrishnan et al. It consists in constructing a space-time mesh made of patches that can be solved independently under a causality constraint. We have obtained very promising numerical results illustrating the potential of Tent Pitcher in particular when coupled with a Trefftz-DG method involving only surface terms. In this way, the space-time mesh is composed of elements which are 3D objects at most. It is also worth noting that this framework naturally allows for local time-stepping which is a plus to increase the accuracy while decreasing the computational burden. The results of this work have been published in the Applicable Analysis Journal [6], in the book of proceedings for European Conference on Numerical Mathematics and Advanced Applications (ENUMATH 2017) (due date April 27, 2019), and in the PhD thesis [2], as well as presented during the International Conference on Spectral and High-Order Methods (ICOSAHOM 2018, London - UK), the 13$^{\text{th}}$ World Congress on Computational Mechanics (WCCM 2018, New-York - USA), and during the annual seminar on Computational Science Engineering and Data Science organized by TOTAL (MATHIAS 2018, Serris - France).

6.4.2. Performance analysis of local time-stepping schemes for wave propagation

Participants: Julien Diaz, Rose-Cloé Meyer.

The efficiency of numerical simulation of wave propagation is highly dependent of the quality of the mesh. For complex simulations, the size of the cells in the mesh can strongly vary, either because of the geometry or because of the different propagation celerity of the waves. To ensure stability, explicit numerical schemes must match with the CFL conditions of every cells of the mesh. When significant disparities appear in the domain, the time step used on big cells is not optimal, which can cause heavy calculation cost and result in a loss of efficiency. To improve the performance of the programs, local time-stepping methods based on a spatial Discontinuous Galerkin discretization have been implemented. In this work, we compared three local time-stepping methods: a conservative method, a recursive method, and an asynchron method. The two first methods use local time steps that are fractions of the global time step, while the third method can use independent time steps on each cell of the mesh. The accuracy of the solution, the computation cost and the speedup of local-time stepping are presented on cases in two and three dimensions on configurations as fine slot or domains with geometric singularities. The results are presented in Rose-Cloé Meyer Master thesis [41]. This work has been achieved in collaboration with Guillaume Dufour and Xavier Ferrières (Onera)

6.4.3. *Construction and analysis of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations.*

Participants: Juliette Chabassier, Julien Diaz.

We submitted a paper to M2AN [37]. This paper deals with the construction of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations. This family of schemes is obtained by replacing the inversion of a matrix, that comes naturally after using the technique of the Modified Equation on the second order Leap Frog scheme applied to dissipative linear wave equations, by an explicit approximation of its inverse. The series can be truncated at different orders, which leads to several schemes. The stability of the schemes is studied. Numerical results in 1D illustrate the good behavior regarding space/time convergence and the efficiency of the newly derived scheme compared to more classical time discretizations. A loss of accuracy is observed for non smooth profiles of dissipation, and we propose an extension of the method that fixes this issue. Finally, we assess the good performance of the scheme for a realistic dissipation phenomenon in Lorentz's materials. This work has been done in collaboration with Sébastien Imperiale (Inria Project-Team M3DISIM).

6.4.4. *Construction and convergence analysis of conservative second order local time discretisation for wave equations*

Participant: Juliette Chabassier.

In this work we present and analyse a time discretisation strategy for linear wave propagation that aims at using locally in space the most adapted time discretisation among a family of implicit or explicit centered second order schemes. The domain of interest being decomposed into several regions, different time discretisations can be chosen depending on the local properties of the spatial discretisations (mesh size and quality) or the physical parameters (high wave speed, low density). We show that, under some conditions on the time step, the family of time discretisations obtained combined with standard finite elements methods in space ensures a second order space-time convergence. This work has been done in collaboration with Sébastien Imperiale (Inria Project-Team M3DISIM). It has been submitted to Numerische Mathematik.

6.4.5. *High-order locally implicit time schemes for linear ODEs*

Participants: Hélène Barucq, Marc Duruflé, Mamadou N'Diaye.

In this work we have proposed a method that combines optimized explicit schemes and implicit schemes to form locally implicit schemes for linear ODEs, including in particular ODEs coming from the space discretization of wave propagation phenomena. This method can be applied to the following ODE

$$M_h \frac{dU}{dt} = K_h U + F(t)$$

Like in the local time-stepping developed by Grote and co-workers, the computational domain is split into a fine region and a coarse region. The matrix A_h is given as

$$A_h = M_h^{-1} K_h = A_h P + A_h (I - P)$$

where P is the projector on the fine region of the computational domain. Then the proposed locally implicit method is obtained from the combination of the A-stable implicit schemes we have developed in 2016 (Padé schemes or Linear-SDIRK schemes detailed in [8]) on the fine region and explicit schemes with optimal CFL number in the coarse region. The developed method has been used to solve the acoustic wave equation and we have checked the convergence in time of these schemes for order 4, 6 and 8.

This work is a chapter of the thesis defended by Mamadou N'diaye under the joint supervision of Hélène Barucq and Marc Duruflé. In 2018, the implemented method has been parallelized in Montjoie and 3-D numerical results have been obtained. An article is in preparation.

SERENA Project-Team

7. New Results

7.1. Unfitted hybrid-high-order methods

Participants: Alexandre Ern, Guillaume Delay.

Our team contributes actively to the development of hybrid high-order (HHO) methods. Such methods support polyhedral meshes with hanging nodes, but one requirement is that the mesh cells have planar faces. This is difficult when it comes to solving with high accuracy a problem posed on a domain with curved boundaries or a problem involving a curved interface separating two materials with different properties. One key idea to treat these problems is to use an unfitted mesh, so that the curved boundary or the curved interface freely cuts through the mesh cells. This greatly simplifies the meshing process, but at the same time poses the question on how the HHO method can address the approximation of functions that are not smooth within some mesh cells. The major idea in our approach, which is inspired from similar approaches developed in the context of the more classical finite element method, is to double the discrete unknowns attached to the cut mesh faces and to introduce a consistent Nitsche-type formulation to enforce either the boundary condition or the jump conditions across the interface in a weak manner. In this context, we started a collaboration with Erik Burman (University College London) and we elaborated in [20] the numerical analysis of HHO methods in an unfitted context; further analysis for Stokes and Helmholtz equations has started recently within the postdoc of Guillaume Delay and a collaboration on the subject with CEA is on the way.

7.2. An exponential time stepping scheme for the simulation of diffusion processes

Participant: Géraldine Pichot.

We present in [56] a new Monte Carlo algorithm to simulate diffusion processes in presence of discontinuous convective and diffusive terms. The algorithm is based on the knowledge of close form analytic expressions of the resolvents of the diffusion processes which are usually easier to obtain than close form analytic expressions of the density. In the particular case of diffusion processes with piecewise constant coefficients, known as Skew Diffusions, such close form expressions for the resolvent are available. Then we apply our algorithm to this particular case and we show that the approximate densities of the particles given by the algorithm replicate well the particularities of the true densities (discontinuities, bimodality, ...) Besides, numerical experiments show a quick convergence.

7.3. Localization of dual and distance norms

Participants: Martin Vohralík, Patrick Ciarlet Jr., Jan Blechta, Josef Málek.

Dual norms like the dual norm of the residual and the distance norm to the Sobolev space H_0^1 seem to be fundamentally global over the entire computational domain. In [23], together with P. Ciarlet, we prove, in extension of some older results, that they are both equivalent to the Hilbertian sums of their localizations over patches of elements. Together with J. Blechta and J. Málek, we extend in [45] this result from the space H_0^1 with Hilbertian structure to the Sobolev space $W_0^{1,p}$, with the exponent p bigger than or equal to one, and to an arbitrary bounded linear functional on $W_0^{1,p}$.

7.4. Adaptivity with guaranteed error contraction

Participants: Martin Vohralík, Alexandre Ern, Patrik Daniel, Iain Smears.

In [26], we conceive novel adaptive refinement strategies which automatically decide between mesh refinement and polynomial degree increase. We numerically observe that the error decreases exponentially as a function of the number of degrees of freedom, for smooth as well as for singular numerical solutions. The salient feature of our approach is, however, that we ensure that the error on the next hp-refinement step will be reduced at least by a factor that is given. We then extend in [53] this result to the case where the underlying algebraic solver is inexact. To the best of our knowledge, these results, obtained in the framework of the Ph.D. thesis of Patrik Daniel, is the first ever where such an error contraction bound is computable and guaranteed. Numerically, its precision turns out to be very high (overestimation by a factor very close to the optimal value of one). It immediately implies convergence of the adaptive method, and we would like to use it in the near future for optimality proofs.

STEEP Project-Team

6. New Results

6.1. Calibration of the Tranus Land Use Module: Optimisation-Based Algorithms, their Validation, and Parameter Selection by Statistical Model Selection

Instantiating land use and transport integrated models (LUTI modelling) is a complicated task, requiring substantial data collection, parameter estimation and expert analysis. In this work, we present a partial effort towards the automation of the calibration of Tranus, one of the most popular LUTI models. First, we give a detailed mathematical description of the activity module and the usual calibration approach. Secondly, we reformulate the estimation of the endogenous parameters called shadow prices as an optimisation problem. We also propose an optimisation algorithm for the calibration of the substitution submodel, setting a base for future fully integrated calibration. We analyse the case of transportable and non-transportable economic sectors and propose a detailed mathematical scheme for each case. We also discuss how to validate calibration results and propose to use synthetic data generated from real world problems in order to assess convergence properties and accuracy of calibration methods. Results of this methodology are presented for realistic scenarios. Finally, we propose a model selection scheme to reduce the number of shadow prices that need to be calibrated, with the aim of reducing the risk of overfitting to data. This work is published in [2].

6.2. Convolutional neural networks for disaggregated population mapping using open data

High resolution population count data are vital for numerous applications such as urban planning, transportation model calibration, and population growth impact measurements, among others. In this work, we present and evaluate an end-to-end framework for computing disaggregated population mapping employing convolutional neural networks (CNNs). Using urban data extracted from the OpenStreetMap database, a set of urban features are generated which are used to guide population density estimates at a higher resolution. A population density grid at a 200 by 200 meter spatial resolution is estimated, using as input gridded population data of 1 by 1 kilometer. Our approach relies solely on open data with a wide geographical coverage, ensuring replicability and potential applicability to a great number of cities in the world. Fine-grained gridded population data is used for 15 French cities in order to train and validate our model. A stand-alone city is kept out for the validation procedure. The results demonstrate that the neural network approach using massive OpenStreetMap data outperforms other approaches proposed in related works. This work is published in [5].

6.3. Uncertainties of Domestic Road Freight Statistics: Insights for Regional Material Flow Studies

Freight statistics are at the core of many studies in the field of industrial ecology because they depict the physical inter-dependencies of territories and allow to link worldwide productions and consumptions. Recent studies have been increasingly focusing on subnational scales, often relying on domestic freight data. In this perspective, this article analyses the uncertainties of the French domestic road freight survey, road being by far the most common mode of transport in the country. Based on a statistical analysis of the survey, we propose a model to estimate the uncertainty of any given domestic road transport flow. We also assess uncertainty reduction when averaging the flows over several years, and obtain for instance a 30% reduction for a 3-year average. We then study the impact of the uncertainties on regional material flow studies such as the Economy-Wide Material Flow Analysis of the Bourgogne region. Overall the case studies advocate for a systematic assessment of freight uncertainties, as neither the disaggregation level nor the quantities traded are good enough predictors. This justifies the need for an easy-to-implement estimation model. Finally, basic comparison with the German and Swedish surveys tend to indicate that the main conclusions presented in this article are likely to be valid in other European countries. This work is published in [3].

6.4. A method for downscaling open population data

To extend our ongoing work on urban sprawl indicators (see above), we have developed a method to perform dissagregated population estimations at building level using open data. Our goal is to estimate the number of people living at the fine level of individual households by using open urban data and coarse-scaled population data. First, a fine scale description of residential land use per building is built using OpenStreetMap. Then, using coarse-scale gridded population data, we perform the down-scaling for each household given their containing area for residential usage. We rely solely on open data in order to ensure replicability, and to be able to apply our method to any city in the world, as long as sufficient data exists. The evaluation is carried out using fine-grained census block data for cities in France as ground-truth.

This work is published in [6] and the associated software implementation is made available as open source code at https://github.com/lgervasoni/urbansprawl.

6.5. Modelling the relationships between urban morphology, pollutant generation and concentration in the air using PLS path modelling

We have simultaneously modelled the factors that contribute to shaping the urban environment in terms of population density and activities and the level of land use mix on the one hand, and the mechanisms through which this urban morphology is linked to the emission of pollutants and their concentration in the air in the municipalities of the Auvergne-Rhône-Alpes region. To do this, we used the PLS path modelling approach, which is a method of estimating structural equations to model the relationships between latent variables obtained by extracting the information contained in the multidimensional data used to measure them. This work was carried out as part of Diop Samba's internship [8].

6.6. Implementation of the World3 model in Python for parametric exploration

The 'World3' model is a digital tool for simulating long-term interactions between population, industrial growth, food production and the boundaries of terrestrial ecosystems. This model was developed in the 1970s. We have ported this model to a modern infrastructure (Python3 + related libraries), in order to be able to apply parameter learning, data analysis and sensitivity study techniques to it. This work was carried out as part of the internship of Aina Rasoldier.

TONUS Team

7. New Results

7.1. Palindromic BGK methods

Since two years we work on implicit relaxation methods to solve hyperbolic PDE without CFL and without matrices to invert. The Palindromic BGK method allows to approximate a hyperbolic system by a larger set of transport equations coupled by a nonlinear source term which relaxes the variables on an equilibrium. Using a splitting scheme, we can solve these transport equations in parallel and solve the local relaxation in a second step. The high-order extension is obtained by a symmetric modified Strang splitting and composition methods.

7.1.1. Boundary condition for Palindromic BGK method

Participants: Florence Drui, Emmanuel Franck, Philippe Helluy, Laurent Navoret.
One of the drawbacks to the Palindromic BGK model is the treatment of the boundary conditions. Indeed the BGK scheme admits more variables than the original one and the boundary conditions for these additional variables are not defined. The classical choice is to impose the equilibrium at the boundary. In this case we obtain instabilities and only the first order convergence. After an analysis of the symmetric modified Strang splitting method, we have identified the dynamic for the non-physical variables and proposed boundary conditions compatible with this dynamic. We obtain stable and second order boundary conditions.

7.1.2. Palindromic BGK scheme for Low-Mach models

Participants: Clémentine Courtès (IRMA), Emmanuel Franck, Philippe Helluy, Laurent Navoret.
Another drawback of the method is the application for "two-scale" problems like Low-Mach flows. Indeed, in this case the BGK representation used generate an large error on the slow scale which is homogeneous to the fast scale. Consequently the slow scale is not well resolved. This problem comes from the fact that the BGK approximation uses a linearization with a constant fast scale to approximate all the systems. We have proposed a new method where we also introduce a slow scale in the BGK approximation. Using this, we obtain accurate results for the Euler equation in the low-Mach regime in 1D. The method gives interesting results also for other applications. In the future we must extend the method in 2D.

7.1.3. Palindromic BGK scheme for diffusion models

Participants: Laura Mendoza, Emmanuel Franck, Laurent Navoret.
In MHD simulations for ITER, we must also discretize with an implicit scheme the anisotropic diffusion. Firstly, we have proposed to extend the previous Palindromic BGK method to the parabolic problems. For that we must use a different Palindromic BGK model with specific parameters. We obtain a second order scheme without CFL for the Heat equation in 1D and 2D. In the future we will consider the high-order schemes and the extension to the anisotropic case.

7.1.4. Semi-Lagrangian on complex geometries for Palindromic BGK scheme

Participants: Laura Mendoza, Emmanuel Franck, Philippe Helluy.
To apply the Palindromic BGK method we must have an advection solver without CFL. In the code Slappy we propose a 3D high-order Semi-Lagrangian solver able to treat blocks-structured meshes with overlapping and non-conformity. This allows to treat complex geometries easily. The solver is written in PYOpenCL and can be used on GPU. In the code the relaxation step is also implemented, which allows to use the Palindromic BGK method on some PDE (Euler, Diffusion etc).

7.1.5. Lattice Boltzmann scheme with PyOpenCL

Participants: Florence Drui, Emmanuel Franck, Philippe Helluy.

In the same idea, another code has been developed to treat hyperbolic systems with the BGK approach. In this case the transport is exact and consequently the method is equivalent to the Lattice Boltzmann scheme. The parallel part is similar and also based on PyOpenCL. This version is less accurate than the previous code, can be used only in Cartesian grids but is more stable and can run more complex problems. The main result is the simulation of 2D resistive MHD instabilities which have the same structure than Tokamak instabilities.

7.2. Numerical methods for Euler/MHD models

7.2.1. Splitting scheme in JOREK code

Participants: Emmanuel Franck
The Jorek code is the main European code for the simulation of Tokamak instabilities. The inversion of the full matrix is based on a Block Jacobi preconditioning which is not efficient in some cases and very greedy in memory. To solve the problem we investigate on splitting scheme which will allow to solve some simple subsystems separately. The splitting scheme have been tested on the first MHD model on JOREK in the quasi-linear case. In this regime the splitting gives good results since the accuracy is close to the original full implicit solver. The nonlinear case is currently studied.

7.2.2. Compatible finite element for MHD

Participants: Emmanuel Franck, Eric Sonnendruecker (IPP), Mustaga Gaja (iPP)
The works on the compatible finite elements for MHD is continued. This method allows to preserve the energy balance or the divergence free constrains with high-order finite element on complex geometries. This method is coupled with a splitting between the different physical parts and a nonlinear solver. The method gives expected results for Maxwell and Acoustic and also gives good results for the nonlinear acoustic part of the MHD model. The magnetic and convective parts of the MHD model are currently studied.

7.2.3. Semi implicit for relaxation model in low-Mach regime

Participants: Emmanuel Franck, Laurent Navoret.
To apply the previous method, we must invert a nonlinear problem. A parabolization method allows to reduce the dimension of the implicit problem. However the problem is still nonlinear and ill-conditioned for strong gradient of the physical quantities. To avoid this, we propose a new relaxation method for the Euler equations (to begin) which allows to linearize the acoustic part preserving the low-Mach limit (which is the relevant regime for our application). This relaxation method allows to obtain a well-conditioned and linear implicit part. The method is validated in 1D/2D in a finite volumes context and will be extended to the high-order scheme and MHD model.

7.3. Eulerian method for Vlasov equation

7.3.1. Recurrence phenomenon for finite element grid based Vlasov solver

Participants: Michel Mehrenberger, Laurent Navoret, Nhung Pham (IRMA)

In this work, we focus on one difficulty arising in the numerical simulation of the Vlasov-Poisson system: when using a regular grid-based solver with periodic boundary conditions, perturbations present at the initial time artificially reappear at a later time. For regular finite-element mesh in velocity, we show that this recurrence time is actually linked to the spectral accuracy of the velocity quadrature when computing the charge density. In particular, choosing trigonometric quadrature weights optimally defers in time the occurrence of the recurrence phenomenon. Numerical results using both the Semi-Lagrangian Discontinuous Galerkin and the Finite Element / Semi-Lagrangian methods have been carried out and confirm the analysis.

7.3.2. Numerical scheme for sheath equilibria

Participants: Mehdi Badsi (Nantes University), Michel Mehrenberger, Laurent Navoret

We are interested in developing a numerical method for capturing stationary sheaths that a plasma forms in contact with a metallic wall. This work is based on a bi-species (ion/electron) Vlasov-Ampère model proposed in [18]. The main question addressed in this work is to know if classical numerical schemes can preserve stationary solutions with boundary conditions, since these solutions are not a priori conserved at the discrete level. In the context of high-order semi-Lagrangian method, due to their large stencil, interpolation near the boundary of the domain also requires a specific treatment. As expected, we numerically observe that the preservation of the equilibria is very sensitive to the prescribed boundary conditions and high order schemes are mandatory to maintain the preservation of the energy in large times.

7.3.3. Realistic geometry for Gysela

Participants: N. Bouzat, C. Bressan, V. Grandgirard, G. Latu, M. Mehrenberger

In magnetically confined plasmas used in Tokamak, turbulence is responsible for specific transport that limits the performance of this kind of reactors. Gyrokinetic simulations are able to capture ion and electron turbulence that give rise to heat losses, but also require state-of-the-art HPC techniques to handle computation costs. Such simulations are a major tool to establish good operating regime in Tokamak such as ITER, which is currently being built. Some of the key issues to address more realistic gyrokinetic simulations are: efficient and robust numerical schemes, accurate geometric description, good parallelization algorithms. The framework of this work is the Semi-Lagrangian setting for solving the gyrokinetic Vlasov equation and the Gysela code. In this paper, a new variant for the interpolation method is proposed that can handle the mesh singularity in the poloidal plane at r = 0 (polar system is used for the moment in Gysela). A non-uniform meshing of the poloidal plane is proposed instead of uniform one in order to save memory and computations. The interpolation method, the gyroaverage operator, and the Poisson solver are revised in order to cope with non-uniform meshes. A mapping that establishes a bijection from polar coordinates to more realistic plasma shape is used to improve realism. Convergence studies are provided to establish the validity and robustness of our new approach.

7.3.4. Parallel computing for kinetic solvers

Participants: Ksander Ejjaaouani, Olivier Aumage, Julien Bigot, Michel Mehrenberger

Existing programming models tend to tightly interleave algorithms and optimizations in HPC simulation codes. This requires scientists to become experts in both the simulated domain and the optimization process and makes the code difficult to maintain and port to new architectures. This paper proposes the InKS programming model that decouples these two concerns with distinct languages for each. The simulation algorithm is expressed in the InKS pia language with no concern for machine-specific optimizations. Optimizations are expressed using both a family of dedicated optimizations DSLs (InKS O) and plain C++. InKS O relies on the InKS pia source to assist developers with common optimizations while C++ is used for less common ones. Our evaluation demonstrates the soundness of the approach by using it on synthetic benchmarks and the Vlasov-Poisson equation. It shows that InKS offers separation of concerns at no performance cost.

7.4. PIC method for Vlasov equation

7.4.1. Parallel computing for PIC method

Participants: Y. Barsamian, A. Chargueraud, S. Hirstoaga, M. Mehrenberger

Particle-in-Cell (PIC) codes are widely used for plasma simulations. On recent multi-core hardware, performance of these codes is often limited by memory bandwidth. We describe a multi-core PIC algorithm that achieves close-to-minimal number of memory transfers with the main memory, while at the same time exploiting SIMD instructions for numerical computations and exhibiting a high degree of OpenMP-level parallelism [6]. Our algorithm keeps particles sorted by cell at every time step, and represents particles from a same cell using a linked list of fixed-capacity arrays, called chunks. Chunks support either sequential or atomic insertions, the latter being used to handle fast-moving particles. To validate our code, called Pic-Vert, we consider a 3d electrostatic Landau-damping simulation as well as a 2d3v transverse instability of magnetized electron holes. Performance results on a 24-core Intel Sky-lake hardware confirm the effectiveness of our algorithm, in particular its high throughput and its ability to cope with fast moving particles.

7.4.2. Two species Vlasov solver

Participants: Y. Barsamian, J. Bernier, S. Hirstoaga, M. Mehrenberger

Thanks to a classical first order dispersion analysis, we are able to check the validity of 1Dx1D two-species Vlasov-Poisson simulations. The extension to second order is performed and shown to be relevant for explaining further details. In order to validate multidimensional effects, we propose a 2Dx2D single species test problem that has true 2D effects coming from the sole second order dispersion analysis. Finally, we perform, in the same code, full 2Dx2D nonlinear two-species simulations with mass ratio around 0.01, and consider the mixing of semi-Lagrangian and Particle-in-Cell methods.

7.5. Other works

7.5.1. Tomography

Participants: Laura Mendoza Virtually all magnetic fusion devices resort to tomography diagnostics for a variety of plasma emissions. All those diagnosis have a lot in common: the plasma is transparent to the observed quantity, such that the signal on a detector is derived from a spatial integration of the local emission. Solving the direct problem (i.e. from simulated emissivity to signals) requires modeling the diagnostic geometry and is used for physics code validation or diagnostic design. Solving the inverse problem (i.e. from experimental signals to reconstruct 2D emissivity) is useful for data interpretation and requires not only geometry modeling but also decomposing the unknown emissivity into basis functions and inversion-regularization routines. In this context, a python library, ToFu, solves the direct and inverse problems for synthetic diagnostics. The project objective for the second part of 2018 is to develop and optimize the existing geometry module in ToFu, with a special focus on the ray-tracing algorithms.

7.5.2. Discontinuous Galerkin solver

Participants: Philippe Helluy, Bruno Weber We have implemented and validated new optimizations in our Discontinuous Galerkin (DG) codes CLAC and SCHNAPS. In CLAC, Bruno Weber, our CIFRE PhD in the AxesSim company, has implemented a local time-step method and optimizations in order to run efficiently the OpenCL kernels both on CPU and GPU. This allows to run a huge electromagnetic simulation of a Bluetooth antenna in interaction with a full volumic human model. The simulation was run on the supercomputer Piz Daint (3rd at the "top 500" ranking in 2017). The computing hours were awarded through a PRACE call dedicated to small companies. In SCHNAPS we were able to assess the efficiency of the StarPU runtime for distributing the computational tasks efficiently on hybrid computers.

7.5.3. Finite volume methods for complex hyperbolic system

Participants: Philippe Helluy, Lucie Quibel In the thesis of Lucie Quibel (started in November 2017), we study numerical methods for solving compressible fluids with complex equation of states. The objective is to simulate liquid-vapor flows that occur in nuclear plants. The pressure behavior of the liquid-vapor mixture is very complex and obtained through measurements and tabulated laws. This sometimes prevent the system from being hyperbolic and leads to instabilities. We are trying to construct simpler but realistic laws that preserve the convexity structure and the scheme robustness.

BIOCORE Project-Team

7. New Results

7.1. Mathematical methods and methodological approach to biology

7.1.1. Mathematical analysis of biogical models

7.1.1.1. Model reduction and sensitivity analysis

Participants: Suzanne Touzeau, Jean-Luc Gouzé, 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. A global sensitivity analysis was performed to check the method robustness against parameter uncertainty and variability ([16]). This work was part of Stefano Casagranda's PhD thesis (2017), and is also a collaboration with Bayer (Sophia Antipolis).

7.1.1.2. Estimation and control

Participants: Suzanne Touzeau, Jean-Luc Gouzé.

Parameter identification in compartmental systems. In collaboration with F. Dayan (Startup Exactcure), we work on practical problems of identifiability of parameters in linear pharmacokinetic models. This was the subject of the internship of Jean-Baptiste Excoffier.

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. To fit a within-host immunological model to a large data set of individual viremia profiles, we developed an ABC-like method, less computationally expensive than standard Bayesian fitting procedures. Rather than reproducing individual profiles, we ientified several parameter sets compatible with the data and reflecting the variability among individuals [59], [26]. This work was part of Natacha Go's post-doctorate, supported by the ANR MIHMES project, in collaboration with the Roslin Institute, Edinburgh, UK. It benefited from the resources and support of NEF computation cluster.

7.1.1.3. Mathematical study of ecological models

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Nicolas Bajeux, Suzanne Touzeau, Israël Tankam Chedjou, Samuel Nilusmas.

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [85]. 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 [11]. External perturbations of interacting populations occur when some individuals are introduced or removed from a natural system, which occurs frequently in pest control applications, either through the direct removal of pests, through the introduction of artificial habitats for the predators or through the introduction of biological control agents in deterministic or stochastic fashion [72].

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; it gives the reference time for crop rotation between resistant and sensitive strains in a mutli-seasonal optimization of root-knot nematodes control [56]. However, it can also arise in tropical environments where, in the absence of weather-related season, time is divided into cropping and fallow seasons, and where the duration of the latter can for example be used as a control method against phytopathogenic nematodes of the plantain plant [46], [58].

7.1.1.4. Analysis of periodic behavior with hybrid models

Participants: Jean-Luc Gouzé, Madalena Chaves.

Periodic orbits in non monotonic negative feedback circuits. In [91], we studied 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. We have now [36] further established uniqueness and stability of the periodic solution under some conditions on the parameters.

Analysis tools for interconnection of Boolean networks. Following the work in [96] and [78], we have generalized the method for computation of the asymptotic graph. In [22], a quantitative dimension is added to the asymptotic graph, through the computation of relative probabilities for each final attractor. In [19], we have extended this methodology for the case of Boolean networks with synchronous updates, in a collaboration with D. Figueiredo and M.A. Martins from the University of Aveiro, Portugal (project PHC Pessoa).

7.1.1.5. Dynamics of complex feedback architectures

Participants: Jean-Luc Gouzé, Madalena Chaves.

To analyze the closed-loop dynamics of metabolic pathways under gene regulation, we propose a method to construct a state transition graph for a given regulatory architecture consisting of a pathway of arbitrary length, with any number of genetic regulators, and under any combination of positive and negative feedback loops [20]. Using this formalism, we analyze a "metabolator"-like mechanism (a pathway with two metabolites and three enzymes) and prove the existence of two co-existing oscillatory behaviors: damped oscillations towards a fixed point or sustained oscillations along a periodic orbit [21].

7.1.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Olivier Bernard, Valentina Baldazzi, Claudia Lopez Zazueta, Lucie Chambon, Agustin Yabo.

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 [12]. This is part of the PhD thesis of Stefano Casagranda (2017), and done in collaboration with Inria IBIS project-team, in particular with D. Ropers.

Analysis and reduction of a model of sugar metabolism in peach fruit. Predicting genotype-to-phenotype relationships is a big challenge for plant biology and breeding. A model of sugar metabolism in peach fruit has been recently developed and applied to 10 peach varieties [25]. A reduction pipeline combining serveral strategies is currently developed to reduce both model size and nonlinearity and allow for further application to virtual breeding (collaboration with B. Quilot-Turion and Mohamed Memmah (INRA Avignon) as part of the PhD thesis of Hussein Kanso).

Analysis of an integrated cell division-endoreduplication and expansion model. The development of a new organ depends on cell-cyle progression and cell expansion, but the interaction and coordination between these processes, under different environments, is still unclear [29]. An integrated model of fruit development has been developed and used to test different interaction schemes, by comparing simulation results to observed cell ditribution data in tomato fruit [65], [47].

7.1.2.1. 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) [80]. We developed different versions of the problem [40], and consider a new problem where the aim is to optimize the production of a product [39],(ANR projects Reset and Maximic, new PhD thesis of A. Yabo, collaboration with McTao Team). We also study variations of the model, including energy (ATP and ADP).

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 [94], [38].

Hybrid control of genetic networks. We design control strategies based on the measurement and control of a unique gene within positive or negative loops of genetic networks, in order to stabilize the system around its unstable fixed point. The quantized nature of genetic measurements and the new synthetic control approaches available in biology encourage the use of piecewise constant control laws. A specific partitioning of the state space and the study of successive repulsive regions allow to show global convergence and global stability for the resulting system [48] . This is part of the thesis of L. Chambon.

7.1.2.2. Slow-Fast analysis of metabolic models

Metabolic modeling generally assumes balanced growth, *i.e.* that there is no accumulation of intermediate compound, and that the metabolism is rapidly at quasi steady state. We go beyond this hypothesis by considering that some metabolic reactions are slow, while other are fast. Then we analyse the differential system using Tikhonov's Theorem. We compare the results obtained using the Drum approach [2], and show that Drum is a reasonable approximation, provided that growth rate stays low. This is part of the PhD thesis of Claudia Lopez Zazueta [31], [30], [54], [55].

7.1.2.3. Large scale metabolic modeling

Metabolic modeling generally assumes balanced growth, *i.e.* that there is no accumulation of intermediate compound, and that the metabolism is rapidly at quasi steady state. We have proposed a new approach called DRUM where this hypothesis is relaxed by splitting the metabolic network into subnetworks and assuming that some compounds can accumulate between the subnetworks [2], [73]. This approach was successfully applied to several cases where the variations in light or nutrient resources induce a strong accumulation in the microalgal cells which could not be represented by the state of the art approaches [74]. More recently we have expended this approach to identify the genomic regulations explaining the change in metabolism especially when considering nitrogen starvation under a light/dark regime.

7.1.3. Biochemical and signaling models

7.1.3.1. Analysis and coupling of biological oscillators

Participants: Sofia Almeida, Madalena Chaves, Eleni Firippi.

Modeling, analysis and coupling of the mammalian cell cycle and clock. Each biological oscillator was modeled by a system of non-linear ordinary differential equations and its parameters calibrated against experimental data (both from the literature and from F. Delaunay's lab). The interactions between the two oscillators are investigated under uni- or bi-directional coupling schemes. Numerical simulations replicate the oscillators' period-lock response and recover observed clock to cell cycle period ratios such as 1:1, 3:2 and 5:4 (as observed in experiments, F. Delaunay's lab) mycitePhD:almeida. This work is in collaboration with F. Delaunay (ANR ICycle) and part of the PhD thesis of Sofia Almeida.

Improving the design of a synthetic oscillator. We analyse a two-variable model (the "Smolen" oscillator) using both numerical simulations and theoretical analysis through a piecewise affine approximation. Our objective is to investigate the existence of oscillatory behaviour and, in particular, to characterize and increase the region of parameters which admits sustained oscillations. This work is part of the PhD thesis of Eleni Firippi (ANR ICycle).

7.1.3.2. Modeling the apoptotic signaling pathway

Participants: Madalena Chaves, Luis Gomes Pereira, Jérémie Roux.

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 [57] (collaboration with J. Roux, for the PhD thesis of Luis Pereira; project Imodrez).

7.2. Fields of application

7.2.1. Bioenergy

Participants: Olivier Bernard, Antoine Sciandra, Walid Djema, Ignacio Lopez Munoz, Ouassim Bara, Jean-Philippe Steyer.

7.2.1.1. Modeling microalgae production

Experimental developments

Running experiments in controlled dynamical environments. The experimental platform made of continuous photobioreactors driven by a set of automaton controlled by the ODIN software is a powerful and unique tool which gave rise to a quantity of very original experiments. Such platform improved knowledge of several biological processes such as lipid accumulation or cell cycle under light fluctuation, etc.

This experimental platform was used to control the long term stress applied to a population of microalgae. This Darwinian selection procedure generated two new strains after more than 6 months in the so called selectiostats.

Other experiments were carried out to reproduce the light signal percept by a cell in a raceway pond [24], derived from Lagrangian hydrodynamical computations. The experiments show that pigments content of the microalgae is highly related to the experimented hydrodynamic regime.

On top of this, we carried out outdoor pilot experiments with solar light. We tested the impact of various temperatures, resulting from different shadowing configurations on microalgal growth rate. This is the topic of Bruno Assis Pessi's master thesis. The impact of process configuration on CO_2 transfer rate has also been tested and quantified [17].

These works have been carried out in collaboration with A. Talec and E. Pruvost (CNRS/Sorbonne Université -Oceanographic Laboratory of Villefranche-sur-Mer LOV).

Metabolism of carbon storage and lipid production. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*, on the basis of the DRUM framework , in order to simulate autotophic, heterotropic and mixotrophic growth, and to determine how to reduce substrate inhibition. The model was extended for other substrates such as glucose or glycerol. A simplified model was developped by I. Lopez to represent the dynamics of polar lipids, especially when faced to higher oxygen concentration.

Modeling the coupling between hydrodynamics and biology. In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on an original multilayer discretisation of Navier-Stokes equations) with microalgae growth was developed [75]. This model is supported by the work of ANGE aiming at improving the discretization scheme of the Navier-Stokes equations and eventually to more accurately represent the hydrodynamics of the raceway and reconstruct Lagrangian trajectories. The accurate reconstruction of the trajectories is verified by a statistical analysis of the probability densities. As a consequence, more relevant experimental protocols have been proposed to more realistically design simplified light signal for experiments [24].

Modeling photosynthetic biofilms. Several models have been developed to represent the growth of microalgae within a biofilm. A first structured physiological model uses mixture theory to represent the microalgae growth, based on the consideration of intracellular reserves triggering the processes of growth, respiration and excretion. We consider separately the intracellular storage carbon (lipids and carbohydrates) and the functional part of microalgae [92]. Another approach accounts for the dynamics of the light harvesting systems when cells are submitted to rapid successions of light and dark phases. A simpler model was developed and used to identify the optimal working mode of a process based on photosynthetic biofilm growing on a conveyor belt [41].

Modeling microalgae production processes. The integration of different models developed within BIOCORE [76] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems. The model was validated at industrial scale with cultivation of the microalgae *Dunaliella salina* [15].

This model was then used to predict productivity in raceway systems, depending on climatic conditions. A Model Predictive Control strategy was developed to on-line adapt influent flow rate and water depth to temperature and light.

We have shown in [87] that a control strategy based on shadowing with solar panel can significantly improve productivity, especially during the early growth stage of the culture.

Modeling thermal adaptation in microalgae. We have studied and compared several models of microalgae growth to different temperatures [82]. Experiments have been carried out in collaboration with A.-C. Baudoux (Biological Station of Roscoff) in order to study growth of various species of the microalgae genus *Micromonas* at different temperatures. After calibration of our models, we have shown that the pattern of temperature response is strongly related to the site where cells were isolated. We derived a relationship to extrapolate the growth response from isolation location. With this approach, we proved that the oceanwide diversity of *Micromonas* species is very similar to the oceanwide diversity of the phytoplankton. We have used Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. We could then predict the evolution of this biodiversity in a warming ocean and show that phytoplankton must be able to adapt within 1000 generation to avoid a drastic reduction in biodiversity [23].

Modeling viral infection in microalgae. Experiments have been carried out in collaboration with A.-C. Baudoux (Biological Station of Roscoff) in order to study the impact of viral infections on the development of populations of *Micromonas* at different temperatures. This work revealed a qualitative change in viral infection when temperature increases. A model was developed to account for the infection of a *Micromonas* population, with population of susceptible, infected and also free viruses. The model turned out to accurately reproduce the infection experiments at various temperatures, and the reduction of virus production above a certain temperature [79].

7.2.1.2. Control and Optimization of microalgae production

Optimization of the bioenergy production systems. A model predictive control algorithm was run based on simple microalgae models coupled with physical models where culture depth influences thermal inertia. Optimal operation in continuous mode for outdoor cultivation was determined when allowing variable culture depth. Assuming known weather forecasts considerably improved the control efficiency.

Interactions between species. We have proposed an optimal control strategy to select the microalgal strain with the lowest pigment content. The control takes benefit from photoinhibition to compute light stresses penalizing the strains with a higher pigment content and finally select microalgae with lower chlorophyll content. This characteristic is of particular interest for maximizing biomass production in dense cultures. The strategy has been carried out at the LOV and eventually the productivity of *Tisochrysis lutea* was improved by 75%.

Finaly, optimal strategies when selecting the strain of interest within two species competing for the same substrate has been proposed, when dynamics is represented by a Droop model [42].

7.2.2. Biological depollution

7.2.2.1. Control and optimization of bioprocesses for depollution

Participants: Olivier Bernard, Carlos Martinez Von Dossow, Jean-Luc Gouzé.

We consider artificial ecosystems including microalgae, cyanobacteria and bacteria in interaction. The objective is to more efficiently remove inorganic nitrogen and phosphorus from wastewater, while producing a microalgal biomass which can be used for biofuel or bioplastic production. Models have been developed including predators grazing the microalgae. Experiments with nitrogen fixing cyanobacteria were carried out, and simple models of the ecosystem where developed to assess the potential of such organisms to support the nitrogen need of microalgae [18].

7.2.2.2. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Carlos Martinez Von Dossow.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Phycover project is aiming at evaluating the potential of this process [95].

We have proposed several models to account for the biodiversity in the microalgal pond and for the interaction between the various species. These models were validated with data from the Saur company. More specifically, we have included in the miroalgae model the impact of the strong turbidity, and derived a theory to better understand the photolimitation dynamics especially when accounting for the photo-inhibition in the illuminated periphery of the reactor [33]. Optimal control strategies playing with the dilution rate, shadowing or modifying depth were then studied [32].

7.2.2.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer, Marjorie Alejandra Morales Arancibia.

Environmental impact assessment. In the sequel of the pioneering life cycle assessment (LCA) work of [83], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable [62].
In the Purple Sun ANR-project, we studied a new paradigm to improve the energy balance by combining biofuel production with photovoltaic electricity. The LCA of a greenhouse with, at the same time, photovoltaic panels and low emissivity glasses is studied. Depending on the period of the year, changing the species can both improve productivity and reduce environmental footprint.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE (Laboratory of Environmental Biotechnology, Narbonne) and Pierre Collet (IFPEN).

7.2.3. Design of ecologically friendly plant production systems

7.2.3.1. Controlling plant arthropod pests

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Nicolas Bajeux, Yves Fotso Fotso.

Optimization of introduction processes. 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 [84], [90] [71], and was one of the key features of L. Mailleret's HDR thesis [11]. A central theoretical result concerns the unveiling of the crucial influence of within-predator density dependent processes. To evaluate this theoretical prediction in a more realistic, stochastic and spatially explicit setting, a stochastic individual based model has been built on the multi-agent programmable modeling environment Netlogo. Extensive simulatory experiments were performed to assess the effects of density dependent processes as well as spatial structure and stochasticity on augmentative biological control performance and variability [67], [68].

In a more general setting, we studied the impact on the introduction success of a population of the interplay of Allee effects, stochasticity in introduction sizes, and occurrence of catastrophes that temporarily wipe out the population. The mean first passage time (MFPT) for a population to reach a viable size was used as a measure of establishment success for the introduction processes [72].

Characteristics of space and the behavior and population dynamics of parasitoids. We studied the influence of the spatial structure and characteristics of the environment on the establishment and spread of biological control agents through computer simulations and laboratory experiments on parasitoids of the genus Trichogramma. This was the topic of Thibaut Morel Journel [89] and is the topic of Marjorie Haond's PhD thesis (ISA, 2015-). The two last articles associated with Thibaut Morel Journel's Thesis appeared this year. In the first one [34], we investigated the effect of habitat fragmentation on the establishment and early spread of an introduced population. We showed that by increasing the risks of dispersal from the introduction site to unfavourable habitat early during the invasion, fragmentation decreased establishment success. However, by decreasing the distance between favourable habitat patches, it also improved the subsequent spread of introduced species over larger areas. In the second paper [35], we explored the influence of different characteristics of the structural connectivity of an invaded habitat on the invading population. We demonstrated how spread

was hindered by habitat clusters and accelerated by the presence of hubs. These results highlight the importance of considering the structure of the invaded area to predict the outcome of invasions. In a different study stemming from Marjorie Haond Thesis, we showed how habitat richness [27] as represented by its local carrying capacity can positively influence the spreading speeed of an expanding population. This work is being performed in collaboration with Elodie Vercken (ISA) and Lionel Roques (BioSP, Avignon).

In a metapopulation context, we studied the invasion success into an environment where part of the patches are sources (favourable environments) and the others are sinks; a criterion has been obtained predicting invasion success when the number of sources is larger than some threshold [70].

Modeling and control of coffee berry borers. We developed a model describing the coffee berry borer dynamics based on the insect life-cycle and the berry availability during a single cropping season. An optimal control problem was formulated by implementing chemical control (insecticides) and/or biological control (entomopathogenic fungi such as *Beauveria bassiana*, microbial parasitoids, traps). The aim was to maximise the yield at the end of the cropping season, while minimising the borer population for the next cropping season and the control costs. The existence of an optimal solution was shown and the problem was solved numerically [49], [44]. This ODE model was extended to integrate the berry maturation age. The well-posedness of the resulting PDE model was shown and an asymptotic analysis was conducted. This research pertains to Yves Fotso Fotso's PhD thesis, who visited BIOCORE during 5 months in 2018 through the EPITAG associate team.

7.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Pauline Clin.

Sustainable management of plant resistance. We studied other plant protection methods dedicated to fight plant pathogens. One such method is the introduction of plant strains that are resistant to one pathogen. This often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants.

Experiments were conducted in INRA Avignon for Potato Virus Y on pepper plants to evaluate the effect of four traits influencing evolutionary forces leading to resistance breakdown: virus effective population sizes, either at plant inoculation or during infection, virus accumulation and differential selection during infection. A generalized linear model showed a strong impact of the second and third one while a positive interection between differential selection and virus accumulation was identified [37]. Also, a stochastic model was developed to help determine the efficiency of pyramiding qualitative resistance and quantitative resistance narrowing population bottlenecks exerted on viruses, the latter aiming at slowing down virus adaptation to the qualitative resistance. It showed the efficiency of pyramiding when the fitness cost of RB virus variants in susceptible plants is intermediate [93]. These studies provide a framework to select plants with appropriate virus-evolution-related traits to avoid or delay resistance breakdown. This was done in collaboration with Frédéric Fabre (INRA Bordeaux) and Benoît Moury (INRA Avignon).

We pursued the calibration of the (spatio-)temporal epidemiological model of the phoma stem canker of oilseed rape, using field data on resistance deployment and virulence of phoma populations. Ongoing work includes the development of a simulation tool designed for researchers as well as non academic partners from technical institutes and agriculture cooperatives, who interact through the MoGeR project. It benefits from the resources and support of NEF computation cluster.

Taking advantage of plant diversity and immunity to minimize disease prevalence. An epidemiological model of gene-for-gene interaction considering a mechanism related to the specific defense response of plants, the systemic acquired resistance (SAR) was developed. SAR provides a sort of immunity to virulent pathogens for resistant plants having undergone an infection attempt by an avirulent pathogen. This model showed that there exists an optimal host mixture that ensures the lowest plant disease prevalence, so as to optimize the crop yield. It is especially efficient for pathogens with a low or intermediate basic reproduction rate and hosts with a high SAR efficiency [51], [52]. This was the topic of Pauline Clin's master thesis and was done in collaboration with Frédéric Hamelin (Agrocampus Ouest).

7.2.3.3. Plant-nematode interactions.

Participants: Valentina Baldazzi, Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Israël Tankam Chedjou, Samuel Nilusmas.

Phytophagous nematodes are small little-mobile worms that feed and reproduce on plant roots, generating considerable losses in numerous crops all over the world. Most eco-friendly plant protection strategies are based on the use of resistant crops, but agricultural practices also contribute to nematode control.

We developed a first physiological model of plant–nematode interactions, explicitly describing resource (water and carbon) allocation between roots and shoots. Indeed, nematodes draw on root carbon pool and reduce plant water uptake from the soil. The consequences on plant growth were analyzed as a function of plant physiological characteristics. In parallel, an experiment was conducted on pepper and tomato plants to monitor plant growth with or without nematodes. Data will be used to calibrate the model. This work was the topic of Thomas Brenière [77] and was done in collaboration with Caroline Djian-Caporalino (ISA, INRA Sophia Antipolis).

We studied the stability of the hybrid interaction model between nematodes and plantain roots [46]. An optimisation problem was formulated to determine the duration between cropping seasons (fallow period) that maximises the farmer's cumulated yield, which is affected by the nematode population, while minimising the costs of nematode control and nursery-bought pest-free suckers, on a fixed time horizon that lasts several cropping seasons. We first considered that the farmer buys and plants pest-free suckers at the beginning of each cropping season. This allows for a fallow period which reduces the nematode population in the soil, as these pests need roots to feed on and reproduce. Two cases were considered: a fixed or a variable fallow period. In the first case, the existence of an optimal solution was proven and its location was computed for small infestations. In the second case, the existence of an optimal strategy was proven and was numerically computed [58]. This research pertains to Israël Tankam Chedjou's PhD thesis, who visited BIOCORE during 5 months in 2018 through the EPITAG associate team.

We studied the resistance-based nematode control. As virulent nematodes exhibit a reduced fitness on susceptible crops, combining both resistant and susceptible plants can help increase the efficacy and sustainability of such control methods. In the *Solanaceae* family, there are two major resistance genes: the first one induces an early reaction when the nematode enters in the root system and the second one induces a late reaction when the nematode creates its feeding site. We used a semi-discrete model describing the plant-nematode interactions within and between cropping season to implement the action of both resistance genes. We computed and compared the optimal deployment strategies of both resistant crops [56]. This research pertains to Samuel Nilusmas' PhD thesis (2016–).

7.2.3.4. Optimality/games in population dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Optimal resource allocation. Mycelium growth and sporulation are considered for phytopathogenic fungi. For biotrophic fungi, a flow of resource is uptaken by the fungus without killing its host; in that case, life history traits (latency-sporulation strategy) have been computed based on a simple model considering a single spore initiating the mycelium, several spores in competition and applying optimal resource allocation, and several spores in competition through a dynamic game through the analytico-numerical solution of the Hamilton-Jacobi-Bellman-Isaacs equation [97]. The solution of this dynamic game has been shown to be the equilibrium of two-trait adaptive dynamics [50]. This work, in the framework of the ANR Funfit project, is done with Fabien Halkett of INRA Nancy.

Optimal foraging and residence times variations. In this work, we built on our re-analysis of the Marginal Value Theorem (MVT) [4] to study the effect on the optimal foraging strategy of habitat conversion, whereby patches are converted from one existing type to another, hence changing the frequency of each type in the environment. We studied how realized fitness and the average rate of movement should respond to changes in the frequency distribution of patch-types, and how they should covary. We found that the initial pattern of patch-exploitation in a habitat can help predict the qualitative responses of fitness and movement rate following habitat conversion. We conclude that taking into account behavioral responses may help better understand the

ecological consequences of habitat conversion. This work was published through the novel preprint reviewing system of Peer Community In Ecology [66].

CARMEN Project-Team

6. New Results

6.1. Exponential Adams Bashforth integrators for cardiac electrophysiology simulations

Models in cardiac electrophysiology are coupled systems of reaction-diffusion PDEs and ODEs. The ODE system displays a very stiff behavior. It is nonlinear and its upgrade at each time step has a high computational cost. A solution that we propose is to develop high-order explicit and stable integration methods. In an article published in 2018 [17] we investigated the use of exponential Adams Bashforth (EAB) integrators in cardiac electrophysiology. The paper demonstrates stability under perturbation (or 0-stability) and provides a new approach for the convergence analysis of the method. The Dahlquist stability properties of the method were also tested in a new framework that incorporates the discrepancy between the stabilizer and the system Jacobian matrix. Provided this discrepancy is small enough, the method was shown to be A(alpha)-stable. This result is interesting for an explicit time-stepping method. Numerical experiments were performed for two classes of stiff models in cardiac electrophysiology. The EAB method was observed to be as stable as implicit solvers and cheaper at the same level of accuracy.

6.2. Inverse models for identification of cellular ionic parameters from macroscopic signals

Traditional inverse models in cardiac electrophysiology have aimed at identifying the activation (and relaxation) order of the cardiac muscle from body surface potentials. However, many cardiac anomalies cannot be reduced to such simple parameters. Underlying the activation and “repolarization” of the cells is a complex interplay of different ion channels, each with its own dynamics. Genetic and other abnormalities express themselves in one or more of these channels. We have therefore taken several initiatives to identify the properties of the channels themselves.

Measurements with micro-electrodes can capture currents generated by a handful of cultured cells. Together with the REO team at the Inria center in Paris we have developed methods to identify properties of individual ionic currents from such measurements [28], [10].

On the level of the whole heart, two studies have attempted to identify cellular properties from surface electrocardiogram (ECG) signals [29], [11]. Ravon et al. have previously developed a method that identifies activation times as well as repolarization properties of a simplified ionic model (the Mitchell-Schaeffer model) [66], but considering only the outer surface of the heart muscle. We now evaluated the capability of this method to distinguish properties of the inner and outer surfaces, which are expected to differ importantly [29]. This turned out to be extremely challenging. Abidi et al. [11], on the other hand, demonstrated that the solution to such problems is unique even for intermediately complex ionic models such as the Beeler-Reuter [46] and Luo-Rudy I [61] models.

Thus, we found that a unique solution exists but is very hard to find in practical situations. This result suggests that further work should aim at removing confounding factors such as limitations in the volume conductor models.

6.3. Cellular and sub-cellular models

The electrical activation of the heart relies on rapid propagation of activation impulses through intercellular connections. Cardiac arrhythmia are often due to damage to these intercellular connections. In various pathologies, loss of individual connections can lead to the formation of a complicated maze in which very slow propagation is possible, leading to reentrant arrhythmias. To investigate such phenomena, the PhD thesis work of P.-E. Bécue was dedicated to the development of a three-dimensional model of the cardiac tissue with sub-cellular resolution [31], [50], [51], [52], [49]. This work builds on our CEPS software, which was specifically extended for this purpose.

Aouadi et al. have specifically developed a model for the connections between the network of cardiac Purkinje fibers and the working myocardium [13] [45]. It handles the myocardium with a standard bidomain model and the Purkinje network, which consists of discrete bundles, with a one-dimensional monodomain model.

On an even smaller scale, the team also worked on a three-dimensional model of calcium release from intracellular organelles and its diffusion inside the cell [42].

6.4. Highly scalable ECG simulation with electrocardiographic lead fields

Currently a monodomain reaction-diffusion model is a well-established method to simulate the electrical activity of the heart [62], [63], even more so because it can be adapted to approximate a bidomain model very closely [48], [54]. Computing the electrocardiogram (ECG) from the results of such models is harder because it requires large linear systems to be solved, and does not scale well to large numbers of processors. A possible solution is to use so-called lead fields, the electrocardiographic term for a linear combination of fundamental solutions that express the ECG potential as an integral over a field of electric current dipoles. A paper published in 2018 has shown that this method gives a huge scaling advantage on highly parallel computers [27]. This result is also of practical importance for our applied work.

DRACULA Project-Team

5. New Results

5.1. Oscillations and asymptotic convergence for a delay differential equation modeling platelet production

In [13], a model for platelet production is introduced for which the platelet count is described by a delay differential equation $P'(t) = -\gamma P(t) + f(P(t))g(P(t-r))$ where f and g are positive decreasing functions. First, the authors study the oscillation of the solutions around the unique equilibrium of the equation above, obtaining an inequality implying such an oscillation. They also obtain provide a condition such that this inequality is necessary and sufficient for oscillation. This result is compared to already existing results and the biological meaning of the inequality is studied. The authors also present a result on the asymptotic convergence of the solutions. This result depends on the behavior of the solution for $t \in [0, r]$, and the authors provide an analysis of the link between this behavior and the initial conditions in the case of a simpler model.

5.2. Meningioma growth dynamics assessed by radiocarbon retrospective birth dating

It is not known how long it takes from the initial neoplastic transformation of a cell to the detection of a tumor, which would be valuable for understanding tumor growth dynamics. We have assessed the age and growth dynamics in patients with WHO grade I meningiomas by combining retrospective birth-dating of cells by analyzing incorporation of nuclear-bomb-test-derived 14C, analysis of cell proliferation, cell density, MRI imaging and mathematical modeling. We provide an integrated model of the growth dynamics of benign meningiomas. The mean age of WHO grade I meningiomas was 22.1$\pm$ 6.5 years. We conclude that WHO grade I meningiomas are very slowly growing brain tumors, which are resected in average two decades after time of origination. [18]

5.3. Existence and stability of periodic solutions of an impulsive differential equation and application to CD8 T-cell differentiation

In this article [16], we study a scalar impulsive differential equation (IDE) with the aim of studying the effects of uneven molecular partitioning upon cell mitosis on CD8 T-cell differentiation. To do so, we introduce mathematical results that stand for a more general class of IDE, then apply them to our IDE and discuss those results with regard to the initial biological problem.

5.4. Investigating the role of the experimental protocol in phenylhydrazine-induced anemia on mice recovery

Erythropoiesis, the process of production of red blood cells, is performed through complex regulatory processes. We proposed an earlier model describing stress erythropoiesis in mice [33]. This model, based on the description of erythroid progenitor and erythrocyte dynamics using delay equations, led us to conclude on the quantitative importance of self-renewal. In [6], we refined this previous approaches by taking into account a more mechanistic description of the induction of anemia via phenylhydrazine injection. This led us to revisit some of our initial hypothesis regarding self-renewal regulation.

5.5. Generalizing a mathematical model of prion aggregation allows strain coexistence and co-stability by including a novel misfolded species

Prions are proteins capable of adopting misfolded conformations and transmitting these conformations to other normally folded proteins. A distinct feature of prion propagation is the existence of different phenotypical variants, called strains. In order to conform to biological observations of strain coexistence and co-stability, we develop in [19] an extension of the classical model by introducing a novel prion species consistent with biological studies.

5.6. Analysis and Numerical Simulation of a Polymerization Model with Possible Agglomeration Process

The purpose of [20] is to provide analytical and numerical results for a general polymerization model with lengthening process by agglomeration. 2D spatial diffusion of monomers is taken into account for the mass transfer between monomers and polymers. The analysis of the model is performed thanks to a double fixed point theorem. Adequate numerical scheme based on a generalization of the anti-dissipative method developed in Goudon (Math. Models Methods Appl. Sci. 23:1177–1215, 2013

5.7. The Origin of Species by Means of Mathematical Modelling

Darwin described biological species as groups of morphologically similar individuals. These groups of individuals can split into several subgroups due to natural selection, resulting in the emergence of new species. Some species can stay stable without the appearance of a new species, some others can disappear or evolve. In [10] we have developed a model which allows us to reproduce the principal patterns in Darwin's diagram. Some more complex evolutionary patterns are also observed. The relation between Darwin's definition of species, stated above, and Mayr's definition of species (group of individuals that can reproduce) is also discussed.

5.8. Improved duality estimates in the time discrete case for cross diffusion models

In [28], time discrete versions of the duality estimates derived by Canizo et al. for parabolic systems have been obtained. They allow the construction of solution with superquadratic reactions terms for cross diffusion models with bounded pressure.

M3DISIM Project-Team

7. New Results

7.1. Mathematical and Mechanical Modeling

7.1.1. Microscopic model of collagen fiber

Participants: Florent Wijanto, Matthieu Caruel, Jean-Marc Allain [correspondant].

Our studies on collagen tissues have shown that the collagen fibers are able to elongate inelastically under stretch. In tendon, this effect has been attributed to the non-permanent cross-bridges which connect the different collagen fibrils (to assemble a fiber). This sliding effect appears experimentally to be reversible (at least partially) if the tissue is left long enough at its initial resting length. However, this sliding is classically included as an irreversible plastic response, or as a damage of the tissue. We are building a model based on a stochastic description of the binding and unbinding of the cross-bridges. This approach will enable us to have a microscopically based picture of the sliding, which will be able to explain some alteration in case of ageing or of pathological alterations of the tissue. At the moment, we have shown the importance of the density of cross-bridges in the cooperative response of the system. A publication is in preparation on the topic.

7.1.2. Stochastic modeling of chemical-mechanical coupling in striated muscles

Participants: Matthieu Caruel, Dominique Chapelle [correspondant], Philippe Moireau.

We propose a chemical-mechanical model of myosin heads in sarcomeres, within the classical description of rigid sliding filaments. In our case, myosin heads have two mechanical degrees-of-freedom (dofs) – one of which associated with the so-called power stroke – and two possible chemical states, i.e. bound to an actin site or not. Our major motivations are twofold: (1) to derive a multiscale coupled chemical-mechanical model, and (2) to thus account – at the macroscopic scale – for mechanical phenomena that are out of reach for classical muscle models. This model is first written in the form of Langevin stochastic equations, and we are then able to obtain the corresponding Fokker-Planck partial differential equations governing the probability density functions associated with the mechanical dofs and chemical states. This second form is important, as it allows to monitor muscle energetics, and also to compare our model with classical ones, such as the Huxley'57 model to which our equations are shown to reduce under two different types of simplifying assumptions. This provides insight, and gives a Langevin form for Huxley'57. We then show how we can calibrate our model based on experimental data – taken here for skeletal muscles – and numerical simulations demonstrate the adequacy of the model to represent complex physiological phenomena, in particular the fast isometric transients in which the power stroke is known to have a crucial role, thus circumventing a limitation of many classical models. This work is accepted for publication in BMMB.

7.1.3. The importance of the pericardium for cardiac biomechanics

Participant: Radomir Chabiniok [correspondant].

The human heart is enclosed in the pericardial cavity. The pericardium consists of a layered thin sac and is separated from the myocardium by a thin film of fluid. It provides a fixture in space and frictionless sliding of the myocardium. The influence of the pericardium is essential for predictive mechanical simulations of the heart. However, there is no consensus on physiologically correct and computationally tractable pericardial boundary conditions. Here we propose to model the pericardial influence as a parallel spring and dashpot acting in normal direction to the epicardium. Using a four-chamber geometry, we compare a model with pericardial boundary conditions to a model with fixated apex. The influence of pericardial stiffness is demonstrated in a parametric study. Comparing simulation results to measurements from cine magnetic resonance imaging reveals that adding pericardial boundary conditions yields a better approximation with respect to atrioventricular plane displacement, atrial filling, and overall spatial approximation error. We demonstrate that this simple model of pericardial-myocardial interaction can correctly predict the pumping

mechanisms of the heart as previously assessed in clinical studies. Utilizing a pericardial model can not only provide much more realistic cardiac mechanics simulations but also allows new insights into pericardial-myocardial interaction which cannot be assessed in clinical measurements yet. The work was accepted for publication in Biomechanics and Modeling in Mechanobiology [26], and is a joint work with Technical University in Munich, Germany (group of W.A. Wall) and Bernoulli Institute for Mathematics at University of Groningen, The Netherlands (C. Bertoglio).

7.1.4. Solving 2D linear isotropic elastodynamics by means of scalar potentials: a new challenge for finite elements

Participants: Sébastien Imperiale, Patrick Joly [Poems].

In this work we present a method for the computation of numerical solutions of 2D homogeneous isotropic elastodynamics equations by solving scalar wave equations. These equations act on the potentials of a Helmholtz decomposition of the displacement field and are decoupled inside the propagation domain. We detail how these equations are coupled at the boundary depending on the nature of the boundary condition satisfied by the displacement field. After presenting the case of rigid boundary conditions, that presents no specific difficulty, we tackle the challenging case of free surface boundary conditions that presents severe stability issues if a straightforward approach is used. We introduce an adequate functional framework as well as a time domain mixed formulation to circumvent these issues. Numerical results confirm the stability of the proposed approach.

7.1.5. Lung multiscale poromechanical modeling, from breathing to pulmonary fibrosis-induced chronic remodeling

Participants: Cécile Patte [correspondant], Martin Genet, Dominique Chapelle.

Pulmonary diseases are about to become the third cause of death in the world. One on them, Idiopathic Pulmonary Fibrosis (IPF), which involves thickening, stiffening and destruction the alveolar walls, remains poorly understood, diagnosed and treated. It has been hypothesized, however, that IPF involves a mechanical vicious circle, where fibrosis induces higher stresses, which in turns favors fibrosis. In this project, we intend to better understand the role of mechanics in the disease progression, in order to improve diagnosis and prognosis. We model the lung behavior during breathing at organ-scale, based on a poromechanical theory, previously established in the team. Then we estimate the regional mechanical properties of the lung, based on clinical data. In the future, the procedure can be used as a prognostic tool by the clinicians.

7.1.6. Mathematical modelling of transient shear wave elastography in the heart

Participants: Federica Caforio [correspondant], Sébastien Imperiale.

The aim of this work is to provide a mathematical model of the excitation and the resulting shear wave propagation in Acoustic Radiation Force (ARF)-based shear wave cardiac elastography. Our approach is based on asymptotic analysis; more precisely, it consists in considering a family of problems, parametrised by a small parameter inversely proportional to the excitation frequency of the probes, the viscosity and the velocity of pressure wave propagation. We derive a simplified model for the expression of the ARF by investigating the limit behaviour of the solution when the small parameter goes to zero. By formal asymptotic analysis - an asymptotic expansion of the solution is used - we show that the leading order term of the expansion is the underlying nonlinear cardiac mechanics. Subsequently, two corrector terms are computed. The first is a fast-oscillating pressure wave generated by the probes, solution of a Helmholtz equation at every time instant. The second corrector term consists in an elastic field with prescribed divergence, having a function of the first corrector as a source term. This field corresponds to the shear acoustic wave induced by the ARF. We also confirm that, in cardiac mechanics, the presence of viscosity in the model is essential to derive an expression of the shear wave propagation from the ARF, and that this phenomenon is related to the nonlinearity of the partial differential equation.

7.1.7. Analysis and calibration of a linear model for structured cell populations with unidirectional motion : Application to the morphogenesis of ovarian follicles

Participants: Frédérique Clément, Frédérique Robin [correspondant], Romain Yvinec [INRA].

In [41], we have analyzed a multi-type age dependent model for cell populations subject to unidirectional motion, in both a stochastic and deterministic framework. Cells are distributed into successive layers; they may divide and move irreversibly from one layer to the next. We have adapted results on the large-time convergence of PDE systems and branching processes to our context, where the Perron-Frobenius or Krein-Rutman theorems cannot be applied. We have derived explicit analytical formulas for the asymptotic cell number moments, and the stable age distribution. We have illustrated these results numerically and we have applied them to the study of the morphodynamics of ovarian follicles. We have proven the structural parameter identifiability of our model in the case of age independent division rates. Using a set of experimental biological data, we have estimated the model parameters to fit the changes in the cell numbers in each layer during the early stages of follicle development.

7.1.8. A multiscale mathematical model of cell dynamics during neurogenesis in the mouse cerebral cortex

Participants: Frédérique Clément [correspondant], Marie Postel [Sorbonne Universités].

Work in collaboration with Sylvie Schneider-Maunoury (Sorbonne Universités), Alice Karam (Sorbonne Universités), Guillaume Pézeron (MNHN). Neurogenesis in the murine cerebral cortex involves the coordinated divisions of two main types of progenitor cells, whose numbers, division modes and cell cycle durations set up the final neuronal output. To understand the respective roles of these factors in the neurogenesis process, we have combined experimental in vivo studies with mathematical modeling and numerical simulations of the dynamics of neural progenitor cells [43]. A special focus has been put on the population of intermediate progenitors (IPs), a transit amplifying progenitor type critically involved in the size of the final neuron pool. Our multiscale formalism describing IP dynamics allows one to track the progression of cells along the subsequent phases of the cell cycle, as well as the temporal evolution of the different cell numbers. Our model takes into account the dividing apical progenitors (AP) engaged into neurogenesis, both neurogenic and proliferative IPs, and the newborn neurons. The transfer rates from one population to another are subject to the mode of division (symmetric, asymmetric, neurogenic) and may be time-varying. The model outputs have been successfully fitted to experimental cell numbers from mouse embryos at different stages of cortical development, taking into account IPs and neurons, in order to adjust the numerical parameters. Applying the model to a mouse mutant for Ftm/Rpgrip1l, a gene involved in human ciliopathies with severe brain abnormalities, has revealed a shortening of the neurogenic period associated with an increased influx of newborn IPs from apical progenitors at mid-neurogenesis. Additional information have been provided on cell kinetics, such as the mitotic and S phase indexes, and neurogenic fraction. Our model can be used to study other mouse mutants with cortical neurogenesis defects and can be adapted to study the importance of progenitor dynamics in cortical evolution and human diseases.

7.1.9. Advances in computational modeling approaches of pituitary gonadotropin signaling

Participants: Frédérique Clément [correspondant], Romain Yvinec [INRA].

Work in collaboration with Pascale Crépieux, Anne Poupon and Éric Reiter (INRA). We have reviewed thoroughly the state-of-the-art in computational modeling approaches of pituitary gonadotropin signaling [30]. Pituitary gonadotropins play an essential and pivotal role in the control of human and animal reproduction within the hypothalamic-pituitary-gonadal (HPG) axis. The computational modeling of pituitary gonadotropin signaling encompasses phenomena of different natures such as the dynamic encoding of gonadotropin secretion, and the intracellular cascades triggered by gonadotropin binding to their cognate receptors, resulting in a variety of biological outcomes. We have overviewed historical and ongoing issues in modeling and data analysis related to gonadotropin secretion in the field of both physiology and neuro-endocrinology. We have mentioned the different mathematical formalisms involved, their interest and limits. We have discussed open statistical questions in signal analysis associated with key endocrine issues. We have also reviewed

recent advances in the modeling of the intracellular pathways activated by gonadotropins, which yields promising development for innovative approaches in drug discovery. The greatest challenge to be tackled in computational modeling of pituitary gonadotropin signaling is the embedding of gonadotropin signaling within its natural multiscale environment, from the single cell level, to the organic and whole HPG level. The development of modeling approaches of G protein-coupled receptor signaling, together with multicellular systems biology may lead to unexampled mechanistic understanding with critical expected fallouts in the therapeutic management of reproduction.

7.1.10. Structured cell population dynamics applied to the early development of ovarian follicles

Participants: Frédérique Clément, Frédérique Robin [correspondant], Romain Yvinec [INRA].

The ovarian follicles are the basic anatomical and functional units of the ovaries, which are renewed from a quiescent pool all along reproductive life. Follicular development involves a finely tuned sequence of growth and maturation processes, involving complex cell dynamics. Understanding follicular development is a crucial issue for the management of reproduction in a clinical or breeding context, and for the preservation of endangered species. In their early stages of development, ovarian follicles are made up of a germ cell (oocyte), whose diameter increases steadily, and of surrounding proliferating somatic cells, which are layered in a globally spherical and compact structure. We have designed a modeling approach dedicated to the initiation phase of follicle development. The initiation phase is described by joint stochastic dynamics accounting for cell shape transitions (from a flattened to a cuboidal shape) and proliferation of reshaped cells. We have then derived the mean time elapsed before all cells have changed shapes and the corresponding increment in the total cell number, which is fitted to experimental data retrieved from primordial follicles (single layered follicle with only flattened cells) and primary follicles (single layered follicles with only cuboidal cells).

7.1.11. Newton-Krylov method for computing the cyclic steady states of evolution problems in non-linear mechanics

Participants: Ustim Khristenko, Patrick Le Tallec [correspondant].

This work is focused on the Newton-Krylov technique for computing the steady cyclic states of evolution problems in nonlinear mechanics with space-time periodicity conditions. This kind of problems can be faced, for instance, in the modeling of a rolling tire with a periodic tread pattern, where the cyclic state satisfies “rolling” periodicity condition, including shifts both in time and space. The Newton-Krylov method is a combination of a Newton nonlinear solver with a Krylov linear solver, looking for the initial state, which provides the space-time periodic solution. The convergence of the Krylov iterations is proved to hold in presence of an adequate preconditioner. After preconditioning, the Newton-Krylov method can be also considered as an observer-controller method, correcting the transient solution of the initial value problem after each period. Using information stored while computing the residual, the Krylov solver computation time becomes negligible with respect to the residual computation time. The method has been analyzed and tested on academic applications and compared with the standard evolution (fixed point) method. Finally, it has been implemented into the Michelin industrial code, applied to a full 3D rolling tire model.

7.1.12. Delayed feedback control method for computing the cyclic steady states of evolution problems

Participants: Ustim Khristenko, Patrick Le Tallec [correspondant].

This work is focused on fast techniques for computing the cyclic steady states of evolution problems in non-linear mechanics with space–time periodicity conditions. In industrial applications, in order to avoid the inversion of very large matrices, such a cyclic solution is usually computed as an asymptotic limit of the associated initial value problem with arbitrary initial data. However, when the relaxation time is high, convergence to the limit cycle can be very slow. In such cases nonetheless, one is not interested in the transient solution, but only in a fast access to the limit cycle. Thus, in this work we modify the problem, introducing the time-delayed feedback control, which is widely used for stabilization of unstable periodic orbits. In our

framework it is applied to an initially stable system in order to accelerate its convergence to the limit cycle. Moreover, the control term, based on the space–time periodicity error, includes both shifts in time and in space. Our main result is the optimal form of the control term for a very general class of linear evolution problems, providing the fastest convergence to the cyclic solution, which has been further extended and studied in the non-linear case. Efficiency of the method increases with the problem's relaxation time. The method has been tested using academic applications and compared to the non-controlled asymptotic convergence as well as to the Newton–Krylov shooting algorithm. Finally, the method has been implemented into the Michelin industrial code, applied to a full 3D rolling tyre model.

7.2. Numerical Methods

7.2.1. Numerical analysis for an energy-preserving total discretization of a poromechanics model with inf-sup stability

Participants: Dominique Chapelle [correspondant], Philippe Moireau.

We consider a previously proposed general nonlinear poromechanical formulation, and we derive a linearized version of this model. For this linearized model, we obtain an existence result and we propose a complete discretization strategy – in time and space – with a special concern for issues associated with incompressible or nearly-incompressible behavior. We provide a detailed mathematical analysis of this strategy, the main result being an error estimate uniform with respect to the compressibility parameter. We then illustrate our approach with detailed simulation results and we numerically investigate the importance of the assumptions made in the analysis, including the fulfillment of specific inf-sup conditions. This work is accepted for publication in Acta Mathematicae Applicatae Sinica.

7.2.2. Efficient estimation of personalized biventricular mechanical function employing gradient-based optimization

Participant: Martin Genet [correspondant].

Individually personalized computational models of heart mechanics can be used to estimate important physiological and clinically-relevant quantities that are difficult, if not impossible, to directly measure in the beating heart. Here, we present a novel and efficient framework for creating patient-specific biventricular models using a gradient-based data assimilation method for evaluating regional myocardial contractility and estimating myofiber stress. These simulations can be performed on a regular laptop in less than 2 hours and produce excellent fit between measured and simulated volume and strain data through the entire cardiac cycle. By applying the framework using data obtained from 3 healthy human biventricles, we extracted clinically important quantities as well as explored the role of fiber angles on heart function. Our results show that steep fiber angles at the endocardium and epicardium are required to produce simulated motion compatible with measured strain and volume data. We also find that the contraction and subsequent systolic stresses in the right ventricle are significantly lower than that in the left ventricle. Variability of the estimated quantities with respect to both patient data and modeling choices are also found to be low. Because of its high efficiency, this framework may be applicable to modeling of patient specific cardiac mechanics for diagnostic purposes.

7.2.3. Equilibrated warping: Finite element image registration with finite strain equilibrium gap regularization

Participant: Martin Genet [correspondant].

In this work, we propose a novel continuum finite strain formulation of the equilibrium gap regularization for image registration. The equilibrium gap regularization essentially penalizes any deviation from the solution of a hyperelastic body in equilibrium with arbitrary loads prescribed at the boundary. It thus represents a regularization with strong mechanical basis, especially suited for cardiac image analysis. We describe the consistent linearization and discretization of the regularized image registration problem, in the framework of the finite elements method. The method is implemented using FEniCS & VTK, and distributed as a freely available python library. We show that the equilibrated warping method is effective and robust: regularization

strength and image noise have minimal impact on motion tracking, especially when compared to strain-based regularization methods such as hyperelastic warping. We also show that equilibrated warping is able to extract main deformation features on both tagged and untagged cardiac magnetic resonance images.

7.2.4. Thermodynamic properties of muscle contraction models and associated discrete-time principles

Participants: François Kimmig [correspondant], Dominique Chapelle, Philippe Moireau.

Considering a large class of muscle contraction models accounting for actin-myosin interaction, we present a mathematical setting in which solution properties can be established, including fundamental thermodynamic balances. Moreover, we propose a complete discretization strategy for which we are also able to obtain discrete versions of the thermodynamic balances and other properties. Our major objective is to show how the thermodynamics of such models can be tracked after discretization, including when they are coupled to a macroscopic muscle formulation in the realm of continuum mechanics. Our approach allows to carefully identify the sources of energy and entropy in the system, and to follow them up to the numerical applications.

7.2.5. A conservative penalisation strategy for the semi-implicit time discretisation of the incompressible elastodynamics equation

Participants: Federica Caforio [correspondant], Sébastien Imperiale.

The principal aim of this work is to provide an adapted numerical scheme for the approximation of elastic wave propagation in incompressible solids. We rely on high-order conforming finite element with mass lumping for space discretisation and implicit/explicit, second-order, energy-preserving time discretisation. The time step restriction only depends on the shear wave velocity and at each time step a Poisson problem must be solved to account for the incompressibility constraint that is imposed by penalisation techniques.

7.2.6. High-order discrete Fourier transform for the solution of the poisson equation

Participants: Federica Caforio [correspondant], Sébastien Imperiale.

The aim of this work is to propose a novel, fast, matrix-free solver for the Poisson problem discretised with High-Order Spectral Element Methods (HO-SEM). This method is based on the use of the Discrete Fourier Transform to reduce the problem to the inversion of the symbol of the operator in frequency space. The solver proposed is endowed with several properties. First, it preserves the efficiency of standard FFT algorithm; then, the matrix storage is minimised; a pseudo- explicit Singular Value Decomposition (SVD) is used for the inversion of the symbols; finally, it can be easily extended to multiple dimensions and non-periodic boundary conditions. In particular, due to the underlying HO-SEM discretisation, the multi-dimensional symbol of the operator can be efficiently computed from the one-dimensional symbol by tensorisation.

7.3. Inverse Problems

7.3.1. Analysis of an observer strategy for initial state reconstruction of wave-like systems in unbounded domain

Participants: Sébastien Imperiale, Philippe Moireau [correspondant].

We are interested in reconstructing the initial condition of a wave equation in an unbounded domain configuration from measurements available in time on a subdomain. To solve this problem, we adopt an iterative strategy of reconstruction based on observers and time reversal adjoint formulations. We prove the convergence of our reconstruction algorithm with perfect measurements and its robustness to noise. Moreover, we develop a complete strategy to practically solve this problem on a bounded domain using artificial transparent boundary conditions to account for the exterior domain. Our work then demonstrates that the consistency error introduced by the use of approximate transparent boundary conditions is compensated by the stabilization properties obtained from the use of the available measurements, hence allowing to still be able to reconstruct the unknown initial condition. This work is accepted with minor revision for publication in COCV.

7.4. Experimental Assessments

7.4.1. Mathematical modeling and experimental validation of flow through aortic valve

Participant: Radomir Chabiniok [correspondant].

Assessment of the valvular diseases by phase-contrast magnetic resonance imaging (MRI) has known limits due to limited spatial-temporal resolution of MRI and artifacts intrinsic to the method. This problem is addressed by the collaborative work of the Institute for Clinical and Experimental Medicine in Prague (IKEM, participants J. Tintera and R. Galabov) and the mathematical modeling group at the Czech Technical University in Prague (CTU, participants P. Paus, R. Fucik), additionally with the combined clinical cardiovascular MRI & modeling expertise of R. Chabiniok (Inria). A flow phantom was constructed at IKEM and used to perform an extensive experimental study targeted to capture the phenomena in valvular stenosis / regurgitation. The Mathematical modeling group at CTU then performed flow simulations by using the techniques of Lattice-Boltzmann method and their high-performance computing GPU implementations. This work is shedding light into possibly significant factors limiting the direct interpretation of PC MRI and opening the way into interaction of PC MRI data with mathematical model as a "smart filtering" of flow exam.

7.4.2. Skin multiscale mechanics

Participant: Jean-Marc Allain [correspondant].

Skin is a complex, multi-layered organ, with important functions in the protection of the body. The dermis provides structural support to the epidermal barrier, and thus has attracted a large number of mechanical studies. As the dermis is made of a mixture of stiff fibres embedded in a soft non-fibrillar matrix, it is classically considered that its mechanical response is based on an initial alignment of the fibres, followed by the stretching of the aligned fibres. Using a recently developed set-up combining multiphoton microscopy with mechanical assay, we imaged the fibres network evolution during dermis stretching. These observations, combined with a wide set of mechanical tests, allowed us to challenge the classical microstructural interpretation of the mechanical properties of the dermis: we observed a continuous alignment of the collagen fibres along the stretching. All our results can be explained if each fibre contributes by a given stress to the global response. This plastic response is likely due to inner sliding inside each fibre. The non-linear mechanical response is due to structural effects of the fibres network in interaction with the surrounding non-linear matrix. This multiscale interpretation explains our results on genetically-modified mice with a simple alteration of the dermis microstructure. Our previous works have led us to write this year one review article and one chapter of book on multiscale skin biomechanics, to be published next year.

7.4.3. Cornea biomechanics

Participants: Chloé Giraudet, Jean-Marc Allain [correspondant], Patrick Le Tallec.

Cornea is the outer part of the eye. It is a curved transparent organ, which gives 2/3 of the focalisation capacity of the eye. Microscopically, it is made mostly of collagen fibres (as skin) organised in cristal-like lamellae of few micrometers of height and a hundred micrometers in length and width. The lamellae are piled up in a plywood structure, creating a millimetre-thick tissue. Between the lamellae, some cells are present to repair and regenerate the tissue. However, this simple image of the organisation of the collagen is in fact too simple and a more complex heterogeneous organisation has been recently described, with in particular some striae (called the Vogt striae). In C. Giraudet's PhD, we propose to explore the link between microstructure organisation of the collagen in the cornea and mechanical properties. To do so, we will first start by proposing an extension of classical mechanical models (such as Holzapfel's law or others) to the specific case of the cornea. This model will be tested against mechanical assays made under advanced optical microscopes to test first if the model can correctly predict the strain field in volume, and secondly if it correctly predicts the evolution of the lamellae microstructure at different stretch levels. At the moment, we have developed the tools to mechanically test the cornea, but also to build a finite element simulation using the real shape of the cornea we are looking at.

7.4.4. Multiscale properties of the passive cardiac muscle

Participants: Nicole Tueni, Jean-Marc Allain [correspondant], Martin Genet.

We are interested in understanding the effect of the remodelling of cardiac tissues after a disease. Cardiac tissues are mostly made of muscle cells. They can remodel themselves in response to an alteration of their normal response by modifying the sizes and the geometries of the cells in the tissue. Nowadays, we are able to describe the active and passive response of a cardiac tissue, assuming we know the main orientation of the cells inside. However, we do not have models which include explicitly the microstructural cellular organization. Such complex models will be strongly beneficial to determine the consequences of local alterations of the muscle behaviour. In N. Tueni's PhD, we are investigating this multi-scale relationship. To do so, we are imaging the organization at the microscale, while measuring the mechanical properties. These results will be the building block to test and develop mechanical models of the cardiac tissues.

7.4.5. Mechano-perception at the cell level

Participant: Jean-Marc Allain [correspondant].

All cells and organisms experience mechanical forces. Plants along their life are submitted from their environment to long lasting sustained stresses and to recurrent cyclic loading/unloading due to wind or water stream. Mechanical stimulations induce short-term cellular responses, leading to mechanoresponsive gene activation followed by long-term responses permitting structural reinforcement at the whole-plant level. We show that the Mechanosensitive channel Small conductance-Like 10 (MSL10) contributes to oscillation perception at the cell level. This channel responds to pulsed membrane stretching with rapid activation and relaxation. Furthermore, oscillatory pressure stimulation modulates its activity, with increased open probability upon oscillatory than during sustained stimulation. Combined with the adequate localization of MSL10 in plant shoot and leaves, its ability to detect oscillatory deformation at the molecular-scale is relevant for a function of this channel in oscillatory perception in plant.

7.5. Clinical Applications

7.5.1. Exploring kinetic energy as a new marker of cardiac function in the single ventricle circulation

Participants: Radomir Chabiniok [correspondant], Tarique Hussain [ToFMOD].

Ventricular volumetric ejection fraction (VV EF) is often normal in patients with single ventricle circulations despite them experiencing symptoms related to circulatory failure. We sought to determine if kinetic energy (KE) could be a better marker of ventricular performance. KE was prospectively quantified using four-dimensional flow MRI in 41 patients with a single ventricle circulation (aged 0.5-28 yr) and compared with 43 healthy volunteers (aged 1.5-62 yr) and 14 patients with left ventricular (LV) dysfunction (aged 28-79 yr). Intraventricular end-diastolic blood was tracked through systole and divided into ejected and residual blood components. Two ejection fraction (EF) metrics were devised based on the KE of the ejected component over the total of both the ejected and residual components using 1) instantaneous peak KE to assess KE EF or 2) summating individual peak particle energy (PE) to assess PE EF. KE metrics are markers of healthy cardiac function. PE EF may be useful in grading dysfunction. The work was published in Journal of Applied Physiology (J Appl Physiol 125: 889-900, 2018), [29]. The work represents a collaboration with King's College London (J. Wong, K. Pushparajah, R. Razavi) and with UT Southwestern Dallas (T. Hussain, the member of Inria Associate team ToFMOD).

7.5.2. Using a patient-specific biomechanical cardiovascular model to estimate continuously Left Ventricular Pressure Volume Loop: A proof of concept study

Participants: Arthur Le Gall, Fabrice Vallée, Philippe Moireau, Dominique Chapelle, Radomir Chabiniok [correspondant].

Pressure Volume loops (PV loops) could contribute to optimise haemodynamic managements. While the invasiveness of PV loop acquisition prevents it from being routinely used during surgery, cardiovascular modelling could represent an alternative. Using continuous recording of aortic pressure and flow, we aimed at calibrating a patient-specific model and at interpreting the simulated PV loop during administration of noradrenaline (NOR). This study is the first to allow continuous PV loop monitoring during general anaesthesia. The work was pursued in the collaboration with Lariboisiere Hospital in Paris (A. Le Gall and F. Vallée, both dually affiliated at Inria and at AP-HP, "poste d'accueil").

7.5.3. Augmenting the interpretation of cardiac MRI by biomechanical modeling: Application to Tetralogy of Fallot

Participants: Marija Gusseva, Philippe Moireau, Tarique Hussain [ToFMod], Gerald Greil [ToFMod], Animesh Tandon [ToFMOD], Dominique Chapelle, Radomir Chabiniok [correspondant].

The particularity of the mixed-valve disease – pulmonary regurgitation often combined with a stenosis – requested to extend our model-representation of the valve to allow the backflow during the heart relaxation. For each patient, biomechanical models of their left and right ventricles (LV, RV) were set up. These models then allowed to investigate the functional properties of dilated right ventricles (RV) with incompetent pulmonary valves and of the pulmonary circulation, properties not directly visible in the clinical data. In particular, immediately after deploying the new valve we could observe a decrease of RV contractility by 15%, while the output of RV into pulmonary circulation has increased. This suggests a positive immediate outcome, as the energy needs for function of RV will decrease. The higher cardiac output also suggests an increase of the filling of LV (preload), which could contribute to an improvement of LV function. The model also uncovered a decrease of resistance in the pulmonary circulation. This very preliminary result might suggest some pathophysiological changes, which are typically not thought of in clinics.

This work is pursued under the objectives of the Inria Associate Team ToFMod (T. Hussain, G. Greil, A. Tandon are members of ToFMOD and affiliated at UT Southwestern Medical Center Dallas, USA), the work was accepted for a conference of International Society of Magnetic Resonance in Medicine 2018 and is in preparation for publication.

7.5.4. Longitudinal study of ventricular remodeling and reverse-remodeling in tetralogy of Fallot patients using CMR coupled with biomechanical modelling

Participants: Marija Gusseva, Tarique Hussain [ToFMod], Animesh Tandon [ToFMod], Dominique Chapelle, Radomir Chabiniok [correspondant].

A preliminary study was performed with the patient-specific models for RV and pulmonary circulations set up from three datasets including the 6-months post-PVR follow-up exams obtained in late 2018 from King's College London. Clinical data analyses show a positive result of pulmonary replacement therapy (PVR) and normalization of the RV size, i.e. the so-called reverse-remodeling of the pathologically dilated RV, in all three patients. The biomechanical modeling suggests a further reduction of the active stress needed to be developed by RV (contractility), i.e. a long-term unloading of the previously overloaded ventricle.

This work is pursued under the objectives of the Inria Associate Team ToFMOD (T. Hussain, A. Tandon are members of ToFMOD and affiliated at UT Southwestern Medical Center Dallas, USA). The main partner in this task is King's College London ("Other Participant" in the ToFMOD Associate team, K. Pushparajah, M. Jones, S. Qureshi) who provided unique clinical data of patients with a long-term follow-up after PVR. The work was submitted to the conference of International Society of Magnetic Resonance in Medicine 2019 – the world-wide major scientific & clinical event when MR data are involved.

7.5.5. Optical flow-based non-rigid registration of cardiac MR images

Participant: Radomir Chabiniok [correspondant].

This work deals with non-rigid registration of cardiac MR images, particularly the MOLLI sequences. MOLLI sequence consists of 11 heart images acquired over 17 cardiac cycles. The images of MOLLI sequence are used for pixel-wise estimation of T1 relaxation time values. In this case the registration is necessary to correct the deformations that occur because of the patient's imperfect breath-holding during the acquisition. The main characteristics of the MOLLI sequence is the evolving intensity of the tissues and also large variations of the image contrast. This characteristics of the sequence make the registration process challenging and make the use of intensity-based registration method impossible. For this purpose, we propose a method based on optical flow, using information obtained by image segmentation. The first step of the registration process, is segmentation of the regions of interest, using the level set method. The segmented objects are represented by distance maps. The transformation between original images is determined by applying the optical flow method to the distance maps. The registration process is independent of the varying intensity and takes into account only the shape and position of the segmented areas, such as the myocardium or the ventricles. The implementation of the proposed method is described and the method is tested on several MOLLI sequences. The results are compared to the results of methods based on maximisation of mutual information, and the proposed method performs better for the images with significant changes in intensity.

The work represents a collaborative project with Institute for Clinical and Experimental Medicine (IKEM) Prague (J. Tintera) and with Czech Technical University in Prague (K. Solovska, T. Oberhuber).

7.5.6. Quantification of biventricular strains in heart failure with preserved ejection fraction using hyperelastic warping method

Participant: Martin Genet.

Heart failure (HF) imposes a major global health care burden on society and suffering on the individual. About 50% of HF patients have preserved ejection fraction (HFpEF). More intricate and comprehensive measurement-focused imaging of multiple strain components may aid in the diagnosis and elucidation of this disease. Here, we describe the development of a semi-automated hyperelastic warping method for rapid comprehensive assessment of biventricular circumferential, longitudinal, and radial strains that is physiological meaningful and reproducible. We recruited and performed cardiac magnetic resonance (CMR) imaging on 30 subjects [10 HFpEF, 10 HF with reduced ejection fraction patients (HFrEF) and 10 healthy controls]. In each subject, a three-dimensional heart model including left ventricle (LV), right ventricle (RV), and septum was reconstructed from CMR images. The hyperelastic warping method was used to reference the segmented model with the target images and biventricular circumferential, longitudinal, and radial strain–time curves were obtained. The peak systolic strains are then measured and analyzed in this study. The ROC analysis indicated LV peak systolic circumferential strain to be the most sensitive marker for differentiating HFpEF from healthy controls. Our results suggest that the hyperelastic warping method with the CMR-derived strains may reveal subtle impairment in HF biventricular mechanics, in particular despite a "normal" ventricular ejection fraction in HFpEF.

7.5.7. Extra corporeal life support for cardiac arrest patients with post-cardiac arrest syndrome: the ECCAR study

Participant: Arthur Le Gall.

Purpose: Post-Cardiac Arrest Shock (PCAS) occurring after resuscitated cardiac arrest (CA), is a main cause of early death. Extra-Corporeal Life Support (ECLS) could be useful pending recovery of myocardial failure. We aimed to describe our PCAS population, and factors associated with ECLS initiation. Materials and Methods: This analysis included 924 patients admitted in two intensive care units (ICU) between 2005 and 2014 for CA and PCAS, and, of those patients, 43 patients for whom an ECLS was initiated. Neurological and ECLS-related outcomes were gathered retrospectively. Conclusions: ECLS, as a salvage therapy for PCAS, could represent an acceptable alternative for highly selected patients.

7.5.8. Evaluation of cardiac output variations with the peripheral pulse pressure to mean arterial pressure ratio.

Participant: Arthur Le Gall.

Cardiac output (CO) optimisation during surgery reduces post-operative morbidity. Various methods based on pulse pressure analysis have been developed to overcome difficulties to measure accurate CO variations in standard anaesthetic settings. Several of these methods include, among other parameters, the ratio of pulse pressure to mean arterial pressure (PP/MAP). The aim of this study was to evaluate whether the ratio of radial pulse pressure to mean arterial pressure (ΔPPrad/MAP) could track CO variations (Δ CO) induced by various therapeutic interventions such as fluid infusions and vasopressors boluses [phenylephrine (PE), norepinephrine (NA) or ephedrine (EP)] in the operating room. Trans-oesophageal Doppler signal and pressure waveforms were recorded in patients undergoing neurosurgery. CO and PPrad/MAP were recorded before and after fluid challenges, PE, NA and EP bolus infusions as medically required during their anaesthesia. ΔPPrad/MAP tracked ΔCO variations during PE and NA vasopressor challenges. However, after positive fluid challenge or EP boluses, ΔPPrad/MAP was not as performant to track ΔCO which could make the use of this ratio difficult in current clinical practice.

7.5.9. Perioperative management of patients with coronary artery disease undergoing non-cardiac surgery: Summary from the French Society of Anaesthesia and Intensive Care Medicine 2017 convention

Participant: Arthur Le Gall.

This review summarises the specific stakes of preoperative, intraoperative, and postoperative periods of patients with coronary artery disease undergoing non-cardiac surgery. All practitioners involved in the perioperative management of such high cardiac risk patients should be aware of the modern concepts expected to decrease major adverse cardiac events and improve short- and long-term outcomes. A multidisciplinary approach via a functional heart team including anaesthesiologists, cardiologists and surgeons must be encouraged. Rational and algorithm-guided management of those patients should be known and implemented from preoperative to postoperative period.

MAMBA Project-Team

7. New Results

7.1. Modelling Polymerization Processes

Nucleation Phenomena.
A new stochastic model of polymerization including the nucleation has been analyzed in [4]. A Functional Central Limit Theorem for the Becker-Döring model in an infinite dimensional state space is established in [25].

An oscillatory model of polymerisation-depolymerisation.
In 2017, we evidenced the presence of several polymeric species by using data assimilation methods to fit experimental data from H. Rezaei's lab [64]. In collaboration with Klemens Fellner from the university of Graz, we now propose a new model, variant of the Becker-Döring system but containing two monomeric species, capable of displaying sustained though damped oscillations [39].

Time asymptotics for nucleation, growth and division equations.
We revisited the well-known Lifshitz-Slyozov model, which takes into account only polymerisation and depolymerisation, and progressively enriched the model. Taking into account depolymerisation and fragmentation reaction term may surprisingly stabilisde the system, since a steady size-distribution of polymers may then emerge, so that "Ostwald ripening" does not happen [8].

Cell population dynamics and its control
The PhD thesis work of Camille Pouchol (co-supervisors Jean Clairambault, Michèle Sabbah, INSERM, and Emmanuel Trélat, Inria CAGE and LJLL) has been continued, leading after his first article published in the J. Maths Pures Appl. [136], summarised in [31], to his PhD defence in June [1], and to a diversification of his research activities in various directions related to population dynamics and optimal control with Antoine Olivier, Emmanuel Trélat and Enrique Zuazua [51], [56] or to more general questions [55].

Measure solutions for the growth-fragmentation equation
As recalled in the section "Foundations", entropy methods for population dynamics have been successfully developed around B. Perthame and co-authors. We recently extend such methods to the growth-fragmentation equation, in collaboration with P. Gwiazda, E. Wiedemann and T. Debiec [40], using the framework of generalised Young measures.

7.2. Large Stochastic Networks

The equilibrium properties of allocation algorithms for networks with a large number of nodes with finite capacity are investigated in [46] and in [60].

7.3. Control Strategies for Sterile Insect Techniques

We proposed different models to serve as a basis for the design of control strategies relying on releases of sterile male mosquitoes (*Aedes spp*) and aiming at elimination of wild vector population. Different types of releases were considered (constant, periodic or impulsive) and sufficient conditions to reach elimination were provided in each case [57], [3], [35]. We also estimated sufficient and minimal treatment times. A feedback approach was introduced, in which the impulse amplitude is chosen as a function of the actual wild population [57], [3], [35].

7.4. Optimal replacement strategies, application to Wolbachia

We modelled and designed optimal release control strategy with the help of a least square problem. In a nutshell, one wants to minimize the number of uninfected mosquitoes at a given time horizon, under relevant biological constraints. We derived properties of optimal controls and studied a limit problem providing useful asymptotic properties of optimal controls [49], [3].

7.5. Oscillatory regimes in population models

Understanding mosquitoes life cycle is of great interest presently because of the increasing impact of vector borne diseases. Observations yields evidence of oscillations in these populations independent of seasonality, still unexplained. We proposed [58], [3] a simple mathematical model of egg hatching enhancement by larvae which produces such oscillations that conveys a possible explanation.

On the other hand, population oscillations may be induced by seasonal changes. We considered a biological population whose environment varies periodically in time, exhibiting two very different "seasons", favorable and unfavorable. We addressed the following question: the system's period being fixed, under what conditions does there exist a critical duration above which the population cannot sustain and extincts, and below which the system converges to a unique periodic and positive solution? We obtained [59], [3] sufficient conditions for such a property to occur for monotone differential models with concave nonlinearities, and applied the obtained criterion to a two-dimensional model featuring juvenile and adult insect populations.

7.6. Feedback control principles for population replacement by Wolbachia

The issue of effective scheduling of the releases of *Wolbachia*-infected mosquitoes is an interesting problem for Control theory. Having in mind the important uncertainties present in the dynamics of the two populations in interaction, we attempted to identify general ideas for building release strategies, which should apply to several models and situations [34]. These principles were exemplified by two interval observer-based feedback control laws whose stabilizing properties were demonstrated when applied to a model retrieved from [76].

7.7. Bacterial motion by run and tumble

Collective motion of chemotactic bacteria such as Escherichia coli relies, at the individual level, on a continuous reorientation by runs and tumbles. It has been established that the length of run is decided by a stiff response to a temporal sensing of chemical cues along the pathway. We describe in [21] a novel mechanism for pattern formation stemming from the stiffness of chemotactic response relying on a kinetic chemotaxis model which includes a recently discovered formalism for the bacterial chemotaxis. We prove instability both for a microscopic description in the space-velocity space and for the macroscopic equation, a flux-limited Keller-Segel equation, which has attracted much attention recently. A remarkable property is that the unstable frequencies remain bounded, as it is the case in Turing instability. Numerical illustrations based on a powerful Monte Carlo method show that the stationary homogeneous state of population density is destabilized and periodic patterns are generated in realistic ranges of parameters. These theoretical developments are in accordance with several biological observations.

This motivates also our study of traveling wave and aggregation in population dynamics of chemotactic cells based on the FLKS model with a population growth term [7]. Our study includes both numerical and theoretical contributions. In the numerical part, we uncover a variety of solution types in the one-dimensional FLKS model additionally to standard Fisher/KPP type traveling wave. The remarkable result is a counter-intuitive backward traveling wave, where the population density initially saturated in a stable state transits toward an un- stable state in the local population dynamics. Unexpectedly, we also find that the backward traveling wave solution transits to a localized spiky solution as increasing the stiffness of chemotactic response. In the theoretical part, we obtain a novel analytic formula for the minimum traveling speed which includes the counter-balancing effect of chemotactic drift vs. reproduction/diffusion in the propagating front. The front propagation speeds of numerical results only slightly deviate from the minimum traveling speeds, except for the localized spiky solutions, even for the backward traveling waves. We also discover an analytic solution of unimodal traveling wave in the large-stiffness limit, which is certainly unstable but exists in a certain range of parameters.

7.8. Numerical methods for cell aggregation by chemotaxis

Three-dimensional cultures of cells are gaining popularity as an in vitro improvement over 2D Petri dishes. In many such experiments, cells have been found to organize in aggregates. We present new results of three-dimensional in vitro cultures of breast cancer cells exhibiting patterns. Understanding their formation is of particular interest in the context of cancer since metastases have been shown to be created by cells moving in clusters. In the paper [37], we propose that the main mechanism which leads to the emergence of patterns is chemotaxis, i.e., oriented movement of cells towards high concentration zones of a signal emitted by the cells themselves. Studying a Keller-Segel PDE system to model chemotactical auto-organization of cells, we prove that it is subject to Turing instability if a time-dependent condition holds. This result is illustrated by two-dimensional simulations of the model showing spheroidal patterns. They are qualitatively compared to the biological results and their variability is discussed both theoretically and numerically.

This motivates to study parabolic-elliptic Keller-Segel equation with sensitivity saturation, because of its pattern formation ability, is a challenge for numerical simulations. We provide in [16] two finite-volume schemes that are shown to preserve, at the discrete level, the fundamental properties of the solutions, namely energy dissipation, steady states, positivity and conservation of total mass. These requirements happen to be critical when it comes to distinguishing between discrete steady states, Turing unstable transient states, numerical artifacts or approximate steady states as obtained by a simple upwind approach. These schemes are obtained either by following closely the gradient flow structure or by a proper exponential rewriting inspired by the Scharfetter-Gummel discretization. An interesting fact is that upwind is also necessary for all the expected properties to be preserved at the semi-discrete level. These schemes are extended to the fully discrete level and this leads us to tune precisely the terms according to explicit or implicit discretizations. Using some appropriate monotonicity properties (reminiscent of the maximum principle), we prove well-posedness for the scheme as well as all the other requirements. Numerical implementations and simulations illustrate the respective advantages of the three methods we compare.

7.9. Focus on cancer

Modelling Acute Myeloid Leukaemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations

This theme has continued to be developed in collaboration with Catherine Bonnet, Inria DISCO (Saclay) [12], [29]. Without control by drugs, but with representation of mutualistic interactions between tumor cells and their surrounding support stroll cells, it has also, in collaboration with Delphine Salort and Thierry Jaffredo (LCQB-IBPS) given rise to a recent work by Thanh Nam Nguyen, hired as HTE and ERC postdoctoral fellow at LCQB, submitted as full article [50].

Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control

The research topic “Evolution and cancer”, designed in the framework of adaptive dynamics to represent and overcome acquired drug resistance in cancer, initiated in [119], [118] and later continued in [90], [89], [117], has been recently summarised in [31] and has been the object of the PhD thesis work of Camille Pouchol, see above “Cell population dynamics and its control” . It is now oriented, thanks to work underway by Cécile Carrère, Jean Clairambault, Tommaso Lorenzi and Grégoire Nadin, in particular towards the mathematical representation of *bet hedging* in cancer, namely a supposed optimal strategy consisting for cancer cell populations under life-threatening cell stress in diversifying their phenotypes according to several resistance mechanisms, such as overexpression of ABC transporters (P-glycoprotein and many others), of DNA repair enzymes or of intracellular detoxication processes. According to different deadly insults the cancer cell population is exposed to, some phenotypes may be selected, any such successful subpopulation being able to store the cell population genome (or subclones of it if the cell population is already genetically heterogeneous) and make it amenable to survival and renewed replication.

Philosophy of cancer biology

This new research topic in Mamba, dedicated to explore possibly underinvestigated, from the mathematical modelling point of view, parts of the field of cancer growth, evolution and therapy, has been the object of a presentation by Jean Clairambault at the recent workshop "Philosophy of cancer biology' (https://www.philinbiomed.org/event/philosophy-of-cancer-biology-workshop/). This workshop gathered most members worldwide of this small, but very active in publishing, community of philosophers of science whose field of research is "philosophy of cancer", as they call it themselves. This topic offers a clear point of convergence between mathematics, biology and social and human sciences.

7.10. Deformable Cell Modeling: biomechanics and Liver regeneration

- Biomechanically mediated growth control of cancer cells The key intriguing novelty was that the same agent-based model after a single parameter has been calibrated with growth data for multicellular spheroids without application of external mechanical stress by adapting a single parameter, permitted to correctly predict the growth speed of multicellular spheroids of 5 different cell lines subject of external mechanical stress. Hereby the same mechanical growth control stress function was used without any modification [44]. The prediction turned out to be correct independent of the experimental method used to exert the stress, whereby once a mechanical capsule has been used, once dextran has been used in the experiments.
- Regeneration of liver with the Deformable Cell Model. The key novelty was the implementation of the model itself, but an interesting novel result is that the DCM permits closure of a pericentral liver lobule lesion generated by drug-induced damage with about 5 times smaller active migration force due to the ability of the cell to strongly deform and squeeze into narrow spaces between the capillaries. This finding stresses that a precise mechanical description is important in view of quantitatively correct modeling results [142]. The deformable cell model however could be used to calibrate the interaction forces of the computationally much cheaper center-based model to arrive at almost the same results.

MONC Project-Team

7. New Results

7.1. Mathematical Modeling of the Proliferation Gradient in MultiCellular Tumor Spheroids

Authors: *Thomas Michel*, J. Fehrenbach, V. Lobjois, J. Laurent, A. Gomes, *Thierry Colin, Clair Poignard*. Paper published in the Journal of Theoretical Biology. https://hal.inria.fr/hal-01883189

MultiCellular Tumor Spheroids are 3D cell cultures that can accurately reproduce the behavior of solid tumors. It has been experimentally observed that large spheroids exhibit a decreasing gradient of proliferation from the periphery to the center of these multicellular 3D models: the proportion of proliferating cells is higher in the periphery while the non-proliferating quiescent cells increase in depth. In this paper, we propose to investigate the key mechanisms involved in the establishment of this gradient with a Partial Differential Equations model that mimics the experimental setup of growing spheroids under different nutrients supply conditions. The model consists of mass balance equations on the two cell populations observed in the data: the proliferating cells and the quiescent cells. The spherical symmetry is used to rewrite the model in radial and relative coordinates. Thanks to a rigorous data postprocessing the model is then fit and compared quantitatively with the experimental quantification of the percentage of proliferating cells from EdU immunodetection on 2D spheroid cryosection images. The results of this calibration show that the proliferation gradient observed in spheroids can be quantitatively reproduced by our model.

7.2. Viscoelastic modeling of the fusion of multicellular tumor spheroids in growth phase

Authors: *Guillaume Dechristé*, Jérôme Fehrenbach, Elena Griseti, Valérie Lobjois, *Clair Poignard*. Paper published in the Journal of Theoretical Biology. https://hal.inria.fr/hal-01786027

Background. Since several decades, the experiments have highlighted the analogy of fusing cell aggregates with liquid droplets. The physical macroscopic models have been derived under incompressible assumptions. The aim of this paper is to provide a 3D model of growing spheroids, which is more relevant regarding embryo cell aggregates or tumor cell spheroids. Methods. We extend the past approach to a compressible 3D framework in order to account for the tumor spheroid growth. We exhibit the crucial importance of the effective surface tension, and of the inner pressure of the spheroid to describe precisely the fusion. The experimental data were obtained on spheroids of colon carcinoma human cells (HCT116 cell line). After 3 or 6 days of culture, two identical spheroids were transferred in one well and their fusion was monitored by live videomicroscopy acquisition each 2hours during 72h. From these images the neck radius and the diameter of the assembly of the fusing spheroids are extracted. Results. The numerical model is fitted with the experiments. It is worth noting that the time evolution of both neck radius and spheroid diameter are quantitatively obtained. The interesting feature lies in the fact that such measurements characterise the macroscopic rheological properties of the tumor spheroids. Conclusions. The experimental determination of the kinetics of neck radius and overall diameter during spheroids fusion characterises the rheological properties of the spheroids. The consistency of the model is shown by fitting the model with two different experiments, enhancing the importance of both surface tension and cell proliferation. General Significance. The paper sheds new light on the macroscopic rheological properties of tumor spheroids. It emphasizes the role of the surface tension and the inner pressure in the fusion of growing spheroid. Under geometrical assumptions, the model reduces to a 2–parameter differential equation fit with experimental measurements. The 3–D partial differential system makes it possible to study the fusion of spheroids in non-symmetrical or more general frameworks.

7.3. Mathematical analysis and 2-scale convergence of a heterogeneous microscopic bidomain model

Authors: *Annabelle Collin*, Sébastien Imperiale. Paper published in Mathematical Models and Methods in Applied Sciences. https://hal.inria.fr/hal-01759914

The aim of this paper is to provide a complete mathematical analysis of the periodic homogenization procedure that leads to the macroscopic bidomain model in cardiac elec-trophysiology. We consider space-dependent and tensorial electric conductivities as well as space-dependent physiological and phenomenological non-linear ionic models. We provide the nondimensionalization of the bidomain equations and derive uniform estimates of the solutions. The homogenization procedure is done using 2-scale convergence theory which enables us to study the behavior of the non-linear ionic models in the homogenization process.

7.4. Pre-treatment magnetic resonance-based texture features as potential imaging biomarkers for predicting event free survival in anal cancer treated by chemoradiotherapy

Authors: Arnaud Hocquelet, Thibaut Auriac, *Cynthia Perier*, Clarisse Dromain, Marie Meyer, Jean-Baptiste Pinaquy, Alban Denys, Hervé Trillaud, *Baudouin Denis de Senneville*, Véronique Vendrely. Paper published in European Radiology. https://hal.archives-ouvertes.fr/hal-01962472

AIM: To assess regular MRI findings and tumour texture features on pre-CRT imaging as potential predictive factors of event-free survival (disease progression or death) after chemoradiotherapy (CRT) for anal squamous cell carcinoma (ASCC) without metastasis.
MATERIALS AND METHODS: We retrospectively included 28 patients treated by CRT for pathologically proven ASCC with a pre-CRT MRI. Texture analysis was carried out with axial T2W images by delineating a 3D region of interest around the entire tumour volume. First-order analysis by quantification of the histogram was carried out. Second-order statistical texture features were derived from the calculation of the grey-level co-occurrence matrix using a distance of 1 (d1), 2 (d2) and 5 (d5) pixels. Prognostic factors were assessed by Cox regression and performance of the model by the Harrell C-index.
RESULTS: Eight tumour progressions led to six tumour-specific deaths. After adjusting for age, gender and tumour grade, skewness (HR = 0.131, 95% CI = 0-0.447, p = 0.005) and cluster shade_d1 (HR = 0.601, 95% CI = 0-0.861, p = 0.027) were associated with event occurrence. The corresponding Harrell C-indices were 0.846, 95% CI = 0.697-0.993, and 0.851, 95% CI = 0.708-0.994.
CONCLUSION: ASCC MR texture analysis provides prognostic factors of event occurrence and requires additional studies to assess its potential in an "individual dose" strategy for ASCC chemoradiation therapy.
KEY POINTS: MR texture features help to identify tumours with high progression risk. Texture feature maps help to identify intra-tumoral heterogeneity. Texture features are a better prognostic factor than regular MR findings.
KEYWORDS: Anal squamous cell carcinoma; Definitive chemoradiotherapy; Imaging biomarkers; Magnetic resonance imaging; Texture analysis

7.5. T2-based MRI Delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy

Authors: *Amandine Crombé, MS Cynthia Périer*, Michèle Kind, *Baudouin Denis De Senneville*, François Le Loarer, Antoine Italiano, Xavier Buy, *Olivier Saut*. Paper published in the Journal of Magnetic Resonance Imaging. https://hal.inria.fr/hal-01929807

Background: Standard of care for patients with high-grade soft-tissue sarcoma (STS) are being redefined since neoadjuvant chemotherapy (NAC) has demonstrated a positive effect on patients' outcome. Yet response evaluation in clinical trials still relies on RECIST criteria.

Purpose: To investigate the added value of a Delta-radiomics approach for early response prediction in patients with STS undergoing NAC.
Study Type: Retrospective.
Population: Sixty-five adult patients with newly-diagnosed, locally-advanced, histologically proven high-grade STS of trunk and extremities. All were treated by anthracycline-based NAC followed by surgery and had available MRI at baseline and after two chemotherapy cycles.
Field Strength/Sequence: Pre- and postcontrast enhanced T1-weighted imaging (T1-WI), turbo spin echo T2-WI at 1.5 T.
Assessment: A threshold of $< 10\%$ viable cells on surgical specimens defined good response (Good-HR). Two senior radiologists performed a semantic analysis of the MRI. After 3D manual segmentation of tumors at baseline and early evaluation, and standardization of voxel-sizes and intensities, absolute changes in 33 texture and shape features were calculated.
Statistical Tests: Classification models based on logistic regression, support vector machine, k-nearest neighbors, and random forests were elaborated using crossvalidation (training and validation) on 50 patients ("training cohort") and was validated on 15 other patients ("test cohort").
Results: Sixteen patients were good-HR. Neither RECIST status (P = 0.112) nor semantic radiological variables were associated with response (range of P-values: 0.134–0.490) except an edema decrease (P = 0.003), although 14 shape and texture features were (range of P-values: 0.002–0.037). On the training cohort, the highest diagnostic performances were obtained with random forests built on three features: Δ_Histogram_Entropy, Δ_Elongation, Δ_Surrounding_Edema, which provided: area under the curve the receiver operating characteristic = 0.86, accuracy = 88.1%, sensitivity = 94.1%, and specificity = 66.3%. On the test cohort, this model provided an accuracy of 74.6% but 3/5 good-HR were systematically ill-classified.
Data Conclusion: A T2-based Delta-radiomics approach might improve early response assessment in STS patients with a limited number of features.

7.6. Revisiting bevacizumab + cytotoxics scheduling using mathematical modeling: proof of concept study in experimental non-small cell lung carcinoma

Authors: Diane-Charlotte Imbs, Raouf El Cheikh, Arnaud Boyer, Joseph Ciccolini, Celine Mascaux, Bruno Lacarelle, Fabrice Barlesi, Dominique Barbolosi, *Sébastien Benzekry*. Paper published in Clinical Pharmacology and Therapeutics: Pharmacometrics and Systems Pharmacology. https://hal.inria.fr/hal-01624423v2

Concomitant administration of bevacizumab and pemetrexed-cisplatin is a common treatment for advanced non-squamous non-small cell lung cancer (NSCLC). Vascular normalization following bevacizumab administration may transiently enhance drug delivery, suggesting improved efficacy with sequential administration. To investigate optimal scheduling, we conducted a study in NSCLC-bearing mice using. First, experiments demonstrated improved efficacy when using sequential versus concomitant scheduling of bevacizumab and chemotherapy. Using a mathematical model of tumor growth under therapy accounting for the normalization effect, we predicted an optimal delay of 2.8 days between bevacizumab and chemotherapy. This prediction was confirmed experimentally, with reduced tumor growth of 38% as compared to concomitant scheduling, and prolonged survival (70 vs. 74 days). Alternate sequencing of 8 days failed in achieving similar increase in efficacy, thus emphasizing the utility of modeling support to identify optimal scheduling. The model could also be a useful tool in the clinic to personally tailor regimen sequences.

NUMED Project-Team (section vide)

REO Project-Team

7. New Results

7.1. Mathematical and numerical analysis of fluid-structure interaction problems

Participants: Muriel Boulakia, Ludovic Boilevin-Kayl, Chen-Yu Chiang, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Damiano Lombardi, Marc Thiriet, Marina Vidrascu.

In [31], we consider a system modeling the interaction between a viscous incompressible fluid and an elastic structure. The fluid motion is represented by the classical Navier-Stokes equations while the elastic displacement is described by the linearized elasticity equation. The elastic structure is immersed in the fluid and the whole system is confined into a bounded domain of dimension 3. Our main result is the local in time existence and uniqueness of a strong solution of the corresponding system. This result holds without any restrictive assumptions on the domains geometry.

The numerical simulation of a thin-walled structure immersed in an incompressible fluid can be addressed by various methods. In [16], three of them are considered: the Arbitrary Lagrangian-Eulerian (ALE) method, the Fictitious Domain/Lagrange multipliers (FD) method and the Nitsche-XFEM method. Taking ALE as a reference, the advantages and limitations of FD and Nitsche-XFEM are carefully discussed on three benchmark test cases which have been chosen to be representative of typical difficulties encountered in valves or living cells simulations.

Fictitious domain approximations of fluid-structure interaction problems are generally discretized in time using strongly coupled schemes. This guarantees unconditional stability but at the price of solving a computationally demanding coupled system at each time-step. The design of loosely coupled schemes (i.e., methods that invoke the fluid and solid solvers only once per time-step) is of fundamental interest, especially for three-dimensional simulations, but the existing approaches are known to suffer from severe stability and/or time accuracy issues. In [28], we propose a new approach that overcomes these difficulties in the case of the coupling with immersed thin-walled structures.

In [27], we derive a Nitsche-based formulation for fluid-structure interaction (FSI) problems with contact. The approach is based on the work of Chouly and Hild [SIAM Journal on Numerical Analysis. 2013;51(2):1295-1307] for contact problems in solid mechanics. We present two numerical approaches, both of them formulating the FSI interface and the contact conditions simultaneously in equation form on a joint interface-contact surface. The first approach uses a relaxation of the contact conditions to allow for a small mesh-dependent gap between solid and wall. The second alternative introduces an artificial fluid below the contact surface. The resulting systems of equations can be included in a consistent fashion within a monolithic variational formulation, which prevents the so-called "chattering" phenomenon. To deal with the topology changes in the fluid domain at the time of impact, we use a fully Eulerian approach for the FSI problem. We compare the effect of slip and no-slip interface conditions and study the performance of the method by means of numerical examples.

7.2. Numerical methods for biological flows

Participants: Ludovic Boilevin-Kayl, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Florian Joly, Alexandre This, Marc Thiriet, Irene Vignon Clementel.

Cirrhosis is the common end-stage of chronic liver disease, with architectural distortion increasing the intrahepatic vascular resistance, leading to portal hypertension and systemic circulatory disorders. In [13] we investigate the impact of the changing vascular resistances on the hepatic and global circulation hemodynamics during cirrhogenesis. Morphological quantification of vascular trees from corosion casts of rats developing the disease provide the input for a lumped parameter model of the liver that was coupled to a model of the entire circulation of the rat. The simulations explain how vascular changes due to cirrhosis severely disrupt both hepatic and global hemodynamics.

Image-based models derived from CT angiography are being used clinically to simulate blood flow in the coronary arteries of individual patients to aid in the diagnosis of disease and planning treatments. However, image resolution limits vessel segmentation to larger epicardial arteries. In [20], we propose an algorithm for the generation of a patient-specific cardiac vascular network from epicardial vessels down to arterioles. We extend a tree generation method based on satisfaction of functional principles, to account for competing vascular trees, with flow-related and geometrical constraints adapting the simultaneous tree growths to patient priors.

Growth and remodeling of the embryo pharyngeal arch artery (PAA) network into the extracardiac great vessels is poorly understood but a major source of clinically serious malformations. In [21] we develop a methodological pipeline from high-resolution nano-computed tomography imaging and live-imaging flow measurements to multiscale pulsatile computational models. We identify local morphological variation along the PAAs and their association with specific hemodynamic changes in embryos of different stages, advancing our understanding of morphogenesis.

In [22] we evaluate atrioventricular valve regurgitation (AVVR) in babies born with an already very challenging heart condition, i.e., with single ventricle physiology. Although the second surgery that single ventricle patients undergo is thought to decrease AVVR, there is much controversy in the clinical literature about AVVR treatment. The effect of AVVR on Stage 1 haemodynamics and resulting acute changes from conversion to Stage 2 circulation in single ventricle patients are analyzed through lumped parameter models. Several degrees of AVVR severity are analyzed, for two types of valve regurgitation: incomplete leaflet closure and valve prolapse.

The medical imaging community is eager to define quantitative biophysical parameters. As part of a book addressing this question, in [26], we give a short overview of the mathematical modeling of blood flow at different resolutions, from the large vessel scale (three-dimensional, one-dimensional, and zero-dimensional modeling) to microcirculation and tissue perfusion.

In order to reduce the complexity of heart hemodynamics simulations, uncoupling approaches are often considered for the modeling of the immersed valves as an alternative to complex fluid-structure interaction (FSI) models. A possible shortcoming of these simplified approaches is the difficulty to correctly capture the pressure dynamics during the isovolumetric phases. In [35], we propose an enhanced resistive immersed surfaces (RIS) model of cardiac valves which overcomes this issue. The benefits of the model are investigated and tested in blood flow simulations of the left heart.

7.3. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi, Fabien Raphel.

In [19] a method to assess the variability of phenomena described by PDEs is proposed. In particular, the probability density distribution of the parameters of a model is estimated, in such a way that the statistics of the model output match the observed ones. The investigated approach is based on a differential entropy regularised moment matching.

In [25] we investigated how, by a semi-empirical design of composite biomarkers, the classification of the action of a drug on the electrical activity of a cell can be improved. The data used are measured with a Micro-Electrodes-Array.

In [33] a method is investigated, to design composite biomarkers by exploiting a database of in silico experiments. In particular, a dictionary approach is proposed. The composite biomarker is expressed as a linear combination of linear and non-linear forms applied to the observable. The coefficients of the combination are determined by solving a ℓ^1 regularised optimisation problem.

7.4. Lung and respiration modeling

Participants: Laurent Boudin, Céline Grandmont, Marina Vidrascu, Marc Thiriet, Irene Vignon Clementel.

In [34] we analyse multiscale models arising in the description of physiological flows such as blood flow in arteries or air flow in the bronchial tree. The fluid in the 3D part is described by the Stokes or the Navier-Stokes system which is coupled to 0D models or so-called Windkessel models. The resulting Navier-Stokes-Windkessel coupled system involves Neumann non-local boundary conditions that depends on the considered applications. We first show that the different types of Windkessel models share a similar formalism. Next we derive stability estimates for the continuous coupled Stokes-Windkessel or Navier-Stokes-Windkessel problem as well as stability estimates for the semi-discretized systems with either implicit or explicit coupling.We exhibit different kinds of behavior depending on the considered 0D model. Moreover even if no energy estimates can be derived in energy norms for the Navier-Stokes-Windkessel system, leading to possible numerical instabilities for large applied pressures, we show that stability estimates for both the continuous and semi-discrete problems, can be obtained in appropriate norms for small enough data by introducing a new well chosen Stokes-like operator. These sufficient stability conditions on the data may give a hint on the order of magnitude of the data enabling stable computations without stabilization method for the problem.

In [17], we consider a multi-species kinetic model which leads to the Maxwell-Stefan equations under a standard diffusive scaling (small Knudsen and Mach numbers). We propose a suitable numerical scheme which approximates both the solution of the kinetic model in rarefied regime and the one in the diffusion limit. We prove some a priori estimates (mass conservation and nonnegativity) and well-posedness of the discrete problem. We also present numerical examples where we observe the asymptotic-preserving behavior of the scheme.

In [30], we are interested in a system of fluid equations for mixtures with a stiff relaxation term of Maxwell-Stefan diffusion type. We use the formalism developed by Chen, Levermore, Liu to obtain a limit system of Fick type where the species velocities tend to align to a bulk velocity when the relaxation parameter remains small.

In [29], we consider the Boltzmann operator for mixtures with cutoff Maxwellian, hard potentials, or hard spheres collision kernels. In a perturbative regime around the global Maxwellian equilibrium, the linearized Boltzmann multi-species operator L is known to possess an explicit spectral gap, in the global equilibrium weighted L^2 space. We study a new operator L_ε obtained by linearizing the Boltzmann operator for mixtures around local Maxwellian distributions, where all the species evolve with different small macroscopic velocities of order $\varepsilon > 0$. This is a non-equilibrium state for the mixture. We establish a quasi-stability property for the Dirichlet form of L_ε in the global equilibrium weighted L^2 space. More precisely, we consider the explicit upper bound that has been proved for the entropy production functional associated to L and we show that the same estimate holds for the entropy production functional associated to L_ε, up to a correction of order ε.

7.5. Miscellaneous

Participants: Damiano Lombardi, Irene Vignon Clementel.

In [32] numerical quadrature schemes for the integration of observable quantities in the Brillouin zone for the periodic Schrödinger operator are investigated.

The indocyanine green (ICG) clearance, presented as plasma disappearance rate is, presently, a reliable method to estimate the hepatic function. However, this technique is not instantaneously available and thus cannot been used intra-operatively (during liver surgery). Near-infrared spectroscopy enables to assess hepatic ICG concentration over time in the liver tissue. In [14], we propose to extract more information from the liver intensity dynamics by interpreting it through a dedicated pharmacokinetics model. Parameters for different liver states are estimated from in-vivo measurements in rabbits (El-Desoky et al. 1999), and their link with liver function is investigated.

The hepatic hemodynamics is an essential parameter in surgical planning as well as in various disease processes. The transit time ultrasound (TTUS) perivascular flow probe technology is widely used in clinical practice to evaluate the hepatic inflow, yet invasive. The phase-contrast-MRI (PC-MRI) is not invasive and potentially applicable in assessing the hepatic blood flow. In [15], we compare the hepatic inflow rates using the PC-MRI and the TTUS probe, and evaluated their predictive value of post-hepatectomy adverse events in a porcine experimental model of partial hepatectomy.

SISTM Project-Team

7. New Results

7.1. Mecanistic modeling

7.1.1. Methodology

Estimation methods in mechanistic models can be seen as an inverse problem, in which we want to recover the individual parameters that have produced some observations. In collaboration with the Inria MONC & M3DISIM team, we propose a method for estimation in ODE with mixed effects on parameters based on Kalman based filter (also known as linear quadratic estimation (LQE)) that consist in correcting the original dynamic at each time by a feedback control.

7.1.2. Applications

7.1.2.1. Applied to IL-7 therapy

We have developed two approaches to optimize the injection of IL-7 in HIV-infected patients: one based on a statistical model and one based on Piecewise Deterministic Markov Processes (PDMP).

Villain L, Commenges D, Pasin C, Prague M, Thiébaut R. Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts. Statistics in Medicine. 2018;38:221-235.

Pasin C, Dufour F, Villain L, Zhang H, Thiébaut R. Controlling IL-7 Injections in HIV-Infected Patients. Bulletin of Mathematical Biology. 2018;80:2349-2377.

7.1.2.2. Applied to development of new HIV immunotherapies and vaccines

HIV infection can be treated but not cured with combination antiretroviral therapy, and new therapies that instead target the host immune response to infection are now being developed. Two recent studies of such immunotherapies, conducted in an animal model (SIV-infected rhesus macaques), have shown that agents which target the innate immune receptor TLR7 along with recombining viral-vector vaccines can prevent or control the rebound in viremia that usually accompanies the discontinuation of antiretroviral drugs. However, the mechanism of action of these therapies remains unknown. In collaboration with Harvard School of public health and Harvard program for evolutionary dynamics, we delineate the best model and the best procedure for treatment effect selection in order to delineate the each of each immunotherapies and design subsequent trials.

7.1.2.3. Applied to development of Ebola vaccines

We have analysed data from three clinical trials conducted under the EBOVAC1 consortium in 4 different countries (UK, Kenya and Uganda/Tanzania) and assessing the safety and immunogenicity of prime-boost vaccine regimens combining one adenovirus-based vector and one modified vaccine Ankara. In particular, we have modeled the dynamics of the humoral immune response following the boost immunization and fitted both linear mixed models and ODEs-based mechanistic models to the antibody concentrations data. This analysis allowed the estimation of the durability of the antibody response, as well as the identification of factors of variability of the response.

Pasin C, Balelli I, Van Effelterre T, Bockstal V, Soloforosi L, Prague M, Douoguih M, Thiébaut R. Dynamics of the humoral immune response to a prime-boost Ebola vaccine: quantification and sources of variation. Under revision in Journal of Virology.

7.2. Statistical learning methods for high-dimensional data

7.2.1. Automatic analysis of cell populations

We have developed two different approaches to classify the cell populations according to the high dimensional data obtained with flow cytometry assays. The first one is based on an interesting development of Dirichlet processes and the second one is based on a simple tree classification providing very high performances:

Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaut R. Sequential Dirichlet process mixtures of multivariate skew t-distributions for model-based clustering of flow cytometry data. Annals of Applied Statistics. In press.

Commenges D, Alkhassim C, Gottardo R, Hejblum B, Thiébaut R. cytometree: A binary tree algorithm for automatic gating in cytometry analysis. Cytometry A. 2018;93:1132-1140.

7.2.2. Missing Value Treatment in Longitudinal High Dimensional Supervised Problems

Poor blood sample quality introduces a large number of missing values in the context of sequencing data production. Furthermore, strong technical biases may force the analyst to remove the considered sequenced samples. Then entire day dependent data are then missing.

We have developed a regularized SVD based method using the temporal structure (through multi-block approach) of the missing values to estimate missing values with the objective of predicting uni-variate or multivariate regression responses but also classification problems. That regularizing method uses soft-thresholding on the co-variance matrices implying natural variable selection of covariate and response through a single hyper-parameter to be tuned.

7.2.3. Left-censored data treatment in High Dimensional Supervised Problems

Data could be censored either by the limit of detection or the limit of quantification. We have developed a regularized method for handling high-dimensional exposure data in the presence of censored values in the field of HIV that could be applied to other fields.

Soret P, Avalos M, Wittkop L, Commenges D, Thiébaut R. Lasso regularization for left-censored Gaussian outcome and high-dimensional predictors. BMC Med Res Methodol. 2018 Dec 4;18(1):159.

7.3. Analysis of results from Clinical trials and cohorts

7.3.1. In the HIV field

We have performed the statistical analyses of the immunogenicity endpoints, including high dimensional assays such as gene expression (RNA Seq), of two HIV vaccine clinical trials: 1) ANRS VRI01, a randomized phase I/II trial evaluating for different prime boost vaccine strategies in healthy volunteers; 2) ANRS 159 LIGHT, a randomized phase II trial comparing a prime-boost therapeutic HIV vaccine strategy to placebo in HIV-infected patients undergoing antiretroviral treatment interruption. The results of each of these two trials have been presented as an oral presentation at the HIV R4P conference in Madrid in October 2018 (Richert L et al. and Lacabaratz C et al). Integrative statistical analyses using sPLS methods (as developed by the team) are currently ongoing to relate markers from different high-dimensional immunogenicity assays or virological assays to each other.

7.3.2. In the Ebola field

We have performed a review of all existing clinical trials available to evaluate Ebola vaccines in macaques and humans.

Gross L, Lhomme E, Pasin C, Richert L, Thiébaut R. Ebola vaccine development: Systematic review of pre-clinical and clinical studies, and meta-analysis of determinants of antibody response variability after vaccination. International Journal of Infectious Diseases. 2018. pii: S1201-9712(18)34457-6.

7.3.3. In cluster randomized trials

Accounting for missing outcome is highly important to recover unbiased results of treatment effects. Weighting approached are less common compared to multilevel multiple imputation to analyse clustered data with missing outcome. >In collaboration with Duke University, we compared the two approaches and evaluated their performances to conclude that weighted appraoch should be considered a viable strategy to account for missing outcomes in cluster randomized trials

In cluster randomized trials, it is often desirable to improve the understanding of intervention effects in the presence of dissemination/spillover. In collaboration with Rhode Island university and Yale University, we aims at proposing innovative approach to analyze the TasP ANRS 12249 trial.We proposed innovative methods to individual, disseminated and overal effect of a clustered intervention based on counterfactuals averages in the presence of dissemination.

7.4. Conferences

Members of the team were involved in several talks during conferences and colloquium.

- Big Data and Information Analytics 2018 BigDIA Conference – 17-19th December 2018, Houston, Texas, USA – Random Forests for high-dimensional longitudinal data [invited talk]
- Big Data and Information Analytics 2018 BigDIA Conference – 17-19th December 2018, Houston, Texas, USA – Analysis of high-dimensional longitudinal data from the French health-administrative databases using machine learning methods: performance comparison between LSTM neural networks and Lasso for the analysis of the risk of road traffic crashes associated with medicinal drug consumption [invited talk]
- Journées Recherche et Santé (JRS) Inserm Phénotypage Clinique et biologie des systèmes – 22 Novembre 2018, Institut Imagine Paris, France - Visualisation des omics et des données cliniques [invited talk]
- Center for Modelling and Simulation in the Biosciences (BIOMS) Symposium 2018 – 1-2nd Octobre 2018, Heidelberg, Germany - Finding the cells in the middle of the data [invited talk]
- Mc Gill University, Departement of Epidemiology Seminar – 24th September 2018, Montreal, Canada - Big Data In Vaccine Clinical Trials: A Dive Into Data Science [invited talk]
- 4th Neurepiomics summer school - 17-20th September 2018, Arcachon, France - Systems biology approaches applied to omics data [invited talk]
- International Workshop in honor of Daniel Commenges' 70th birthday- June 4-5, 2018, Bordeaux, France - The mechanistic model point of view of causality [invited talk]
- Population Approach Group Europe (PAGE meeting) - May 29 - June 1, 2018, Montreux, Switzerland - Use of mathematical modeling for optimizing and adapting immunotherapy protocols in HIV-infected patients [oral contribution]
- International Symposium on HIV and Emerging Infectious Diseases (ISHEID) - May 16-18, 2018, Marseille, France - In silico clinical research, keep it dynamic! [invited talk]
- International Conference « Statistics and Health » - 11-12 January 2018, Toulouse Institute of Mathematics, Toulouse - Use of mathematical modeling for accelerating and personalizing clinical trials [invited talk]
- Workshop Developments in cluster randomised and stepped wedge designs, London, 21-22 Nov. 2018 Performance of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Optimizing the administration of IL7. [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Fitting pharmacokinetics data with a population-based Kalman filters [oral contribution]
- International Biometrics Society, Barcelona, Spain, 09-13 July 2018. Random Forests for high-dimensional longitudinal data. [oral contribution]
- ENBIS European Network for Business and Industrial Statistics conference, 2-6 Sept 2018. Mechanistic modeling for in silico trials [oral contribution]
- IMI 10th Anniversary Scientific Symposium, Brussels, Belgium, 22-23 October 2018. Modelling the humoral immune response to Ebola vaccine [oral contribution]
- Montreal University, Faculté de Pharmacie, Departement of Mathematical pharmacology Seminar – 21th December 2018, Montreal, Canada - Mechanistic modeling for in silico trials [invited talk]

XPOP Project-Team

7. New Results

7.1. Normalizing constants of log-concave densities

We derive explicit bounds for the computation of normalizing constants Z for log-concave densities $\pi = \mathrm{e}^{-U}/Z$ w.r.t. the Lebesgue measure on $\mathbb{R}^d$. Our approach relies on a Gaussian annealing combined with recent and precise bounds on the Unadjusted Langevin Algorithm (Durmus, A. and Moulines, E. (2016). High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm). Polynomial bounds in the dimension d are obtained with an exponent that depends on the assumptions made on U. The algorithm also provides a theoretically grounded choice of the annealing sequence of variances. A numerical experiment supports our findings. Results of independent interest on the mean squared error of the empirical average of locally Lipschitz functions are established.

7.2. The Tamed Unadjusted Langevin Algorithm

We consider the problem of sampling from a probability measure π having a density on $\mathbb{R}^d$ known up to a normalizing constant, $x \to \mathrm{e}^{-U(x)}/Z$. The Euler discretization of the Langevin stochastic differential equation (SDE) is known to be unstable in a precise sense, when the potential U is superlinear. Based on previous works on the taming of superlinear drift coefficients for SDEs, we introduce the Tamed Unadjusted Langevin Algorithm (TULA) and obtain non-asymptotic bounds in V-total variation norm and Wasserstein distance of order 2 between the iterates of TULA and π, as well as weak error bounds. Numerical experiments support our findings.

7.3. Development and performance of npde for the evaluation of time-to-event models

Normalised prediction distribution errors (npde) are used to graphically and statistically evaluate mixed-effect models for continuous responses. Our aim was to extend npde to time-to-event (TTE) models and evaluate their performance. We extended npde to TTE models using imputations to take into account censoring. We then evaluated their performance in terms of type I error and power to detect model misspecifications for TTE data by means of a simulation study with different sample sizes. Type I error was found to be close to the expected 5% significance level for all sample sizes tested. The npde were able to detect misspecifications in the baseline hazard as well as in the link between the longitudinal variable and the survival function. The ability to detect model misspecifications increased as the difference in the shape of the survival function became more apparent. As expected, the power also increased as the sample size increased. Imputing the censored events tended to decrease the percentage of rejections.

7.4. Low-rank Interaction with Sparse Additive Effects Model for Large Data Frames

Many applications of machine learning involve the analysis of large data frames-matrices collecting heterogeneous measurements (binary, numerical, counts, etc.) across samples-with missing values. Low-rank models are popular in this framework for tasks such as visualization, clustering and missing value imputation. Yet, available methods with statistical guarantees and efficient optimization do not allow explicit modeling of main additive effects such as row and column, or covariate effects. We introduced a low-rank interaction and sparse additive effects (LORIS) model which combines matrix regression on a dictionary and low-rank design, to estimate main effects and interactions simultaneously. We provide statistical guarantees in the form of upper bounds on the estimation error of both components. Then, we introduced a mixed coordinate gradient descent (MCGD) method which provably converges sub-linearly to an optimal solution and is computationally efficient for large scale data sets. Simulated and survey data showed that the method has a clear advantage over current practices, which consist in dealing separately with additive effects in a preprocessing step.

7.5. Diffusion approximations and control variates for MCMC

A new methodology was developed for the construction of control variates to reduce the variance of additive functionals of Markov Chain Monte Carlo (MCMC) samplers. Our control variates are defined as linear combinations of functions whose coefficients are obtained by minimizing a proxy for the asymptotic variance. The construction is theoretically justified by two new results. We first show that the asymptotic variances of some well-known MCMC algorithms, including the Random Walk Metropolis and the (Metropolis) Unadjusted/Adjusted Langevin Algorithm, are close to the asymptotic variance of the Langevin diffusion. Second, we provide an explicit representation of the optimal coefficients minimizing the asymptotic variance of the Langevin diffusion. Several examples of Bayesian inference problems demonstrate that the corresponding reduction in the variance is significant, and that in some cases it can be dramatic.

7.6. Density estimation for random walks in random environment

We consider the problem of non-parametric density estimation of a random environment from the observation of a single trajectory of a random walk in this environment. We first construct a density estimator using the beta-moments. We then show that the Goldenshluger-Lepski method can be used to select the beta-moment. We prove nonasymptotic bounds for the supremum norm of these estimators for both the recurrent and the transient to the right case. A simulation study supports our theoretical findings.

7.7. Imputation of mixed data with multilevel singular value decomposition

Statistical analysis of large data sets offers new opportunities to better understand many processes. Yet, data accumulation often implies relaxing acquisition procedures or compounding diverse sources. As a consequence, such data sets often contain mixed data, i.e. both quantitative and qualitative and many missing values. Furthermore, aggregated data present a natural multilevel structure, where individuals or samples are nested within different sites, such as countries or hospitals. Imputation of multilevel data has therefore drawn some attention recently, but current solutions are not designed to handle mixed data, and suffer from important drawbacks such as their computational cost. In this article, we propose a single imputation method for multilevel data, which can be used to complete either quantitative, categorical or mixed data. The method is based on multilevel singular value decomposition (SVD), which consists in decomposing the variability of the data into two components, the between and within groups variability, and performing SVD on both parts. We show on a simulation study that in comparison to competitors, the method has the great advantages of handling data sets of various size, and being computationally faster. Furthermore, it is the first so far to handle mixed data. We apply the method to impute a medical data set resulting from the aggregation of several data sets coming from different hospitals. This application falls in the framework of a larger project on Trauma patients. To overcome obstacles associated to the aggregation of medical data, we turn to distributed computation. The method is implemented in an R package

7.8. Logistic Regression with Missing Covariates – Parameter Estimation, Model Selection and Prediction

Logistic regression is a common classification method in supervised learning. Surprisingly , there are very few solutions for performing it and selecting variables in the presence of missing values. We develop a complete approach, including the estimation of parameters and variance of estimators, derivation of confidence intervals and a model selection procedure, for cases where the missing values can be anywhere in covariates. By well organizing different patterns of missingness in each observation , we propose a stochastic approximation version of the EM algorithm based on Metropolis-Hasting sampling, to perform statistical inference for logistic regression with incomplete data. We also tackle the problem of prediction for a new individual with missing values, which is never addressed. The methodology is computationally efficient, and its good coverage and variable selection properties are demonstrated in a simulation study where we contrast its performances to other methods. For instance, the popular multiple imputation by chained equation can lead to biased estimates while our method is unbiased. The method was applied on a dataset of severely traumatized patients from Paris

hospitals to predict the occurrence of hemorrhagic shock, a leading cause of early preventable death in severe trauma cases. The aim is to consolidate the current red flag procedure, a binary alert identifying patients with a high risk of severe hemorrhage. The methodology is implemented in the R package misaem.

7.9. A fast Stochastic Approximation of the EM algorithm for nonlinear mixed effects models

The ability to generate samples of the random effects from their conditional distributions is fundamental for inference in mixed effects models. Random walk Metropolis is widely used to perform such sampling, but this method is known to converge slowly for high dimensional problems, or when the joint structure of the distributions to sample is spatially heterogeneous. We propose an independent Metropolis-Hastings (MH) algorithm based on a multidimensional Gaussian proposal that takes into account the joint conditional distribution of the random effects and does not require any tuning. Indeed, this distribution is automatically obtained thanks to a Laplace approximation of the incomplete data model. We show that such approximation is equivalent to linearizing the structural model in the case of continuous data. Numerical experiments based on simulated and real data demonstrate the good performance of the proposed methods. In particular, we show that the suggested MH algorithm can be efficiently combined with a stochastic approximation version of the EM algorithm for maximum likelihood estimation in nonlinear mixed effects models.

7.10. Incomplete graphical model inference via latent tree aggregation

Graphical network inference is used in many fields such as genomics or ecology to infer the conditional independence structure between variables, from measurements of gene expression or species abundances for instance. In many practical cases, not all variables involved in the network have been observed, and the samples are actually drawn from a distribution where some variables have been marginalized out. This challenges the sparsity assumption commonly made in graphical model inference, since marginalization yields locally dense structures, even when the original network is sparse. We developed a procedure for inferring Gaussian graphical models when some variables are unobserved, that accounts both for the influence of missing variables and the low density of the original network. Our model is based on the aggregation of spanning trees, and the estimation procedure on the Expectation-Maximization algorithm. We treat the graph structure and the unobserved nodes as missing variables and compute posterior probabilities of edge appearance. To provide a complete methodology, we also propose several model selection criteria to estimate the number of missing nodes. A simulation study and an illustration flow cytometry data reveal that our method has favorable edge detection properties compared to existing graph inference techniques. The methods are implemented in an R package.