RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2019

Section New Results

Edition: 2020-03-21

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. 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. In this work [16], we correct these limitations with the so-called *combined RMSD*, 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 lRMSD 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 of choice to perform positive and negative discrimination of degree of freedom, with applications to the design of move sets and collective coordinates.

Executables to compute combined RMSD are available within the Structural Bioinformatics Library (http://sbl.inria.fr).

5.2. Modeling the flexibility of macro-molecules

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

5.2.1. 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, and also performing numerical integration in high dimensional spaces. 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.

In this work [19], we propose 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 distributions. These improvements are especially well suited to improve calculations on a per basin basis – included anharmonic ones.

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.

5.2.2. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy

Participant: F. Cazals.

In collaboration with a group of colleagues led by J. M. Carazo, CSIC, Biocomputing Unit, National Center for Biotechnology, Spain.

Single-particle analysis by electron microscopy is a well established technique for analyzing the three-dimensional structures of biological macromolecules. Besides its ability to produce high-resolution structures, it also provides insights into the dynamic behavior of the structures by elucidating their conformational variability. In this work [17], the different image-processing methods currently available to study continuous conformational changes are reviewed.

5.3. Algorithmic foundations

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

5.3.1. Comparing two clusterings using matchings between clusters of clusters

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

In collaboration with R. Watrigant, University Lyon I.

Clustering is a fundamental problem in data science, yet, the variety of clustering methods and their sensitivity to parameters make clustering hard. To analyze the stability of a given clustering algorithm while varying its parameters, and to compare clusters yielded by different algorithms, several comparison schemes based on matchings, information theory and various indices (Rand, Jaccard) have been developed. In this work [15], 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 heuristic 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).

5.3.2. Low-Complexity Nonparametric Bayesian Online Prediction with Universal Guarantees

Participant: F. Cazals.

In collaboration with A. Lhéritier, Amadeus SA.

In this work [18], we propose a novel nonparametric online predictor for discrete labels conditioned on multivariate continuous features. The predictor is based on a feature space discretization induced by a full-fledged k-d tree with randomly picked directions and a recursive Bayesian distribution, which allows to automatically learn the most relevant feature scales characterizing the conditional distribution. We prove its pointwise universality, i.e., it achieves a normalized log loss performance asymptotically as good as the true conditional entropy of the labels given the features. The time complexity to process the n-th sample point is $O(\log n)$ in probability with respect to the distribution generating the data points, whereas other exact nonparametric methods require to process all past observations. Experiments on challenging datasets show the computational and statistical efficiency of our algorithm in comparison to standard and state-of-the-art methods.

5.3.3. 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 (Coati project-team), A. Chaintreau (Columbia University), and G. Ducoffe (National Institute for Research and Development in Informatics, Bucharest).

In this work [14], 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. In this paper we disprove this. Specifically, we prove that for any $k \geq 4$, the maximum time of convergence is in $\Omega(|V|^{\Theta(\log |V|)})$.

BEAGLE Project-Team

7. New Results

7.1. Computational Glioscience: A book to review the existing mathematical models of the glial cells

[participant: H. Berry]

Over the last two decades, the recognition that astrocytes - the predominant type of cortical glial cells - could sense neighboring neuronal activity and release neuroactive agents, has been instrumental in the uncovering of many roles that these cells could play in brain processing and the storage of information. These findings initiated a conceptual revolution that leads to rethinking how brain communication works since they imply that information travels and is processed not just in the neuronal circuitry but in an expanded neuron-glial network. On the other hand the physiological need for astrocyte signaling in brain information processing and the modes of action of these cells in computational tasks remain largely undefined. This is due, to a large extent, both to the lack of conclusive experimental evidence, and to a substantial lack of a theoretical framework to address modeling and characterization of the many possible astrocyte functions. This book [https://hal.inria.fr/hal-01995842] aims at filling this gap, providing the first systematic computational approach to the complex, wide subject of neuron-glia interactions. The organization of the book is unique insofar as it considers a selection of "hot topics" in glia research that ideally brings together both the novelty of the recent experimental findings in the field and the modelling challenge that they bear. A chapter written by experimentalists, possibly in collaboration with theoreticians, will introduce each topic. The aim of this chapter, that we foresee less technical in its style than in conventional reviews, will be to provide a review as clear as possible, of what is "established" and what remains speculative (i.e. the open questions). Each topic will then be presented in its possible different aspects, by 2-3 chapters by theoreticians. These chapters will be edited in order to provide a "priming" reference for modeling neuron-glia interactions, suitable both for the graduate student and the professional researcher.

7.2. The impact of tracers on lipid digestion kinetics

[Participant: Carole Knibbe]

Dietary fats are present in the diet under different types of structures, such as spread vs emulsions (notably in processed foods and enteral formula), and interest is growing regarding their digestion and intestinal absorption. In clinical trials, there is often a need to add stable isotope-labeled triacylglycerols (TAGs) as tracers to the ingested fat in order to track its intestinal absorption and further metabolic fate. Because most TAG tracers contain saturated fatty acids, they may modify the physicochemical properties of the ingested labeled fat and thereby its digestion. However, the actual impact of tracer addition on fat crystalline properties and lipolysis by digestive lipases still deserves to be explored. In this context, we monitored the thermal and polymorphic behavior of anhydrous milk fat (AMF) enriched in homogeneous TAGs tracers and further compared it with the native AMF using differential scanning calorimetry and power X-ray diffraction. As tracers, we used a mixture of tripalmitin, triolein and tricaprylin at 2 different concentrations (1.5 and 5.7wt%, which have been used in clinical trials). The addition of TAG tracers modified the AMF melting profile, especially at the highest tested concentration (5.7 wt%). Both AMF and AMF enriched with 1.5wt% tracers were completely melted around 37°C, i.e. close to the body temperature, while the AMF enriched with 5.7wt% tracers remained partially crystallized at this temperature. Similar trends were observed in both bulk and emulsified systems. Moreover, the kinetics of AMF polymorphic transformation was modified in the presence of tracers. While only β' form was observed in the native AMF, the β-form was clearly detected in the AMF containing 5.7wt% tracers. We further tested the impact of tracers on the lipolysis of AMF in bulk using a static in vitro model of duodenal digestion. Lipolysis of AMF enriched with 5.7wt% tracers was delayed compared

with that of AMF and AMF enriched with 1.5wt% tracers. Therefore, low amounts of TAG tracers including tripalmitin do not have a high impact on fat digestion, but one has to be cautious when using higher amounts of these tracers.

7.3. The control of synaptic plasticity by external factors

[participant: H. Berry]

The dorsal striatum exhibits bidirectional corticostriatal synaptic plasticity, NMDAR- and endocannabinoids-(eCB)-mediated, necessary for the encoding of procedural learning. Therefore, characterizing factors controlling corticostriatal plasticity is of crucial importance. Brain-derived neurotrophic factor (BDNF) and its receptor, the tropomyosine receptor kinase- B (TrkB), shape striatal functions and their dysfunction deeply affects basal ganglia. BDNF/TrkB signaling controls NMDAR-plasticity in various brain structures including the striatum. However, despite cross-talk between BDNF and eCBs, the role of BDNF in eCB- plasticity remains unknown. In https://hal.inria.fr/hal-02076121, we show that BDNF/TrkB signaling promotes eCB-plasticity (LTD and LTP) induced by rate-based (low-frequency stimulation) or spike-timing- based (spike-timing-dependent plasticity, STDP) paradigm in striatum. We show that TrkB activation is required for the expression and the scaling of both eCB-LTD and eCB-LTP. Using two-photon imaging of dendritic spines combined with patch-clamp recordings, we show that TrkB activation prolongs intracellular calcium transients, thus increasing eCB synthesis and release. We provide a mathematical model for the dynamics of the signaling pathways involved in corticostriatal plasticity. Finally, we show that TrkB activation enlarges the domain of expression of eCB-STDP. Our results reveal a novel role for BDNF/TrkB signaling in governing eCB-plasticity expression in striatum, and thus the engram of procedural learning.

7.4. A new model for calcium signals in tiny sub-cellular domains

[participants: A. Denizot, H. Soula, H. Berry]

Astrocytes, a glial cell type of the central nervous system, have emerged as detectors and regulators of neuronal information processing. Astrocyte excitability resides in transient variations of free cytosolic calcium concentration over a range of temporal and spatial scales, from sub-microdomains to waves propagating throughout the cell. Despite extensive experimental approaches, it is not clear how these signals are transmitted to and integrated within an astrocyte. The localization of the main molecular actors and the geometry of the system, including the spatial organization of calcium channels IP3R, are deemed essential. However, as most calcium signals occur in astrocytic ramifications that are too fine to be resolved by conventional light microscopy, most of those spatial data are unknown and computational modeling remains the only methodology to study this issue. In https://hal.inria.fr/hal-02184344v2, we propose an IP3R-mediated calcium signaling model for dynamics in such small sub-cellular volumes. To account for the expected stochasticity and low copy numbers, our model is both spatially explicit and particle-based. Extensive simulations show that spontaneous calcium signals arise in the model via the interplay between excitability and stochasticity. The model reproduces the main forms of calcium signals and indicates that their frequency crucially depends on the spatial organization of the IP3R channels. Importantly, we show that two processes expressing exactly the same calcium channels can display different types of calcium signals depending on the spatial organization of the channels. Our model with realistic process volume and calcium concentrations successfully reproduces spontaneous calcium signals that we measured in calcium micro-domains with confocal microscopy and predicts that local variations of calcium indicators might contribute to the diversity of calcium signals observed in astrocytes. To our knowledge, this model is the first model suited to investigate calcium dynamics in fine astrocytic processes and to propose plausible mechanisms responsible for their variability.

7.5. Evolution of genome size

Using the Aevol software, we investigated the dynamics of genome size under different evolutionary pressures (variation of mutation rates and variation of population sizes). The dynamics of the model enabled us to identify a new mutational pressure on genome size that spontaneously increase the fraction of non-coding

sequences. We showed that this mutational pressure interact with the selective pressure for robustness (knibbe et al., 2007), resulting in an equilibrium of genome size and non-coding proportion. Moreover, we showed that this equilibrium can change depending on the size of the population due to the resulting effect on selection intensity. A paper has been published in the proceedings of the ALife 2019 conference (cardes et al, 2019) and an article in in preparation.

7.6. Dynamics of evolutionary innovation

Using a combination of mathematical and computational models (NK-Fitness-Landscapes and Aevol), we investigated the dynamics of innovation in evolving systems. We showed that innovation is often triggered by specific mutational events, typically structural variation of the genome (e.g. duplications, inversions, ...). We further studied this effect and showed that innovation is due to the differences of time scale between the different kinds of mutations: fast mutations (typically point mutations) are rapidly exhausted, resulting in a fitness plateau. However, slow mutations (typically structural variations) can open new evolutionary paths, resulting in the population escaping from the fitness plateau. An article is in preparation in collaboration with Santiago F. Elena (CSIC, Spain).

7.7. Evolution of biological complexity

Using a modified version of the Aevol platform, we studied the evolution of complex features. By evolving population of organisms in conditions where complexity is counter-selected, we showed that complexity accumulates even in these conditions, i.e. even when complex organisms are less fit than simple ones. Moreover we showed that complex organisms are not more robust and not more evolvable than simple ones. This shows that evolution spontaneously initiate a "complexity ratchet" that forces complexity to grow. An article is in press in the Artificial Life Journal (to be published in 2020).

7.8. Dynamics of mutator strains

In a long-lasting collaboration with Utrecht University, we studied the dynamics of mutator strains in constant environments (mutator strains being individuals which mutation rate is increased by several orders of magnitude). Contrary to what is generally admitted, we showed that, although mutators initially suffer from a mutational burden (in coherence with the theory), they are able to quickly recover and avoid the burden. Moreover, we showed that they do so by contracting their coding genome compartment and expanding their non-coding compartment. This result show that mutators can thrive even in a constant environment (rutten et al., 2019).

7.9. Mutiscale phylogenetics models

[Participant: Eric Tannier]

We progressed in the modeling of multi-scale phylogenetic events: we gave an algorithm to infer gene conversions according to a phylogeny [7], a complexity result and an algorithm for transfers with replacements [6], and we devised a simulation tool integrating extinct species and horizontal inheritance [3].

7.10. Evolution of the Drosophila melanogaster Chromatin Landscape and Its Associated Proteins

[participant: A. Crombach]

In the nucleus of eukaryotic cells, genomic DNA associates with numerous protein complexes and RNAs, forming the chromatin landscape. Through a genome-wide study of chromatin-associated proteins in Drosophila cells, five major chromatin types were identified as a refinement of the traditional binary division into hetero- and euchromatin. These five types were given color names in reference to the Greek word chroma. They are defined by distinct but overlapping combinations of proteins and differ in biological and biochemical properties, including transcriptional activity, replication timing, and histone modifications. We assessed the evolutionary relationships of chromatin-associated proteins and presented an integrated view of the evolution and conservation of the fruit fly Drosophila melanogaster chromatin landscape. We combined homology prediction across a wide range of species with gene age inference methods to determine the origin of each chromatin-associated protein. This provided insight into the evolution of the different chromatin types. Our results indicate that for the euchromatic types, YELLOW and RED, young associated proteins are more specialized than old ones; and for genes found in either chromatin type, intron/exon structure is lineage-specific. Next, we provide evidence that a subset of GREEN-associated proteins is involved in a centromere drive in D. melanogaster. Our results on BLUE chromatin support the hypothesis that the emergence of Polycomb Group proteins is linked to eukaryotic multicellularity. In light of these results, we discuss how the regulatory complexification of chromatin links to the origins of eukaryotic multicellularity.

BIGS Project-Team

7. New Results

7.1. Stochastic modelling

Participants: A. Gégout-Petit, S. Mézières, 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), have proposed estimators of the parameters and showed their properties and proposed a way to choose between several neighborhood models. These results were published in Spatial Statistics [4].

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 paper published in Scientific Reports [8].

After preliminary suggestions on the building of models for low-grade gliomas [3], we focused our attention on the diffuse character of such tumors. We characterized the infiltrating phenotype (infiltration rate, direction of infiltration, evolution of morphology over time) as a new variable to consider in a context of multifactorial modelling (submitted article). A monocentric retrospective study has been conducted on the local database, estimating survival paramaters and comparing the effects of treatments (writing article). A brain cartography obtained by sensorial simulations during awake surgery with the aid of clustering analysis has been published in "Brain - A Journal of Neurology" [6].

7.2. Optimal Control of Markov Processes

Participants: B. Scherrer

Finite-horizon lookahead policies are abundantly used in Reinforcement Learning and demonstrate impressive empirical success. Usually, the lookahead policies are implemented with specific planning methods such as Monte Carlo Tree Search (e.g. in AlphaZero). Referring to the planning problem as tree search, a reasonable practice in these implementations is to back up the value only at the leaves while the information obtained at the root is not leveraged other than for updating the policy. Here, we question the potency of this approach. Namely, the latter procedure is non-contractive in general, and its convergence is not guaranteed. Our proposed enhancement, in [9], published in AAAI'2019, is straightforward and simple: use the return from the optimal tree path to back up the values at the descendants of the root. This leads to a γ^h-contracting procedure, where γ is the discount factor and h is the tree depth. To establish our results, we first introduce a notion called *multiple-step greedy consistency*. We then provide convergence rates for two algorithmic instantiations of the above enhancement in the presence of noise injected to both the tree search stage and value estimation stage.

Value iteration is a method to generate optimal control inputs for generic nonlinear systems and cost functions. Its implementation typically leads to approximation errors, which may have a major impact on the closed-loop system performance. We talk in this case of approximate value iteration (AVI). In [24], published in CDC'2019, we investigate the stability of systems for which the inputs are obtained by AVI. We consider deterministic discrete-time nonlinear plants and a class of general, possibly discounted, costs. We model the closed-loop system as a family of systems parameterized by tunable parameters, which are used for the approximation of the value function at different iterations, the discount factor and the iteration step at which we stop running the algorithm. It is shown, under natural stabilizability and detectability properties as well as mild conditions on the approximation errors, that the family of closed-loop systems exhibit local practical

stability properties. The analysis is based on the construction of a Lyapunov function given by the sum of the approximate value function and the Lyapunov-like function that characterizes the detectability of the system. By strengthening our conditions, asymptotic and exponential stability properties are guaranteed.

Many recent successful (deep) reinforcement learning algorithms make use of regularization, generally based on entropy or Kullback-Leibler divergence. In [10], published in ICML'2019, we propose a general theory of regularized Markov Decision Processes that generalizes these approaches in two directions: we consider a larger class of regularizers, and we consider the general modified policy iteration approach, encompassing both policy iteration and value iteration. The core building blocks of this theory are a notion of regularized Bellman operator and the Legendre-Fenchel transform, a classical tool of convex optimization. This approach allows for error propagation analyses of general algorithmic schemes of which (possibly variants of) classical algorithms such as Trust Region Policy Optimization, Soft Q-learning, Stochastic Actor Critic or Dynamic Policy Programming are special cases. This also draws connections to proximal convex optimization, especially to Mirror Descent.

7.3. Algorithms and Estimation for graph data

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

We consider the problem of graph estimation in a zero-inflated Gaussian model. In this model, zero-inflation is obtained by double truncation (right and left) of a Gaussian vector. The goal is to recover the latent graph structure of the Gaussian vector with observations of the zero-inflated truncated vector. We propose a two step estimation procedure. The first step consists in estimating each term of the covariance matrix by maximising the corresponding bivariate marginal log-likelihood of the truncated vector. The second one uses the graphical lasso procedure to estimate the sparsity of the precision matrix, which encodes the graph structure. We then state some theoretical results about the convergence rate of the covariance matrix and precision matrix estimators. These results allow us to establish consistency of our procedure with respect to graph structure recovery. We also present some simulation studies to corroborate the efficiency of our procedure. It is the object of the submitted paper [29], a part of the PhD thesis [1] and the communications [16] [15].

7.4. Regression and machine learning

Participants: E. Albuisson, T. Bastogne, S. Ferrigno, A. Gégout-Petit, F. Greciet, P. Guyot, C. Karmann, J.-M. Monnez, N. Sahki, S. Mézières, B. Lalloué

Through a collaboration with the pharmaceutical company Transgene (Strasbourg, France), we have developed a method for selecting covariates. The problem posed by Transgene was to establish patient profiles on the basis of their response to a treatment developed by Transgene. We have then proposed a new methodology for selecting and ranking covariates associated with a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology successively intertwines the clustering of covariates, decorrelation of covariates using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking. A simulation study shows the interest of the decorrelation inside the different clusters of covariates. We have applied our method to the data of Transgene. For instance, transcriptomic data of 37 patients with advanced non-small-cell lung cancer who have received chemotherapy, to select the transcriptomic covariates that explain the survival outcome of the treatment. Our method has also been applied in another collaboration with biologists (CRAN laboratory, Nancy, France). In that case, our method has been applied to transcriptomic data of 79 breast tumor samples, to define patient profiles for a new metastatic biomarker and associated gene network. Our developed method is a contribution to the development of personalized medicine. We have published the method, as well as the two applications in [27].

In order to detect change of health state for lung-transplanted patient, we have begun to work on breakdowns in multivariate physiological signals. We consider the score-based CUSUM statistic and propose to evaluate the detection performance of some thresholds on simulation data. Two thresholds come from the literature: Wald's constant and Margavio's instantaneous threshold, and three contribution thresholds built by a simulation-based procedure: the first one is constant, the second instantaneous and the third is a dynamical version of

the previous one. The threshold's performance is evaluated for several scenarii, according to the detection objective and the real change in the data. The simulation results show that Margavio's threshold is the best at minimizing the detection delay while maintaining the given false alarm rate. But on real data, we suggest to use the dynamic instantaneous threshold because it is the easiest to build for practical implementation. It is the purpose of the communication [11] and the submitted paper [35].

We consider the problem of variable selection in regression models. In particular, we are interested in selecting explanatory covariates linked with the response variable and we want to determine which covariates are relevant, that is which covariates are involved in the model. In this framework, we deal with L1-penalised regression models. To handle the choice of the penalty parameter to perform variable selection, we develop a new method based on knockoffs. This revisited knockoffs method is general, suitable for a wide range of regressions with various types of response variables. Besides, it also works when the number of observations is smaller than the number of covariates and gives an order of importance of the covariates. Finally, we provide many experimental results to corroborate our method and compare it with other variable selection methods. It is the object of communication [17], the submitted paper [30] and a chapter of the PhD thesis [1].

In order to model crack propagation rate, continuous physical phenomenon that presents several regimes, we proposed a piecewise polynomial regression model under continuity and/or derivability assumptions as well as a statistical inference method to estimate the transition times and the parameters of each regime. We proposed several algorithms and studied their efficiency. The most efficient algorithm relies on dynamic programming. It is the object of the communication [14] and the PhD thesis of Florine Greciet.

Let consider a regression model $Y = m(X) + \sigma(X)\varepsilon$ to explain Y from X, 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 validate 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 the cvmgof R package, an easy-to-use tool for many users. The use of the package is described and illustrated using simulations to compare the three implemented tests in a paper in progress.

In epidemiology, we are working with INSERM clinicians and biostatisticians to study fetal development in the last two trimesters of pregnancy. Reference or standard curves are required in this kind of biomedical problems. Values which lie outside the limits of these reference curves may indicated the presence of disorder. Data are from the French EDEN mother-child cohort (INSERM). It is a mother-child cohort study investigating the prenatal and early postnatal determinants of child health and development. 2002 pregnant women were recruited before 24 weeks of amenorrhoea in two maternity clinics from middle-sized French cities (Nancy and Poitiers). From May 2003 to September 2006, 1899 newborns were then included. The main outcomes of interest are fetal (via ultra-sound) and postnatal growth, adiposity development, respiratory health, atopy, behaviour and bone, cognitive and motor development. We are studying fetal weight that depends on the gestional age in the second and the third trimesters of mother's pregnancy. Some classical empirical and parametric methods as polynomials are first used to construct these curves. Polynomial regression is one of the most common parametric approach for modelling growth data espacially during the prenatal period. However, some of them require strong assumptions. So, we propose to work with semi-parametric LMS method, by modifying the response variable (fetal weight) with a Box-cox transformation. Nonparametric methods as Nadaraya-Watson kernel estimation or local polynomial estimation are also proposed to construct these curves. It is the object of the communication [28] and a paper is in progress. 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).

Many articles were devoted to the problem of recursively estimating eigenvectors corresponding to eigenvalues in decreasing order of the expectation of a random matrix using an i.i.d. sample of it. The present study makes the following contributions: the convergence of processes to normed eigenvectors is proved under two sets of more general assumptions, the observed random matrices are no more supposed i.i.d.; moreover, the scope of these processes is widened. The application to online principal component analysis of a data stream is treated, assuming that data are realizations of a random vector Z whose expectation is unknown and is estimated online, as well as possibly the metric used when it depends on unknown characteristics of Z; two types of processes are studied: we are no more bound to use a data mini-batch at each step, but we can use all previously observed data up to the current step without storing them, thus taking into account all the information contained in previous data. The conducted experiments have shown that processes of the second type are faster than those of the first type. It is the object of the submitted paper [32] and the communication [21].

The study addresses the problem of constrained binary logistic regression, particularly in the case of a data stream, using a stochastic approximation process. To avoid a numerical explosion which can be encountered, we propose to use a process with online standardized data instead of raw data. This type of process can also be used when we have to standardize the explanatory variables, for example in the case of a shrinkage method such as LASSO. Herein, we define and study the almost sure convergence of an averaged constrained stochastic gradient process with online standardized data. Moreover we propose to use a piecewise constant step-size in order that the step-size does not decrease too quickly and reduce the speed of convergence. Processes of this type are compared to classical processes on real and simulated datasets. The results of conducted experiments confirm the validity of the choices made. This will be used in an ongoing application to online updating of a score in heart failure patients. It is the object of the submitted paper [31] and the communications [20],[19].

CAPSID Project-Team

7. New Results

7.1. Axis 1 : New Approaches for Knowledge Discovery in Structural Databases

7.1.1. Biomedical Knowledge Discovery

Our collaboration with clinicians at the CHRU Nancy in the framework of the RHU FIGHT-HF program and of the Contrat d'Interface has lead to two publications demonstrating the added value of database and knowledge graph exploitation when analyzing observational or prospective cohorts. In a retrospective observational study, we have identified and characterized patient subgroups presenting stable or unstable positivity to anti-phospholipid antobodies assays [15]. In the European FibroTarget cohort study, we have contributed to the characterization of at-risk phenotypic groups using proteomic biomarkers [16].

Another application is carried out in collaboration with the Orpailleur Team and concerns the PraktikPharma ANR project. We aim at building explanations for severe drug side effects (such as drug-induced liver injury or severe cutaneous adverse reaction) from pharmacogenomics RDF graph (PGXlod). We obtained a podium abstract at the MedInfo 2019 conference for providing molecular characterization for unexplained adverse drug reactions using pharmacogenomics RDF graph (PGXlod) [30].

7.1.2. Stochastic Decision Trees for Similarity Computation

In the frame of Kévin Dalleau's PhD thesis, we have designed a method to compute similarities on unlabeled data using stochastic decision trees [31], [27]. 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 extended the approach to the computation of pairwise similarity for graph nodes. The experimental results are competitive with state of the art methods. We are currently working on merging the two similarity methods (on attribute-value objects and on graph nodes) to attributed graphs where the nodes are described by attributes.

We plan to study the application of this 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.

7.1.3. Protein Annotation and Machine Learning

We have been involved in the 3rd international CAFA Challenge ("Critical Assessment of Functional Annotation") through our work on (i) domain functional annotation (Zia Alborzi's PhD thesis) and (ii) label propagation in graphs (Bishnu Sarker's PhD thesis). We were therefore contributors of the general report published this year [23].
As part of his PhD work, Bishnu Sarker developed and tested on UniProt/SwissProt a new method for functional annotation of proteins using domain embedding-based sequence classification [25].

Multiple Instance Learning (MIL) is a machine learning strategy that can be applied to sets of sequences describing organisms displaying a given property. The purpose here is to be able to classify a new organism with respect to this property based on its sequences and their similarity to the sequences of classified organisms. New MIL algorithms have been described and tested in the framework of a collaboration [26], [24]. Another collaborative work has lead to the development of a distributed algorithm for large-scale graph clustering [34].

7.2. Axis 2 : Integrative Multi-Component Assembly and Modeling

7.2.1. EROS-DOCK algorithm and its extensions

We have adapted our EROS-DOCK protein-protein docking software [35], [19] to account for experimental knowledge on the protein-protein interface to be modeled. Indeed, structural biology experiments can identify pairs of amino-acids from each protein in a protein-protein interface that are likely to be in close contact. This additional restraint is used to pre-prune the 3D rotational space of one protein toward another, by eliminating cones of rotations that cannot fulfill the distance between the two points at the protein surfaces. Using a single restraint permits to decrease the average execution time by at least 90 percent.

We also developed a new version of EROS-DOCK for multi-body docking (modeling assemblies of more than 2 proteins), using a combinatorial approach. We assembled trimers by docking in a first stage all possible combinations of pairs of proteins involved in the multi-body complex. Possible trimer solutions are assembled by fixing one protein, the "root-protein" (protein A, say) at the origin and by placing the other two around it using the transformations,T[AB] and T[AC], from the corresponding pairwise solution lists returned by EROS-DOCK. If the three transformations together form a near-native (biologically relevant) trimer structure, then it is natural to suppose that T[BC] should be found in the list of B-C pairwise solutions.

Both extensions of the EROS-DOCK algorithm reported last year and published early this year [19] have been presented by Maria-Elisa Ruiz Echartea at the 2019 CAPRI meeting in april 2019 (http://www.capri-docking.org/events/) and at the MASIM meeting in november 2019 [28]. These results are part of her PhD Thesis that was defended on december 18, 2019 (the thesis will soon be available on HAL). A paper describing EROS-DOCK adaptation to multi-body docking in under revision in *Proteins*.

7.2.2. Protein docking

The regular participation of the Capsid team to the CAPRI challenge is acknowledged through its contribution to the review published this year on CAPRI round 46 [17].

We also contributed to an evaluation of docking software performance in protein-glycosaminoglycan systems [22].

7.2.3. 3D modeling and virtual screening

We have built a 3D model by homology of a new class of relaxase involved in the horizontal transfer of DNA in a group of bacteria called Firmicutes [21].

We also built a 3D model of a chemosensory GPCR as a potential target to control a parasite in plants [13].

Virtual screening was applied on various targets in a re-purposing strategy and led to the discovery of small molecules active against invasive fungal disease [14], [18].

DYLISS Project-Team

7. New Results

7.1. Scalable methods to query data heterogenity

Participants: Emmanuelle Becker, Lucas Bourneuf, Olivier Dameron, Xavier Garnier, Vijay Ingalalli, Marine Louarn, Yann Rivault, Anne Siegel.

Increasing life science resources re-usability using Semantic Web technologies [*E. Becker, O. Dameron, X. Garnier, V. Ingallali, M. Louarn, Y. Rivault, A. Siegel*] [25], [18], [29], [31], [23], [27], [28]. Our work was focused on assessing to what extent Semantic Web technologies also facilitate reproducibility and reuse of life sciences studies involving pipelines that compute associations between entities according to intermediary relations and dependencies.

- We followed on 2018 action exploratoire Inria by studying possible optimizations for federated SPARQL queries [31]
- We considered a case-study in systems biology ([Regulatorycircuits link]), which provides tissue-specific regulatory interaction networks to elucidate perturbations across complex diseases. We relied on this structure and used Semantic Web technologies (i) to integrate the Regulatory Circuits data, and (ii) to formalize the analysis pipeline as SPARQL queries. Our result was a 335,429,988 triples dataset on which two SPARQL queries were sufficient to extract each single tissue-specific regulatory network.
- A second case-study concerned public health data for reusing electronic health data, selecting patients, identifying specific events and interpreting results typically requires biomedical knowledge [64]. We developed the queryMed R package [18], [29]. It aims to facilitate the integration of medical and pharmacological knowledge stored in formats compliant with the Linked Data paradigm (e.g. OWL ontologies and RDF datasets) into the R statistical programming environment. We showed that linking a medical database of 1003 critical limb ischemia (CLI) patients to ontologies allowed us to identify all the drugs prescribed for CLI and also to detect one contraindicated prescription for one patient. We also investigated temporal models of care sequences for the exploration of medico-administrative data as part of Johanne Bakalara's PhD, supervized with Thomas Guyet (Lacodam) and Emmanuel Oger (Repères).
- We pursued the development of AskOmics [27]. Version 3 adds the capability to generate the graph of entity types (aka abstraction) from typed RDF datasets, improved management of entity hierarchies and support for federated queries on external SPARQL endpoints such as UniProt and neXtProt.

Graph compression and analysis [*L. Bourneuf*] [26], [24]. Because of the increasing size and complexity of available graph structures in experimental sciences like molecular biology, techniques of graph visualization tend to reach their limit.

- We developed the Biseau approach, a programming environment aiming at simplifying the visualization task. Biseau takes advantage of Answer Set Programming and shows as a use-case how Formal Concept Analysis can be efficiently described at the level of its properties, without needing a costly development process. It reproduces the core results of existing tools like LatViz or In-Close.
- We formalized a graph compression search space in order to provide approximate solutions to the NP-complete problem of computing a lossless compression of the graph based on the search of cliques and bicliques. Our conclusion is that the search for graph compression can be usefully associated with the search for patterns in a concept lattice and that, conversely, confusing sets of objects and attributes brings new interesting problems for FCA.

7.2. Metabolism: from enzyme sequences to systems ecology

Participants: Méziane Aite, Arnaud Belcour, Mael Conan, François Coste, Clémence Frioux, Jeanne Got, Anne Siegel, Hugo Talibart.

Modelling proteins with long distance dependencies [*F. Coste, H. Talibart*] [15], [30], [30]

- We proposed to use information on protein contacts to train probabilistic context-free grammars representing families of protein sequences. We developed the theory behind the introduction of contact constraints in maximum-likelihood and contrastive estimation schemes and implemented it in a machine learning framework for protein grammars. Evaluation showed high fidelity of grammatical descriptors to protein structures, improved precision in recognizing sequences and the ability to model a meta-family of proteins that could not be modeled by classical approaches [15].
- We then investigated the problem of modeling proteins with crossing dependencies. Motivated by their success on contact prediction, we propose to use Potts models for the purposes of modeling proteins and searching. We developed ComPotts a tool for optimal alignment and comparison of Potts models, enabling to take into account the coevolution of residues for the search of protein homologs [30], [40].

Large-scale eukaryotic metabolic network reconstruction [*A. Siegel, C. Frioux, M. Aite, A. Belcour, J. Got, N. Théret, M. Conan*] [17], [14], [38]. Metabolic network reconstruction has attained high standards but is still challenging for complex organisms such as eukaryotes.

- *Large-scale eukaryotic metabolic network reconstruction*: We participated to the reconstruction of a genome–scale metabolic network for the brown Algae *Saccharina japonica* and *Cladosiphon okamuranus* in order to shed light of the specificities on the carotenoid biosynthesis Pathway.
- *Metabolic pathway inference from non genomic data*: We designed methods for the identification of metabolic pathways for which enzyme information is not precise enough. As an application study, we focused on Heterocyclic Aromatic Amines (HAAs), which are environmental and food contaminants classified as probable carcinogens. Our approach based on a refinement of molecular predictions with enzyme activity scores allows to accurately predict HAAs biotransformation and their potentials DNA reactive compounds [54].

Systems ecology: design of microbial consortia [*C. Frioux, A. Belcour, J. Got, M. Aite, A. Siegel*] [21], [22], [34], [33].

- We participated to the application of our methods to algal-microbial consortia, with good preliminary results, and presented them as an invited conference [22].

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

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

Creation of predictive functional signaling networks [*M. Folschette, N. Théret*] [16].

- Integrating genome-wide gene expression patient profiles with regulatory knowledge is a challenging task because of the inherent heterogeneity, noise and incompleteness of biological data. We proposed an automatic pipeline to extract automatically regulatory knowledge from pathway databases and generate novel computational predictions related to the state of expression or activity of biological molecules. We applied it in the context of hepatocellular carcinoma (HCC) progression, and evaluated the precision and the stability of these computational predictions. Our computational model predicted the shifts of expression of 146 initially non-observed biological components. Our predictions were validated at 88% using a larger experimental dataset and cross-validation techniques.

Experimental evidences of transcript predictions [*C. Belleannée, S. Blanquart, N. Guillaudeux*] [13].

- We designed comparative-genomics based models of gene structures through genes comparisons across species. These models enable to predict putative transcript isoforms in a species given the knowledge available in other species [46]. We recently published a first experimental validation of such a predicted transcriptome [13]. In this work, transcript isoforms of the human TRPM8 gene yield transcript predictions in the mouse TRPM8 gene, which are experimentally validated using targeted PCR in mouse tissues. This work also provides a first attempt to estimate origin of new isoforms during the gene evolution.
- In another collaboration with IGDR, we considered a multi-species gene comparison including human, mouse and dog [39]. This work reveals global trends of the gene isoform sets evolution, suggesting a extremely high plasticity of alternative transcription and alternative splicing propensities in those three species. This work moreover provides experimental evidences of the predicted transcripts based on public RNAseq data, highlighting the tissue specificity of isoform expression across species.

Formalizing and enriching phenotype signatures using Boolean networks [*C. Biane-Fourati, M. Wéry, A. Siegel, O. Dameron*] [20], [42], [22]

- We used Formal Concept Analysis as a symbolic bi-clustering techniques to classify and sort the steady states of a Boolean network according to biological signatures based on the hierarchy of the roles the network components play in the phenotypes. We applied our approach on a T helper lymphocyte (Th) differentiation network with a set of signatures corresponding to the sub-types of Th. This led to the identification and prediction of a new hybrid sub-type later confirmed by the literature.

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2019.

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 chose not to detail here the results on more theoretical aspects of computer science when these are initially addressed in contexts not directly related to computational biology even though those on string [11], [36], [40], [41], [23], [45] and graph algorithms in general [35], [39], [38], [17], [43] are relevant for life sciences, such as for instance pan-genome analysis, or could become more specifically so in a near future. One important example of the latter concerns enumeration algorithms that has always been at the heart of the computer science and mathematics interests of the team. In such context, the so-called reconfiguration problem which asks whether one solution can be transformed into the other in a step-by-step fashion such that each intermediate solution is also feasible is of particular relevance. This was explored in the context of a perfect matching problem [37].

A few other results of 2019 are not mentioned in this report, not because the corresponding work is not important, but because it was likewise more specialised [8], [9], [12], [44]. In the same way, also for space reasons, we chose not to detail the results presented in some biological papers of the team when these did not require a mathematical or algorithmic input [16], [22].

On the other hand, we do mention a couple of works that were in preparation or about to be submitted towards the end of 2018.

6.2. Axis 1: Genomics

Transcriptome profiling using Nanopore sequencing Our vision of DNA transcription and splicing has changed dramatically with the introduction of short-read sequencing. These high-throughput sequencing technologies promised to unravel the complexity of any transcriptome. Generally gene expression levels are well-captured using these technologies, but there are still remaining caveats due to the limited read length and the fact that RNA molecules had to be reverse transcribed before sequencing. Oxford Nanopore Technologies has recently launched a portable sequencer which offers the possibility of sequencing long reads and most importantly RNA molecules. In [28], we generated a full mouse transcriptome from brain and liver using such Oxford Nanopore device. As a comparison, we sequenced RNA (RNA-Seq) and cDNA (cDNA-Seq) molecules using both long and short reads technologies and tested the TeloPrime preparation kit, dedicated to the enrichment of full-length transcripts. Using spike-in data, we confirmed in [28] that expression levels are efficiently captured by cDNA-Seq using short reads. More importantly, Oxford Nanopore RNA-Seq tends to be more efficient, while cDNA-Seq appears to be more biased. We further showed that the cDNA library preparation of the Nanopore protocol induces read truncation for transcripts containing internal runs of T's. This bias is marked for runs of at least 15 T's, but is already detectable for runs of at least 9 T's and therefore concerns more than 20% of the expressed transcripts in mouse brain and liver. Finally, we outlined that bioinformatic challenges remain ahead for quantifying at the transcript level, especially when reads are not full-length. Accurate quantification of repeat-associated genes such as processed pseudogenes also remains difficult, and we show in the paper that current mapping protocols which map reads to the genome largely over-estimate their expression, at the expense of their parent gene.

Genotyping and variant detection The amount of genetic variation discovered and characterised in human populations is huge, and is growing rapidly with the widespread availability of modern sequencing technologies. Such a great deal of variation data, that accounts for human diversity, leads to various challenging computational tasks, including variant calling and genotyping of newly sequenced individuals. The standard pipelines for addressing these problems include read mapping, which is a computationally expensive procedure. A few mapping-free tools were proposed in recent years to speed up the genotyping process. While such tools have highly efficient run-times, they focus on isolated, bi-allelic SNPs, providing limited support for multi-allelic SNPs, indels, and genomic regions with high variant density. To address these issues, we introduced MALVA, a fast and lightweight mapping-free method to genotype an individual directly from a sample of reads [10]. MALVA is the first mapping-free tool that is able to genotype multi-allelic SNPs and indels, even in high density genomic regions, and to effectively handle a huge number of variants such as those provided by the 1000 Genome Project. An experimental evaluation on whole-genome data shows that MALVA requires one order of magnitude less time to genotype a donor than alignment-based pipelines, providing similar accuracy. Remarkably, on indels, MALVA provides even better results than the most widely adopted variant discovery tools.

Still on the issue of SNP detection, in [25], we developed the positional clustering theory that (i) describes how the extended Burrows–Wheeler Transform (eBWT) of a collection of reads tends to cluster together bases that cover the same genome position, (ii) predicts the size of such clusters, and (iii) exhibits an elegant and precise LCP array based procedure to locate such clusters in the eBWT. Based on this theory, we designed and implemented an alignment-free and reference-free SNP calling method, and we devised a SNP calling pipeline. Experiments on both synthetic and real data show that SNPs can be detected with a simple scan of the eBWT and LCP arrays as, in agreement with our theoretical framework, they are within clusters in the eBWT of the reads. Finally, our tool intrinsically performs a reference-free evaluation of its accuracy by returning the coverage of each SNP. Based on the results of the experiments on synthetic and real data, we conclude that the positional clustering framework can be effectively used for the problem of identifying SNPs, and it appears to be a promising approach for calling other types of variants directly on raw sequencing data.

Finally, variant detection and various related algorithmic problems were extensively explored in the PhD of Leandro I. S. de Lima [2] defended in April 2019.

Bubble generator Bubbles are pairs of internally vertex-disjoint (s, t)-paths in a directed graph, which have many applications in the processing of DNA and RNA data such as variant calling as presented above. Listing and analysing all bubbles in a given graph is usually unfeasible in practice, due to the exponential number of bubbles present in real data graphs. In [4], we proposed a notion of bubble generator set, *i.e.*, a polynomial-sized subset of bubbles from which all the other bubbles can be obtained through a suitable application of a specific symmetric difference operator. This set provides a compact representation of the bubble space of a graph. A bubble generator can be useful in practice, since some pertinent information about all the bubbles can be more conveniently extracted from this compact set. We provided a polynomial-time algorithm to decompose any bubble of a graph into the bubbles of such a generator in a tree-like fashion. Finally, we presented two applications of the bubble generator on a real RNA-seq dataset.

Genome assembly The continuous improvement of long-read sequencing technologies along with the development of ad-doc algorithms has launched a new *de novo* assembly era that promises high-quality genomes. However, it has proven difficult to use only long reads to generate accurate genome assemblies of large, repeat-rich human genomes. To date, most of the human genomes assembled from long error-prone reads add accurate short reads to further improve the consensus quality (polishing). In a paper to be submitted before the end of 2019 (with as main authors A. di Genova and M.-F. Sagot), we report the development of an algorithm for hybrid assembly, WENGAN, and its application to hybrid sequence datasets from four human samples. WENGAN implements efficient algorithms that exploit the sequence information of short and long reads to tackle assembly contiguity as well as consensus quality. We show that the resulting genome assemblies have high contiguity (contig NG50:16.67-62.06 Mb), few assembly errors (contig NGA50:10.9-45.91 Mb), good consensus quality (QV:27.79-33.61), high gene completeness (BUSCO complete: 94.6-95.1%), and consume few computational resources (CPU hours:153-1027). In particular, the WENGAN assembly of the

haploid CHM13 sample achieved a contig NG50 of 62.06 Mb (NGA50:45.91 Mb), which surpasses the contiguity of the current human reference genome (GRCh38 contig NG50:57.88 Mb). Because of its lower cost, WENGAN is an important step towards the democratisation of the *de novo* assembly of human genomes. WENGAN is available at https://github.com/adigenova/wengan.

On assembly still, although haplotype-aware genome assembly plays an important role in genetics, medicine and various other disciplines, the generation of haplotype-resolved *de novo* assemblies remains a major challenge. Beyond distinguishing between errors and true sequential variants, one needs to assign the true variants to the different genome copies. Recent work has pointed out that the enormous quantities of traditional NGS read data have been greatly underexploited in terms of haplotig computation so far, which reflects the fact that the methodology for reference independent haplotig computation has not yet reached maturity. We presented in [7] a new approach, called POLYploid genome fitTEr (POLYTE) for a *de novo* generation of haplotigs for diploid and polyploid genomes of known ploidy. Our method follows an iterative scheme where in each iteration reads or contigs are joined, based on their interplay in terms of an underlying haplotype-aware overlap graph. Along the iterations, contigs grow while preserving their haplotype identity. Benchmarking experiments on both real and simulated data demonstrate that POLYTE establishes new standards in terms of error-free reconstruction of haplotype-specific sequences. As a consequence, POLYTE outperforms state-of-the-art approaches in various relevant aspects, notably in polyploid settings.

Others Besides the above, we have also explored a proteogenomics workflow for the expert annotation of eukaryotic genomes [18], as well as a technology- and species-independent simulator of sequencing data and genomic variants [42].

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, who was until the end of 2018 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 Analysis of differential expression of genes is often performed to understand how the metabolic activity of an organism is impacted by a perturbation. However, because the system of metabolic regulation is complex and all changes are not directly reflected in the expression levels, interpreting these data can be difficult. In [26], we presented a new algorithm and computational tool that uses a genome-scale metabolic reconstruction to infer metabolic changes from differential expression data. Using the framework of constraint-based analysis, our method produces a qualitative hypothesis of a change in metabolic activity. In other words, each reaction of the network is inferred to have increased, decreased, or remained unchanged in flux. In contrast to similar previous approaches, our method does not require a biological objective function and does not assign on/off activity states to genes. An implementation is provided and is available online at the address https://github.com/htpusa/moomin. We applied the method to three published datasets to show that it successfully accomplishes its two main goals: confirming or rejecting metabolic changes suggested by differentially expressed genes based on how well they fit in as parts of a coordinated metabolic change, as well

as inferring changes in reactions whose genes did not undergo differential expression. The above work was also part of the PhD of Taneli Pusa [3] defended in February 2019.

Metabolic games Game theory is a branch of applied mathematics originally developed to describe and reason about situations where two or more rational agents, the "homo economicus", are faced with choices and have potentially conflicting goals. All participants want to maximise their own well-being, but are doing so taking into account that everyone else is doing the same. Thus paradoxical, suboptimal, outcomes are possible and even common. Evolutionary game theory was born out of the realisation that rational choice can be replaced by natural selection: in the course of evolution the strategy (phenotype) that would "win" the game would prevail by simply proliferating more successfully thanks to its success in the "game". It turns out that phenotype prediction in the context of metabolic networks is exactly the type of problem that evolutionary game theory was meant to answer: given a set of choices (as defined by a metabolic network reconstruction), what will be the actual metabolism observed? In other words, if we culture a set of organisms together in a given medium, which are the phenotype(s) that emerge as winners? In [27], we sought to provide a short introduction to both evolutionary game theory and its use in the context of metabolic modelling. This work was also part of the PhD of Taneli Pusa [3].

6.4. Axis 3: (Co)Evolution

Modelling invasion Nowadays, the most used model in studies of the coevolution of hosts and symbionts is phylogenetic tree reconciliation. A crucial issue in this model is that from a biological point of view, reasonable cost values for an event-based reconciliation are not easily chosen. Different methods have been developed to infer such cost values for a given pair of host and symbiont trees, including one we established in the past. However, a major limitation of these methods is their inability to model the "invasion" of different host species by a same symbiont species (referred to as a spread event), which is often observed in symbiotic relations. Indeed, many symbionts are generalist. For instance, the same species of insect may pollinate different species of plants. In a paper currently in preparation, we propose a method, called AMOCOALA, which for a given pair of host and symbiont trees, estimates the frequency of the cophylogenetic events, in presence of spread events, based on an approximate Bayesian computation (ABC) approach that may be more efficient than a classical likelihood method. The algorithm that we propose on one hand provides more confidence in the set of costs to be used for a given pair of host and symbiont trees, while on the other hand, it allows to estimate the frequency of the events even in the case of large datasets. We evaluated our method on both synthetic and real datasets.

Co-divergence and tree topology In reconstructing the common evolutionary history of hosts and symbionts, the current method of choice is the phylogenetic tree reconciliation. In this model, we are given a host tree H, a symbiont tree S, and a function σ mapping the leaves of S to the leaves of H and the goal is to find, under some biologically motivated constraints, a reconciliation, that is a function from the vertices of S to the vertices of H that respects σ and allows the identification of biological events such as co-speciation, duplication and host switch. The maximum co-divergence problem consists in finding the maximum number of co-speciations in a reconciliation. This problem is NP-hard for arbitrary phylogenetic trees and no approximation algorithm is known. In [14], we considered the influence of tree topology on the maximum co-divergence problem. In particular, we focused on a particular tree structure, namely caterpillar, and showed that in this case the heuristics that are mostly used in the literature provide solutions that can be arbitrarily far from the optimal value. We then proved that finding the max co-divergence is equivalent to computing the maximum length of a subsequence with certain properties of a given permutation. This equivalence leads to two consequences: (i) it shows that we can compute efficiently in polynomial time the optimal time-feasible reconciliation, and (ii) it can be used to understand how much the tree topology influences the value of the maximum number of co-speciations.

6.5. Axis 4: Human and animal health

Rare disease studies Minor intron splicing plays a central role in human embryonic development and survival. Indeed, biallelic mutations in RNU4ATAC, transcribed into the minor spliceosomal U4atac snRNA, are responsible for three rare autosomal recessive multimalformation disorders named Taybi-Linder (TALS/MOPD1), Roifman (RFMN), and Lowry-Wood (LWS) syndromes, which associate numerous overlapping signs of varying severity. Although RNA-seq experiments have been conducted on a few RFMN patient cells, none have been performed in TALS, and more generally no in-depth transcriptomic analysis of the 700 human genes containing a minor (U12-type) intron had been published as yet. We thus sequenced RNA from cells derived from five skin, three amniotic fluid, and one blood biosamples obtained from seven unrelated TALS cases and from age- and sex-matched controls. This allowed us to describe for the first time the mRNA expression and splicing profile of genes containing U12-type introns, in the context of a functional minor spliceosome. Concerning RNU4ATAC-mutated patients, we showed in [15] that as expected, they display distinct U12-type intron splicing profiles compared to controls, but that rather unexpectedly the mRNA expression levels are mostly unchanged. Furthermore, although U12-type intron missplicing concerns most of the expressed U12 genes, the level of U12-type intron retention is surprisingly low in fibroblasts and amniocytes, and much more pronounced in blood cells. Interestingly, we found several occurrences of introns that can be spliced using either U2, U12, or a combination of both types of splice site consensus sequences, with a shift towards splicing using preferentially U2 sites in TALS patients' cells compared to controls.

This work is part of the PhD of Audric Cologne [1] defended in October 2019.

Cancer studies Circular RNAs (circRNAs) are a class of RNAs that is under increasing scrutiny, although their functional roles are debated. In [30], we analysed RNA-seq data of 348 primary breast cancers and developed a method to identify circRNAs that does not rely on unmapped reads or known splice junctions. We identified 95,843 circRNAs, of which 20,441 were found recurrently. Of the circRNAs that match exon boundaries of the same gene, 668 showed a poor or even negative ($R < 0.2$) correlation with the expression level of the linear gene. An *In silico* analysis showed that only a minority (8.5%) of circRNAs could be explained by known splicing events. Both these observations suggest that specific regulatory processes for circRNAs exist. We confirmed the presence of circRNAs of CNOT2, CREBBP, and RERE in an independent pool of primary breast cancers. We identified circRNA profiles associated with subgroups of breast cancers and with biological and clinical features, such as amount of tumour lymphocytic infiltrate and proliferation index. siRNA-mediated knockdown of circCNOT2 was shown to significantly reduce viability of the breast cancer cell lines MCF-7 and BT-474, further underlining the biological relevance of circRNAs. Furthermore, we found that circular, and not linear, CNOT2 levels are predictive for progression-free survival time to aromatase inhibitor (AI) therapy in advanced breast cancer patients, and found that circCNOT2 is detectable in cell-free RNA from plasma. We showed that circRNAs are abundantly present, show characteristics of being specifically regulated, are associated with clinical and biological properties, and thus are relevant in breast cancer.

Other cancer studies have concerned the automatic discovery of the 100-miRNA signature for cancer classification [21], an Integrative and comparative genomic analysis to identify clinically relevant pulmonary carcinoid groups and unveil the supra-carcinoids [5], [complete with 2 papers not yet entered in Hal], and finally the investigation of new therapeutic interventions that are needed to increase the immunogenicity of tumours and overcome the resistance to these immuno-therapies [29].

Infection studies *Mycoplasma hyopneumoniae* is an economically devastating pathogen in the pig farming industry, however little is known about its relation with the swine host. To improve our understanding on this interaction, we infected epithelial cells with *M. hyopneumoniae* to identify the effects of the infection on the expression of swine genes and miRNAs. In addition, we identified miRNAs differentially expressed (DE) in the extracellular milieu and in exosome-like vesicles released by infected cells. A total of 1,268 genes and 170 miRNAs were DE post-infection ($p<0.05$). We identified the up-regulation of genes related to redox homeostasis and antioxidant defense, most of them putatively regulated by the transcription factor NRF2. Down-regulated genes were enriched in cytoskeleton and ciliary function, which could partially explain *M. hyopneumoniae* induced ciliostasis. Our predictions showed that DE miRNAs could be regulating the

aforementioned functions, since we detected down-regulation of miRNAs predicted to target antioxidant genes and up-regulation of miRNAs targeting ciliary and cytoskeleton genes. Based on these observations, *M. hyopneumoniae* seems to elicit an antioxidant response induced by NRF2 in infected cells; in addition, we propose that ciliostasis caused by this pathogen might be related to down-regulation of ciliary genes. The paper presenting these results has been submitted and is in revision.

Others Besides the above, a first step towards deep learning assisted genotype-phenotype association in whole genome-sized data has been explored in the context of predicting amyotrophic lateral sclerosis [34].

GENSCALE Project-Team

7. New Results

7.1. Algorithms & Methods

7.1.1. SV genotyping

Participants: Dominique Lavenier, Lolita Lecompte, Claire Lemaitre, Pierre Peterlongo.

Structural variations (SV) are genomic variants of at least 50 base pairs (bp) that can be rearranged within the genome and thus can have a major impact on biological processes. Sequencing data from third generation technologies have made it possible to better characterize SVs. Although many SV callers have been published recently, there is no published method to date dedicated to genotyping SVs with this type of data. Variant genotyping consists in estimating the presence and ploidy or absence of a set of known variants in a newly sequenced individual. Thus, in this paper, we present a new method and its implementation, SVJedi, to genotype SVs with long reads. From a set of known SVs and a reference genome, our approach first generates local sequences representing the two possible alleles for each SV. Long read data are then aligned to these generated sequences and a careful analysis of the alignments consists in identifying only the informative ones to estimate the genotype for each SV. SVJedi achieves high accuracy on simulated and real human data and we demonstrate its substancial benefits with respect to other existing approaches, namely SV discovery with long reads and SV genotyping with short reads [23], [24], [35]. SVJedi is implemented in Python and available at https://github.com/llecompte/SVJedi.

7.1.2. Genome assembly of targeted organisms in metagenomic data

Participants: Wesley Delage, Fabrice Legeai, Claire Lemaitre.

In this work, we propose a two-step targeted assembly method tailored for metagenomic data, called MinYS (for MineYourSymbiont). 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 novel 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 MinYS 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 10 times faster than classical *de novo* metagenomics assemblers and it is able to recover large structural variations co-existing in a sample. MinYS is a Python3 pipeline, distributed on github (https://github.com/cguyomar/MinYS) and as a conda package in the bio-conda repository [22].

7.1.3. SimkaMin: subsampling the kmer space for efficient comparative metagenomics

Participants: Claire Lemaitre, Pierre Peterlongo.

SimkaMin [12] is a quick comparative metagenomics tool with low disk and memory footprints, thanks to an efficient data subsampling scheme used to estimate Bray-Curtis and Jaccard dissimilarities. One billion metagenomic reads can be analyzed in less than 3 minutes, with tiny memory (1.09 GB) and disk ($\sim$0.3 GB) requirements and without altering the quality of the downstream comparative analyses, making of SimkaMin a tool perfectly tailored for very large-scale metagenomic projects.

7.1.4. Haplotype reconstruction: phasing co-localized variants

Participants: Mohammed Amin Madoui, Pierre Peterlongo.

In collaboration with Amin Madoui from the Genoscope (CEA), we develop a new methodology to reconstruct haplotypes or strain genomes directly from raw sequencing set of (metagenomic) reads. The goal is to propose long assembled sequences (i.e. complete genomes are not mandatory) such that each assembled sequence belongs to only one sequenced chromosome and is not a consensus of several similar sequences. Downstream, this enables to perform population genomics analyses.

The key idea is to use the DiscoSnp [10] output, detecting set of variant alleles that are co-localized on input reads or pairs of input reads. Then we finally reconstruct set of sequences that are as parsimonious as possible with those observations.

7.1.5. Finding all maximal perfect haplotype blocks in linear time

Participant: Pierre Peterlongo.

Recent large-scale community sequencing efforts allow at an unprecedented level of detail the identification of genomic regions that show signatures of natural selection. However, traditional methods for identifying such regions from individuals' haplotype data require excessive computing times and therefore are not applicable to current datasets. In 2019, Cunha et al. (Proceedings of BSB 2019) suggested the maximal perfect haplotype block as a very simple combinatorial pattern, forming the basis of a new method to perform rapid genome-wide selection scans. The algorithm they presented for identifying these blocks, however, had a worst-case running time quadratic in the genome length. It was posed as an open problem whether an optimal, linear-time algorithm exists. We gave two algorithms that achieve this time bound, one which is conceptually very simple and uses suffix trees and a second one using the positional Burrows-Wheeler Transform, that is very efficient also in practice [20].

7.1.6. Short read correction

Participant: Pierre Peterlongo.

We propose a new approach for the correction of NGS reads. This approach is based on the construction of a clean de Bruijn graph in which the correction is made at the contig level. In a second step, original reads are mapped on this graph, allowing to correct the original reads [16].

7.1.7. Large-scale kmer indexation

Participants: Téo Lemane, Pierre Peterlongo.

In the SeqDigger ANR project framework (see dedicated Section), we aim to index TB or PB of genomic sequences, assembled or not. The central idea is to assign any kmer (word of length k) to the set of indexed dataset it belongs to. For doing this we have proposed a method that improves one of the state of the art algorithm (HowDeSBT [38]) by optimizing the way kmers are counted and represented [36].

7.1.8. Proteogenomics workflow for the expert annotation of eukaryotic genomes

Participant: Pierre Peterlongo.

Accurate structural annotation of genomes is still a challenge, despite the progress made over the past decade. The prediction of gene structure remains difficult, especially for eukaryotic species, and is often erroneous and incomplete. In [15], we proposed a proteogenomics strategy, taking advantage of the combination of proteomics datasets and bioinformatics tools, to identify novel protein coding-genes and splice isoforms, to assign correct start sites, and to validate predicted exons and genes.

7.1.9. Gap-filling with linked-reads data

Participants: Anne Guichard, Fabrice Legeai, Claire Lemaitre, Arthur Le Bars, Pierre Peterlongo.

We develop a novel approach for filling assembly gaps with linked reads data (typically 10X Genomics technology). The approach is based on local assembly using our tool MindTheGap [9], and takes advantage of barcode information to reduce the input read set in order to reduce the de Bruijn graph complexity. The approach is applied to recover the genomic structure of a 1.3 Mb locus of interest in a dozen of re-sequenced butterfly genomes (*H. numata*) in the Supergene ANR project context.

7.1.10. Statistically Significant Discriminative Patterns Searching

Participants: Gwendal Virlet, Dominique Lavenier.

We propose a novel algorithm, called SSDPS, to discover patterns in two-class datasets. The algorithm, developed in collaboration with the LACODAM Inria team, owes its efficiency to an original enumeration strategy of the patterns, which allows to exploit some degrees of anti-monotonicity on the measures of discriminance and statistical significance. Experimental results demonstrate that the performance of the algorithm is better than others. In addition, the number of generated patterns is much less than the number of the other algorithms. An experiment on real data also shows that SSDPS efficiently detects multiple SNPs combinations in genetic data [27].

7.2. Optimisation

7.2.1. Chloroplast genome assembly

Participants: Sébastien Francois, Roumen Andonov, Dominique Lavenier.

This research focuses on the last two stages of *de novo* genome assembly, namely, scaffolding and gap-filling, and shows that they can be solved as part of a single optimization problem. Our approach is based on modeling genome assembly as a problem of finding a simple path in a specific graph that satisfies as many distance constraints as possible encoding the read-pair insert-size information. We formulate it as a mixed-integer linear programming (MILP) problem and apply an optimization solver to find the exact solutions on a benchmark of chloroplast genomes. We show that the presence of repetitions in the set of unitigs is the main reason for the existence of multiple equivalent solutions that are associated to alternative subpaths. We also describe two sufficient conditions and we design efficient algorithms for identifying these subpaths. Comparisons of the results achieved by our tool with the ones obtained with recent assemblers are also presented [11].

7.2.2. Integer Linear Programming for Metabolic Networks

Participants: Kerian Thuillier, Roumen Andonov.

Metabolic networks are a helpful tool to represent and study cell metabolisms. They contain information about every reaction occurring inside an organism. However, metabolic networks of poorly studied species are often incomplete. It is possible to complete these networks with knowledge of other well-known species.

In this study, we present a new linear programming approach for the problem of topological activation in metabolic networks based on flows and the Miller, Tucker and Zemlin (MTZ) formulation for solving the longest path problem. We developed a tool called *Flutampl* with AMPL (A Mathematical Programming Language). It returns optimal solutions for the hybrid completion directly from *sbml* files (the data format used for modelling metabolic networks) [37].

7.2.3. Integer Linear Programming for De novo Long Reads Assembly

Participants: Victor Epain, Roumen Andonov, Dominique Lavenier.

To tackle the de novo long read assembly problem, we investigate a new 2-step method based on integer linear programming. The first step orders the long reads and the second one generates a consensus sequence. Each step is based on a different IPL specification. In 2019, we focused on step 1: long reads are first compared to build an overlapping graph. Then we use integer linear programming to find the heaviest path in a graph $G = (V, E, \lambda)$, where V is the vertices set corresponding to the long reads, E the edge set associated to the overlaps between long reads and λ the overlap length. For large graph, V is partitioned into several parts, each one is solved independently, and the solutions are merged together. Preliminary experimentation show that bacteria assemblies can be successfully solved in a few minutes [31].

7.3. Experiments with the MinION Nanopore sequencer

7.3.1. Storing information on DNA Molecules

Participants: Dominique Lavenier, Emeline Roux, Ariane Badoual.

In 2019, we started a new research activity aiming to explore the possibility to use the DNA molecules as a storage medium. We designed a complete DNA storage system based on the Oxford Nanopore sequencing technology and performed several experimentations by sequencing several synthetized DNA fragments ranging from 500 to 1,000 bp. These sequences have been designed with ad-hoc coding to prevent specific sequencing errors of the nanopore technology such as indel errors in homo-polymer sequences [29] [34]. These real experiments demonstrate that a text encoded into the DNA alphabet, then synthetized into DNA molecules, sequenced with the MinION, and finally processed using bioinformatics approaches can be fully recovered [28].

7.3.2. Identification of bacterial strains

Participants: Jacques Nicolas, Emeline Roux, Grégoire Siekaniec, Clara Delahaye.

Our aim is to provide rapid algorithms for the identification of bacteria at the finest taxonomic level. We have developed an expertise in the use of the MinIon long read technology and have produced and assembled many genomes for lactic bacteria in cooperation with INRA STLO, which have been made available on the Microscope platform at Genoscope (http://www.genoscope.cns.fr/agc/microscope/). We have developed a first classifier that demonstrates the possibility to identify isolated strains with spaced seed indexing of the noisy long reads produced by the MinIon.

7.4. Benchmarks and Reviews

7.4.1. Evaluation of error correction tools for long Reads

Participants: Lolita Lecompte, Pierre Peterlongo.

Long read technologies, such as 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 [17]. ELECTOR is freely available at https://github.com/kamimrcht/ELECTOR.

7.4.2. Evaluation of insertion variant callers on real human data

Participants: Wesley Delage, Claire Lemaitre.

Insertion variants are one of the most common types of structural variation. Although such variants have many biological impacts on species evolution and health, they have been understudied because they are very difficult to detect with short read re-sequencing data. Recently, with the commercialization of novel long reads technologies, insertion variants are finally being discovered and referenced in human populations. Thanks to several international efforts, some gold standard call sets have been produced in 2019, referencing tens of thousands insertions. On these datasets, all existing short-read insertion variant callers, including our own method MindTheGap [9] which overtook others on simulated data, can reach at most 5 to 10 % of the referenced insertion variants. In this work, we propose a classification of the different types of insertion variants, based on the genomic context of the insertion site and the levels of duplication contained in the inserted sequence or within its breakpoints. In a detailed benchmark, we then analyze which of these types are the most impacted by the low recall of existing methods. Finally, by simulating various identified factors of difficulty, we investigate the causes of low recall and how these can be bypassed or improved in existing algorithms.

7.4.3. Modeling activities in cooperation with Inria project Dyliss

Participant: Jacques Nicolas.

J. Nicolas has maintained a partial activity with its previous research team Dyliss. In this framework, we have explored the use of Formal Concept Analysis (FCA) to ease the analysis of biological networks. The PhD thesis of L. Bourneuf on graph compression using FCA, defended this year, has introduced a new extension of FCA for this purpose, working on triplet concepts, which correspond to overlapping bicliques in graphs. The search space of concepts for graph compression has been presented in [21]. FCA applied to data on the steady states of a Boolean network and the dependencies between its proteins allowed to build a classifier used to analyze the states according to the phenotypic signatures of its network components. We have identified variants to the phenotypes and characterized hybrid phenotypes [19].

7.5. Bioinformatics Analysis

7.5.1. Genomics of Brassicaceae and agro-ecosystems insects

Participant: Fabrice Legeai.

Through its long term collaboration with INRA IGEPP, and its support to the BioInformatics of Agroecosystems Arthropods platform (http://bipaa.genouest.org), GenScale is involved in various genomic projects in the field of agricultural research. First, on plant genomics, we helped to identify duplicated copies of genes and repeated elements in the Brassica genomes [14]. Then, on major agricultural pests or their natural ennemies such as parasitoids, we conducted large scale analyses on the expression of effector genes involved in the adaptation of pea aphids to their host-plants [13]. Finally, we explored the expression of genes related to the virus machinery of bathyplectes parasitoids wasp of the alfalfa weevil [18].

7.5.2. Structural genome analysis of S. pyogenes strains

Participants: Emeline Roux, Dominique Lavenier.

The *S. pyogenes* bacteria is responsible for many human infections. With the increase in the prevalence of infections (750 million infections per year worldwide and 4th in terms of mortality from bacterial infection), a better understanding of adaptive and evolutionary mechanisms at play in this bacteria is essential. The molecular characterization of the different strains is done by the *emm* gene. A statistical analysis of the different types of *emm* on the Brittany population shows 3 main dynamics: sporadic types, endemic types or epidemic types. The last case was observed in Brittany for the type *emm75* between 2009 and 2017. Two hypotheses can be considered: (1) the emergence of a new subtype or winning clone in an unimmunized population; (2) increased pathogenicity through genetic evolution of the strains, including the acquisition of new virulence factors. In collaboration with the microbiology department of the Rennes Hospital, we sequenced more than 30 *S. pyogenesemm75* strains (Oxford Nanopore MinION sequencing) in order to study the dynamic of the epidemic thought their structural genomic variation.

7.5.3. Linking allele-specific expression and natural selection in wild populations

Participants: Mohammed Amin Madoui, Pierre Peterlongo.

Allele-specific expression (ASE) is now a widely studied mechanism at cell, tissue and organism levels. However, population-level ASE and its evolutionary impacts have still never been investigated. Here, we hypothesized a potential link between ASE and natural selection on the cosmopolitan copepod *Oithona similis*. We combined metagenomic and metatranscriptomic data from seven wild populations of the marine copepod *O. similis* sampled during the Tara Oceans expedition. We detected 587 single nucleotide variants (SNVs) under ASE and found a significant amount of 152 SNVs under ASE in at least one population and under selection across all the populations. This constitutes a first evidence that selection and ASE target more common loci than expected by chance, raising new questions about the nature of the evolutionary links between the two mechanisms [33].

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 [13], 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. 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. An article 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, was published in *BMC Bioinformatics* this year [22]. We notably show that the expression pattern in different growth media, supporting different growth rates, corresponds to the pattern expected for a constitutive gene.

6.2. 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 6.1 . 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. The investigation of the problem of reconstructing promoter activity statistics from reporter gene population snapshot data has led 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. The method has been demonstrated on simulated data. Theoretical as well as simulation resuts have been published in *Automatica* this year [15], and will be the object of application to real data.

One of the key limitations of the method is the assumption of stationary promoter activation statistics. In the context of controlled gene expression processes, this may hamper applicability of the method. In response to this, an extension of the method for so-called modulated processes (stationary processes reshaped by a time-varying control input), has been developed and demonstrated on simulations of controlled gene expression. Results were submitted for possible presentation and publication in the proceedings of the IFAC world congress 2020.

6.3. Mathematical analysis of structured branching populations

The investigation of cellular populations at the single-cell level has led to the discovery of important phenomena, such as the co-occurrence of different phenotypes in an isogenic population. Novel experimental techniques, such as time-lapse fluorescence microscopy combined with the use of microfluidic devices (Section 3.2), enable one to take the investigation further by providing time-course profiles of the dynamics of individual cells over entire lineage trees. The development of models that take into account the genealogy of individual cells is an important step in the study of inheritance in bacterial population. As a prerequisite, the efficient analysis of single-cell data relies on the mathematical analysis of those models.

Structured branching processes allow for the study of populations, where the lifecycle of each cell is governed by a given characteristic or trait, such as the concentration of a specific protein inside the cell. The dependence of bacterial phenotypes like cell division times or ageing on such characteristics has been investigated by Aline Marguet using mathematical analysis of the underlying processes. To understand the long-time behavior of structured branching populations, the process describing the trait of a typical individual along its ancestral lineage, called auxiliary process [21] and its asymptotic behavior play a key role. In a publication in *ESAIM: Probability and Statistics* that appeared this year [20], we proved that the empirical measure of the structured branching process converges to the mean value of this auxiliary process. The approach relies on ergodicity arguments for the time-inhomogeneous auxiliary Markov process. The novelty compared to existing spectral methods is that our method allows to consider processes with time-varying rates for the modeling of changing environments. For example, we studied the case of a size-structured population in a varying environment and proved the convergence of the empirical measure in this specific case.

In collaboration with Charline Smadi (IRSTEA Grenoble), Aline Marguet also investigated the long-time behavior of a general class of branching Markov processes. This work, which has been submitted for publication [27], aims at understanding the link between the dynamic of the trait and the dynamic of the population. In the case of a trait modelling the proliferation of a parasite infection in a cellular population, we exhibit conditions on the dynamics of the parasites to survive in the population, despite the cellular divisions that dilute the number of parasites in each cell.

The study of the asymptotic behavior of general non-conservative semigroups is important for several aspects of branching processes, especially to prove the efficiency of statistical procedures. Vincent Bansaye from École Polytechnique, Bertrand Cloez from INRA Montpellier, Pierre Gabriel from Université Versailles Saint-Quentin, and Aline Marguet obtained necessary and sufficient conditions for uniform exponential contraction in weighted total variation norm of non-conservative semigroups. It ensures the existence of Perron eigenelements and provides quantitative estimates of spectral gaps, complementing Krein-Rutman theorems and generalizing recent results relying on probabilistic approaches. This work was submitted for publication this year [26].

6.4. Inference of gene expression parameters on lineage trees

As explained in the previous section, recent technological developments have made it possible to obtain time-course single-cell measurements of gene expression as well as the associated lineage information. However,

most of the existing methods for the identification of mathematical models of gene expression are not well-suited to single-cell data and make the simplifying assumptions that cells in a population are independent, thus ignoring cell lineages. The development of statistical tools taking into account the correlations between individual cells will allow in particular for the investigation of inheritance of traits in bacterial populations.

In the framework of structured branching processes, we studied the statistical reconstruction of parameters. We considered the problem of estimating the division rate from the observations of the trait of the cells at birth. Previous works on the subject considered deterministic dynamics for the evolution of the trait. In collaboration with Marc Hoffmann (Université Paris Dauphine), Aline Marguet investigated the case of a trait evolving according to a diffusion process. The study of the asymptotic behavior of the tagged-chain, corresponding to the trait of a uniformly chosen individual, allowed us to prove the convergence of the empirical measure of the branching process, and the asymptotic minmax efficiency of nonparametric estimators for the density of the transition kernel and the invariant measure of the tagged-chain. For the estimation of the division rate, we proved in a parametric framework the asymptotic efficiency of a standard maximum likelihood proxy estimation. Finally, we demonstrate the validity of our approach on simulated datasets. The results of this work were published in *Stochastic Processes and their Applications* [17].

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], Eugenio Cinquemani, Marc Lavielle (XPOP, Inria Saclay–Île-de-France) and Aline Marguet developed a new model and a method for inference from data for gene expression along tree where the kinetic expression parameters are assumed to be inherited from the mother cell in an autoregressive way. This model generalizes the state-of-the-art mixed-effect models to the case of lineage trees. We implemented the inference procedure in Julia and proved that it provides unbiased estimates of the parameters. The application to the data of osmotic shock response by yeast show that the correlation between the parameter of a cell and its daughter is of 0.6 according to our model, leading to new biological questions such as the understanding of the origin of this inheritance. The results of this study were presented at the major bioinformatics conference ISMB/ECCB 2020 and published in the associated special issue of *Bioinformatics* [19].

6.5. Modeling and inference of RNA degradation

The ability to rapidly respond to changing nutrient availability is crucial for *E. coli* to survive in many environments including the gut. Reorganization of gene expression is the first step for bacteria to adjust their metabolism accordingly. It involves fine-tuning of both transcription and mRNA stability by dedicated regulatory interactions. While transcriptional regulation has been largely studied, the role of mRNA stability during a metabolic switch is poorly understood.

This question was addressed in the framework of the PhD thesis of Manon Morin funded by an INRA-Inria grant. Using combined genome-wide transcriptome and mRNA decay analyses, Manon Morin, Delphine Ropers and colleagues from the Toulouse Biotechnology Institute (ex-LISBP, INRA/INSA Toulouse) investigated the role of mRNA stability in the response of *E. coli* to nutrient changes. They demonstrated that transcript stability increases along metabolic transitions representative of the carbon source fluctuations, the glucose-acetate-starvation transition [9], [10]. Most of the stabilization occurs at glucose-acetate transition when glucose is exhausted. Stabilized mRNAs remain stable during acetate consumption and carbon starvation. Meanwhile, expression of most genes is downregulated. Metabolic control analysis showed that most of gene expression regulation is driven by changes in transcription. Post-transcriptional regulations appear to be important for genes involved in bacterial response to nutrient starvation. These results have been further developed in a paper recently submitted to a biology journal.

The observation of a global stabilization of cellular mRNAs during adaptation to carbon source depletion raises questions about the regulatory mechanisms at work. Known regulators of mRNA stability such as the protein Hfq, the carbon storage regulator Csr, and several small regulatory RNAs, specifically target mRNAs. Are these regulatory mechanisms sufficient to explain the systematic adjustment of mRNA half-lives? The collaboration with Muriel Cocaign-Bousquet and colleagues from the Toulouse Biotechnology Institute has been pursued to answer these questions, in the context of the PhD thesis of Thibault Etienne,

funded by an INRA-Inria PhD grant. The objective is to develop models able to explain how cells coordinate their physiology and the functioning of the degradation machinery following environmental changes. In a paper submitted this year, Thibault Etienne, Delphine Ropers and Muriel Cocaign-Bousquet investigate the possibility that competition between mRNAs for their binding to the degradation machinery is an important mechanism for the regulation of mRNA half-lives. They develop a mathematical model of mRNA degradation and assess the role of competitive effects on mRNA degradation kinetics by numerical simulation and sensitivity analysis. Competition appears to globally increase the stability of cellular mRNAs and to amplify the effect of post-transcriptional regulation. In a follow-up study, the model is currently being used to interpret large data sets corresponding to the degradation kinetics of 4254 mRNAs in *E. coli* cells growing in four different environmental conditions.

6.6. 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 in combination with methods from optimal control theory. In a paper published in the *Journal of Mathematical Biology* this year [24], 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, were presented at the European Control Conference (ECC 2019) in Naples this year and published in the proceedings [25].

6.7. Bacterial growth inhibition by acetate

High concentrations of organic acids such as acetate inhibit growth of *Escherichia coli* and other bacteria. This phenomenon is of interest for understanding bacterial physiology but is also of practical relevance. Growth inhibition by organic acids underlies food preservation and causes problems during high-density fermentation in biotechnology. The development of new approaches for the relief of growth inhibition by acetate during high-density fermentation of *E. coli* is one of the motivating assumptions for the work of IBIS in the IPL project COSY (Sections 7.2 and 6.8 below).

What causes growth inhibition by acetate? Classical explanations invoke the uncoupling effect of acetate and the establishment of an anion imbalance. During his PhD thesis, Stéphane Pinhal investigated an alternative hypothesis: the perturbation of acetate metabolism due to the inflow of excess acetate. In an experimental and modelling study published in the *Journal of Bacteriology* [23], Stéphane Pinhal, Delphine Ropers, Hans Geiselmann, and Hidde de Jong developed a set of isogenic strains that remove different parts of the metabolic network involved in acetate metabolism. Analysis of these strains revealed that the inflow of acetate accounts for 20% of the growth-inhibitory effect through a modification of the acetyl phosphate concentration. While the study does not provide a definite answer to the question of what accounts for the remaining 80% of the reduction in growth rate, some of the observations argue against a prominent role of uncoupling in growth inhibition by acetate in the conditions tested.

6.8. Modeling synthetic microbial communities for improving productivity

Modelling, analysis and control of microbial community dynamics is a fast-developing subject with great potential implications in the understanding of natural processes and the enhancement of biootechnological processes. Within the IPL COSY (Section 7.2), we picked up the challenge to design and investigate the dynamics of synthetically engineered microbial communities with a consortium of Inria partners. In IBIS, in particular, we are addressing the design of a bacterial community of two *E.coli strains*, mimicking mutualistic relationships found in nature, and with the potential to outperform a single producer strain in the production of a heterologous protein. During the post-doctoral stay of Marco Mauri, we developed an ODE model of the key growth phenotypes of the community and their interactions, calibrated the model on literature data, and analysed the model for an in-depth understanding of the conditions supporting coexistence and of the tradeoffs encountered in this production process. The results are presented in a paper submitted for publication this year and will be tested experimentally in the framework of the recently-started PhD project of Maaike Sangster. Analysis of optimal community control problems as well as design and deployment of optimal control strategies will follow in synergy with other IPL COSY partners.

6.9. Detection of small non-coding RNAs

Small non-coding RNAs (sRNAs) regulate numerous cellular processes in all domains of life. Several approaches have been developed to identify them from RNA-seq data, which are efficient for eukaryotic sRNAs but remain inaccurate for the longer and highly structured bacterial sRNAs. Together with colleagues from INSA de Lyon, Stéphan Lacour developed APERO, a new algorithm to detect small transcripts from paired-end bacterial RNA-seq data. This algorithm is based on a novel approach, which does not start from the read coverage distribution, but analyzes boundaries of individual sequenced fragments to infer the 5´ and 3´ ends of all transcripts. Validation of the algorithm on *Escherichia coli* and *Salmonella enterica* datasets, based on experimentally validated sRNAs, showed it to outperform all existing methods in terms of sRNA detection and boundary precision. Moreover, APERO was able to identify the small transcript repertoire of *Dickeya dadantii* including putative intergenic RNAs, 5´ UTR or 3´ UTR-derived RNA products and antisense RNAs. This work was published in *Nucleic Acids Research* this year [18]. APERO is freely available as an open source R package (https://github.com/Simon-Leonard/APERO). In other work, together with colleagues from the University of Salento, Lecce (Italy), Stéphan Lacour contributed to RHOTERMPREDICT, an algorithm for predicting Rho-dependent transcription terminators in bacterial genomes [16].

LIFEWARE Project-Team

7. New Results

7.1. CRN design by program compilation

Participants: Elisabeth Degrand, François Fages, Mathieu Hemery, Wei-Chih Huang [NTU Taiwan], Sylvain Soliman.

One goal of synthetic biology is to implement useful functions with biochemical reactions, either by reprogramming living cells or programming artificial vesicles. In this perspective, we consider Chemical Reaction Networks (CRN) as a programming language, and investigate the CRN program synthesis problem. Recent work has shown that CRN interpreted by differential equations are Turing-complete and can be seen as analog computers where the molecular concentrations play the role of information carriers. Any real function that is computable by a Turing machine in arbitrary precision can thus be computed by a CRN over a finite set of molecular species. The proof of this result gives a numerical method to generate a finite CRN for implementing a real function presented as the solution of a Polynomial Initial Values Problem (PIVP).

The compilation of high-level imperative programming languages in CRN requires however an efficient implementation of program control flows using threshold functions. The biochemical threshold function is also a crucial component in the biosensor circuits to be deployed in living cells or synthetic vesicles for disease diagnosis. In [5], based on the zero-order ultrasensitivity, we propose an economic biochemical implementation of threshold functions with reconfigurable threshold values. We show that the so-constructed threshold function module well approximates the unit step function and allows robust composition with other function modules for complex computation tasks. This is now implemented in BIOCHAM-4 for the compilation of sequentiality and conditionals in CRNs.

7.2. CRN design by artificial evolution

Participants: Elisabeth Degrand, François Fages, Mathieu Hemery, Sylvain Soliman.

In [4], [12], we study an alternative method based on artificial evolution to build a CRN that approximates a real function given on finite sets of input values. We present a nested search algorithm that evolves the structure of the CRN and optimizes the kinetic parameters at each generation. We evaluate this algorithm on the Heaviside and Cosine functions both as functions of time and functions of input molecular species. We then compare the CRN obtained by artificial evolution both to the CRN generated by the numerical method from a PIVP definition of the function, and to the natural CRN found in the BioModels repository for switches and oscillators.

On a Heaviside function of time, the results obtained by artificial evolution lead to a remarkably simple CRN of 3 molecular species and 5 reactions with double catalysts which provide a very stiff transition although using mass action law kinetics. This solution is more economical than the CRN generated by the PIVP method for sigmoid functions. On a Heaviside function of input, the CRN found by evolution are slightly more complicated than the bistable switch found in cell cycle CRN for instance, but much less complex than the MAPK signaling network that plays a similar role.

On the cosine function of time, the best CRN found by evolution contains an annihilation reaction similar to the CRN generated by the numerical method for positive and negative variables, but one less reaction thanks to an intriguing non symmetric use of the two variables which preserves the limit cycle. Interestingly, the evolved and the PIVP generated structures could be compared to prokaryote and eukaryote models of the circadian clock found in BioModels.

On the cosine function of input, a CRN surprisingly emerges with the structure of the CRN for cosine function of time, using the same trick as for PIVP compilation to stop time at the desired input value.

In [2], we use a genetic algorithm to evolve biochemical networks displaying a direct logarithmic response. Numerous biological systems are known to harbour a form of logarithmic behaviour, from Weber's law to bacterial chemotaxis. Working on a logarithmic scale allows the organism to respond appropriately to large variations in a given input at a modest cost in terms of metabolism. Interestingly, a quasi-perfect log-response implemented by the same simple core network evolves in a convergent way across our different replications. The best network is able to fit a logarithm over 4 order of magnitude with an accuracy of the order of 1%. At the heart of this network, we show that a logarithmic approximation may be implemented with one single non-linear interaction, that can be interpreted either as a phosphorylation or as a ligand induced multimerization and provide an analytical explanation of the effect. Biological log-response might thus be easier to implement than usually assumed.

7.3. CRN learning from data time series

Participants: François Fages, Jeremy Grignard, Julien Martinelli, Sylvain Soliman.

With the automation of biological experiments and the increase of quality of single cell data that can now be obtained by phosphoproteomic and time lapse videomicroscopy, automating the building of mechanistic models from these data time series becomes conceivable and a necessity for many new applications. While learning numerical parameters to fit a given model structure to observed data is now a quite well understood subject, learning the structure of the model is a more challenging problem that previous attempts failed to solve without relying quite heavily on prior knowledge about that structure. In [8], [7], we consider mechanistic models based on chemical reaction networks (CRN) with their continuous dynamics based on ordinary differential equations, and finite time series about the time evolution of concentration of molecular species for a given time horizon and a finite set of perturbed initial conditions. We present a greedy heuristics unsupervised statistical learning algorithm to infer reactions with a time complexity for inferring one reaction in $O(t.n^2)$where n is the number of species and t the number of observed transitions in the traces. We evaluate this algorithm both on simulated data from hidden CRNs, and on real videomicroscopy single cell data about the circadian clock and cell cycle progression of NIH3T3 embryonic fibroblasts. In all cases, our algorithm is able to infer meaningful reactions, though generally not a complete set for instance in presence of multiple time scales or highly variable traces.

7.4. CRN reductions

Participants: Orianne Bargain, Eléonore Bellot, François Fages, Eva Philippe, Sylvain Soliman.

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 using constraint programming or SAT solving techniques. Nevertheless, establishing whether two models are linked through a SEPI is an NP-complete problem which can be computationally expensive in some practical cases. Furthermore, the number of SEPIs can be huge, and some of them may not have a biological interpretation. In [11], we have improved the SEPI framework in this respect in three ways: by introducing optimization criteria to restrict the set of solutions, by restricting merge operations to some notion of neighborhood, and by preprocessing the CRN graphs in normal form in order to eliminate some common model reduction patterns.

Furthermore, in the framework of the ANR-DFG SYMBIONT project we investigate mathematical justification of SEPI reductions based on Tikhonov's theorem and their computation using tropical algebra methods and constraint programming techniques [0].

7.5. CRN modeling of biological systems

Participants: Auriane Cozic, Elisabeth Degrand, François Fages, Eléa Greugny, Jeremy Grignard, Constance Le Gac, Léna Le Quellec, Paul Remondeau, Sylvain Soliman.

[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.

This year, beyond implementation work on hybrid simulations in BIOCHAM and on antithetic feedback control in CRNs, we have started the computational modelling of three biological systems with important potential applications in biomedicine.

The first is about erythrocytes (i.e. red blood cells). Their most obvious function concerns the respiratory system since erythrocytes allow gas exchanges at the level of the organism by transporting dioxygen and carbon dioxide between the lungs and the tissues. However, red blood cells also have an important buffer function in the blood, which is necessary to keep blood pH in the physiological range. Modelling the red blood cells with CRNs using BIOCHAM gives us insight as to which biological objects are necessary to allow the cell to process its functions correctly. At the level of Systems Biology, it also allows us to understand the links between the different biological functions of erythrocytes.

The second concerns microtubules and their post-translational modifications involved in major cellular processes such as: mitosis, cardiomyocyte contraction, and neuronal differentiation. More precisely, in neurons, the post-translational modifications of detyrosination and tyrosination are crucial for neuronal plasticity, axon regeneration, recruitment and transports of proteins and correct neuronal wiring. We hypothesize that the decrease of density and length of microtubules and the loss of neuronal structures such as synapses, dendritic spine and growth cone which are correlated with the progressive cognitive decline [9,10] may be the consequence of the dysregulation of the cycle detyrosination/tyrosination in neurodegenerative disorder. This hypothesis is investigated in collaboration with Servier by combining experimental approaches with mathematical modelling.

The third concerns inflammation processes in skin. Skin protects the body against external agents, for instance pathogens, irritants, or UV radiation, that can trigger inflammation. Inflammation is a complex phenomenon that is classified in two main types, acute and chronic. They are distinguished by different parameters such as the duration, the underlying mechanisms, the components involved like the type of immune cells, and the nature and intensity of the associated clinical signs. The computational models developed in collaboration with Johnson&Johnson France, combine mathematical and multi-agent modelling using BIOCHAM and EPISIM modelling tools.

7.6. Automated Inference of Boolean models from molecular interaction maps

Participant: Sylvain Soliman.

Molecular interaction maps have emerged as a meaningful way of representing biological mechanisms in a comprehensive and systematic manner. However, their static nature provides limited insights to the emerging behavior of the described biological system under different conditions. Computational modeling provides the means to study dynamic properties through *in silico* simulations and perturbations.

In collaboration with Anna Niarakis (Université d'Évry, GenHotel) we have started developing the CaSQ Python package, by defining simplification rules and logical formulas for the inferred Boolean models according to the topology and the annotations of the starting molecular interaction maps. We used CaSQ to produce executable files of existing molecular maps notably a big map of the Rheumatoid Arthritis that is at the core of Évry team's work.

A publication on the inference process has already been submitted to Bioinformatics but work continues on the applications side to fine-tune the automatically generated model and analyze its dynamical properties.

7.7. Optimal control of an artificial microbial differentiation system for protein bioproduction

Participants: Élise Weill Duflos, Virgile Andréani, Chetan Aditya, Pierre Martinon [EPI Commands], Jakob Ruess, Grégory Batt, J. Frédéric Bonnans [EPI Commands].

The growth of microorganisms is controlled by strategies for the dynamical allocation of available resources over different cellular functions. Synthetic biology approaches are considered nowadays to artificially modify these strategies and turn microbial populations into biotechnological factories for the production of metabolites of interest. In our recent work, we have studied dynamics of microbial resource allocation and growth in terms of coarse-grained self-replicator models described by ordinary differential equations, and proposed artificial control strategies for the optimization of metabolite production based on the reengineering of resource allocation. In this contribution, we elaborated on our earlier results and further investigate synthetic resource allocation control strategies [9]. Using numerical simulation, we studied 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. Results provided novel insight into the most favorable synthetic control strategies.

7.8. Can optimal experimental design serve as a tool to characterize highly non-linear synthetic circuits?

Participants: Maxim Kryukov [Pasteur Institute], Arthur Carcano, Grégory Batt, Jakob Ruess.

One of the most crippling problems in quantitative and synthetic biology is that models aiming to describe the real mechanisms of biochemical processes inside cells typically contain too many unknown parameters to be reliably inferable from available experimental data. Recent years, however, have seen immense progress in the development of experimental platforms that allow not only to measure biological systems more precisely but also to administer external control inputs to the cells. Optimal experimental design has been identified as a tool that can be used to decide how to best choose these control inputs so as to excite the systems in ways that are particularly useful for learning the biochemical rate constants from the corresponding data. Unfortunately, the experiment that is best to learn the parameters of a system depends on the precise values of these parameters, which are naturally unknown at the time at which experiments need to be designed. Here, we used a recently constructed genetic toggle switch as a case study to investigate how close to the best possible experiment we can hope to get with the most widely used optimal design approaches in the field. We found that, for strongly nonlinear systems such as the toggle switch, reliably predicting the information that can be gained from a priori fixed experiments can be difficult if the system parameters are not known very precisely [6]. This suggests that a better strategy to guarantee informative experiments might be to use feedback control and to adjust the experimental plan in real time.

7.9. Molecular noise of innate immunity shapes bacteria-phage ecologies

Participant: Jakob Ruess.

Mathematical models have been used successfully at diverse scales of biological organization, ranging from ecology and population dynamics to stochastic reaction events occurring between individual molecules in single cells. Generally, many biological processes unfold across multiple scales, with mutations being the best studied example of how stochasticity at the molecular scale can influence outcomes at the population scale. In many other contexts, however, an analogous link between micro-and macro-scale remains elusive, primarily due to the challenges involved in setting up and analyzing multi-scale models. In [3], we employed such a model to investigate how stochasticity propagates from individual biochemical reaction events in the bacterial innate immune system to the ecology of bacteria and bacterial viruses. We showed analytically how the dynamics of bacterial populations are shaped by the activities of immunity-conferring enzymes in single cells and how the ecological consequences imply optimal bacterial defense strategies against viruses. Our results suggest that bacterial populations in the presence of viruses can either optimize their initial growth rate or their population size, with the first strategy favoring simple immunity featuring a single restriction modification system and the second strategy favoring complex bacterial innate immunity featuring several simultaneously active restriction modification systems.

7.10. Estimating information in time-varying signals

Participant: Jakob Ruess.

Across diverse biological systems - ranging from neural networks to intracellular signaling and genetic regulatory networks - the information about changes in the environment is frequently encoded in the full temporal dynamics of the network nodes. A pressing data-analysis challenge has thus been to efficiently estimate the amount of information that these dynamics convey from experimental data. In [1], we developed and evaluated decoding-based estimation methods to lower bound the mutual information about a finite set of inputs, encoded in single-cell high-dimensional time series data. For biological reaction networks governed by the chemical Master equation, we derived model-based information approximations and analytical upper bounds, against which we benchmarked our proposed model-free decoding estimators. In contrast to the frequently-used k-nearest-neighbor estimator, decoding-based estimators robustly extract a large fraction of the available information from high-dimensional trajectories with a realistic number of data samples. We applied these estimators to previously published data on Erk and Ca^{2+} signaling in mammalian cells and to yeast stress-response, and found that substantial amount of information about environmental state can be encoded by non-trivial response statistics even in stationary signals. We argued that these single-cell, decoding-based information estimates, rather than the commonly-used tests for significant differences between selected population response statistics, provide a proper and unbiased measure for the performance of biological signaling networks.

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 as compressed sensing, source separation, and super-resolution imaging, for example.

Based on the results of [31], 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 two minimization algorithm, CoBic (Constrained Biconvex) and PeBic (Penalized Biconvex).

We apply the algorithms to the problem of Single-Molecule Localization Microscopy and compare the results with the well-known IHT algorithm [22]. Both visually and numerically the biconvex reformulations perform better. Furthermore, the algorithm has been compared to the IRL1-CEL0 [23] and Deep-STORM [25]. The IRL1-CEL0 minimizes an exact relaxation [29] 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 presented at the ISBI 2019 conference [6], as well as a more mathematical article was presented as a poster at GRETSI 2019 [12]. A full journal article has been submitted to the Biomedical Optics Express for a feature issue: Superresolution Microscopy on the 25th Anniversary of STED Microscopy and the 20th Anniversary of SIM.

6.2. Biological Image Super-resolution Enhanced with Tensor

Participants: Jose Henrique de Morais Goulart, Laure Blanc-Féraud, Eric Debreuve, Sébastien Schaub.

This work is part of the BISET project, funded by the académie 1 RISE (Réseaux, Information et Société numErique) of Idex UCA JEDI.

Fluorescence microscopy imaging has numerous applications in biological sciences, but has limited resolution due to light diffraction. Recently proposed super-resolution techniques acquire an image time series at a high frame rate and exploit independent random fluorophore blinking for reconstruction. This approach holds great potential for observing live-cell sub-cellular phenomena, which is a challenging scenario with strict constraints over the deployed excitation levels and the acquisition time.

The BISET project aimed to develop tensor-based super-resolution fluorescence microscopy algorithms based on this approach. Assuming a known separable PSF $h(x, y) = g(x)g(y)$, a third-order tensor model with two spatial diversities and one temporal diversity was proposed. The model unknowns are high-dimensional fluorophore spatial profiles along x and y directions and temporal fluorophore profiles modeling blinking. Our formulation employs a least-squares loss term and penalty functions promoting spatial profile sparsity (necessary for fluorophore locality) and temporal profile group sparsity (which controls the number of fluorophores).

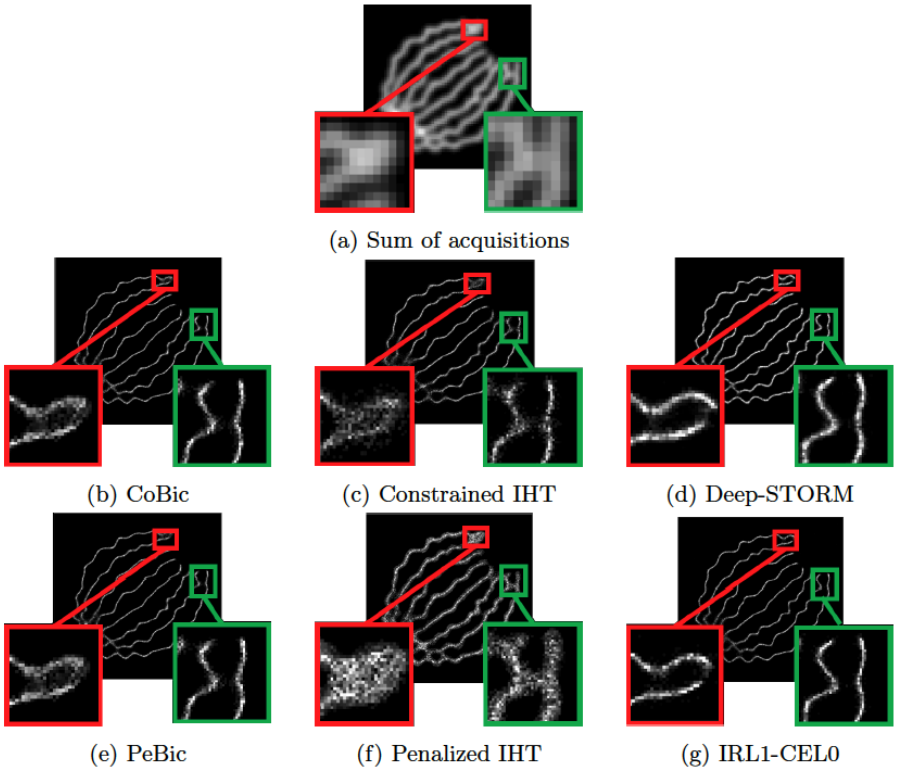


Figure 1. Reconstructed images from the simulated ISBI dataset [28], 99 non-zero pixels on average. Top: Sum of the acquisitions. Middle: From left to right: CoBic, Constrained IHT and Deep-STROM. Bottom: From left to right: PeBic, Penalized IHT and IRL1-CEL0.

The formulation is nonconvex but block-convex in the unknown profiles and thus can be solved by alternating minimization. It has a significantly smaller number of unknowns in comparison with a matrix-based convex one (with frames as columns), in consonance with the current trend of employing nonconvex formulations rather than overparameterized convex ones which are often too costly. However, its resolution is numerically challenging for high-density acquisitions. Indeed, even though the proposed algorithm is able to reveal the overall target structure in our simulations, it produces a "dotted" reconstruction. For comparison, we developed a matrix-based formulation with nonconvex group-sparsity regularization, which is more costly to solve but achieves better results. These findings were published in the IEEE CAMSAP 2019 conference [11], and were also presented on October 2019 in a GdR ISIS (Information, Signal, Image et Vision)/MIA (Mathématiques de l'Imagerie et de ses Applications)/ONDES meeting [0] held in Paris and entitled "Co-conception: hybrid sensors and algorithms for innovative systems". An illustration of the results produced by the developed tensor and matrix methods is given in Figure 2 , along with outcomes of other state-of-art methods. In conclusion, though our tensor approach is innovative and was shown to be promising, further research is needed to overcome the model estimation difficulties.

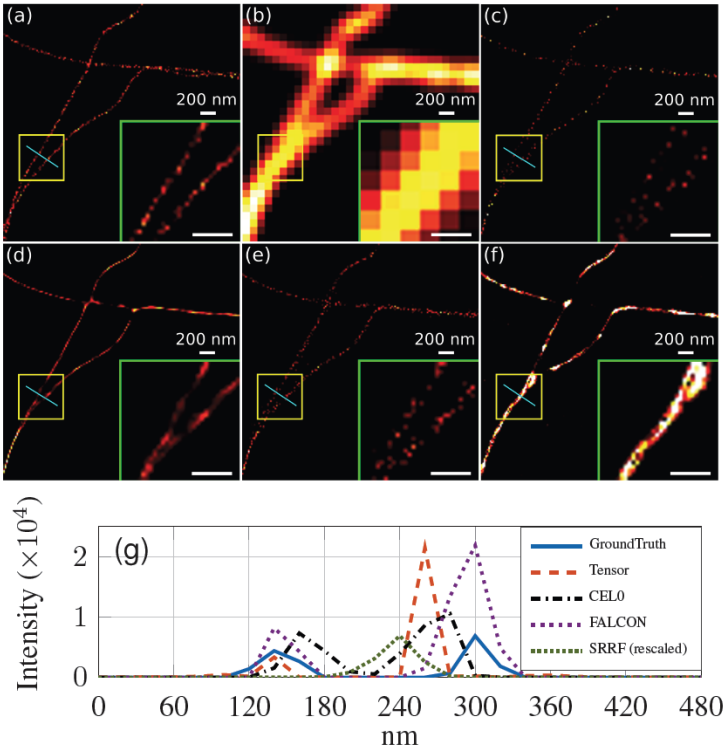


Figure 2. Results for reconstruction of simulated microtubules: (a) integrated ground truth; (b) integrated observed stack (5× zoom); (c) proposed tensor approach; (d) proposed matrix approach; (e) FALCON; (f) SRRF; (g) intensity profiles along the shown blue line. The frame in the bottom right corner shows a 2.66× *zoom of the smaller yellow frame.*

6.3. Classification and Modeling of the Fibronectin Networks in Extracellular Matrices

[0]Meeting webpage: `http://www.gdr-isis.fr/index.php?page=reunion&idreunion=401`.

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

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) networks, a major extracellular matrix (ECM) molecule, expressed in pathological states (fibrosis, cancer, etc). Our goal is to develop numerical quantitative biomarkers that describe the geometrical organization of the different four variants of the FN fiber networks, from 2D confocal microscopy images. Since the functions of these variants are not well defined in the context of their role within the tumour microenvironment, we hope that a computational model might be able to provide a meaningful description that incorporates the structural differences among the variants.

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 [24] and show that the curvelet coefficients are capable of discerning among the four FN variants with similar performances to those of a human annotator.

The second step of our work consisted 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). The graph parameters corresponding to the geometrical and topological features of the improved skeletonizations (i.e. median pore circularity, ratio of fiber thinness, fiber thickness kurtosis, fiber connectivity) of the four FN variants, are then classified by a DAG-SVM. It is thus shown through the analysis of the feature distribution over the four variants, features PCA analysis and SVM-based classification, that graph features can discriminate among the FN variants almost as well as our first work. This proves that the graph representation embeds the most relevant information provided by the image.

The next step focused on the development of a metric between graphs that takes into account their topology and geometry. This distance is bound to provide a quantitative but also a qualitative comparison of the four FN variants as well as a differentiation between normal and tumour-like FN fibers. In order to evaluate the distance among graphs, we have referred to graph-matching techniques, which are considered standard problems that deal with graph comparison. The main idea is to obtain an evaluation of the similarity between two graphs, by finding the optimal correspondence between their nodes, such as to align their structure, i.e their adjacency matrices. We expect to obtain invariance with respect to translation, rotation and scale.

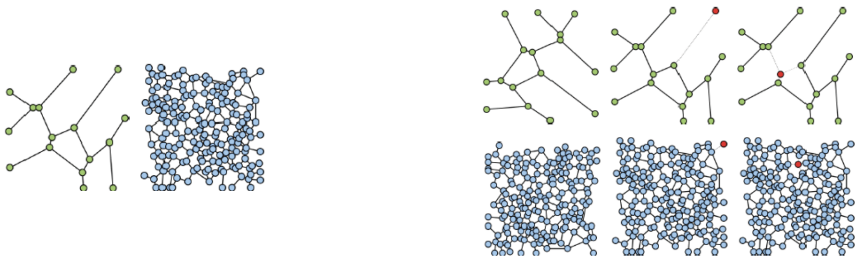


Figure 3. Generated toy-graphs with different dimensions: $16, 181$ *nodes (left side). Right side illustrates the database of toy-graphs derived from the initial ones, but having small modifications in terms of node order (first column) and different number of nodes (second and third column). The purpose is to match the nodes of every pair of graph (initial-modified) using the graph-matching and optimal assignment framework and compare the performances of the two methods.*

More specifically, we are interested in one of the various techniques to perform many-to-many graph matching [32], where the merging of multiple nodes to match another one is allowed, especially in the case of graphs with different dimensions (i.e. different number of total vertices). Alternatively, we considered a different line of work, based on optimal transport for the comparison of structured objects (e.g. graphs) with associated probability distributions. We focus on the work of Peyré et al. [26] that have considered a metric called Gromov-Wasserstein, capable of comparing objects that lie in spaces with different dimensions, by minimizing the cost of mass transport from one discrete distribution to the other. In the context of graph matching techniques, this can be regarded as a probabilistic assignment problem.

In [13], we have compared the two aforementioned approaches from a graph-matching perspective, on randomly generated graphs (Figure 3), in the context of a preliminary study for the future modeling of FN graph-based representations. We have tested different graph scenarios, with various information captured by the adjacency matrix (binary adjacency matrix, shortest path between nodes). Moreover, we have slightly modified the second method by optimal transport, to make it feasible for direct one-to-one matching, by adding dummy masses. We have concluded that the graph matching by many-to-many assignment, captures a meaningful distance between two given graphs with good performances, while the Gromov-Wasserstein discrepancy is computed faster but with lower performances.

One advantage of using graph-matching techniques for comparing fiber networks, comes from the possibility of defining a median graph that will be representative of a FN class. Currently, we are developing methodologies for deriving the representative graph for FN variants, using the metric provided by the many-to-many graph assignment problem. The challenges range from deciding a good technique to perform a meaningful matching among the graphs, to determining the adjacency of the median graph and the corresponding physical localization of the nodes.

A second advantage is given by the possibility of computing various deformation maps between FN fiber networks: the matching serves as a registration between the graphs, and once after having obtained an assignment between the corresponding graph edges, we can compute the differences in terms of fiber length, orientation, etc (see Figure 4 for a example of a deformation map in terms of fiber length - which can be regarded as a local stretching of the fibers that should be applied to first graph in order to obtain the second one)).

The deformation maps can subsequently be analyzed in a test hypothesis framework that decides whether the variation of a certain parameter (e.g. length) is due to the the variance within the same class or not.

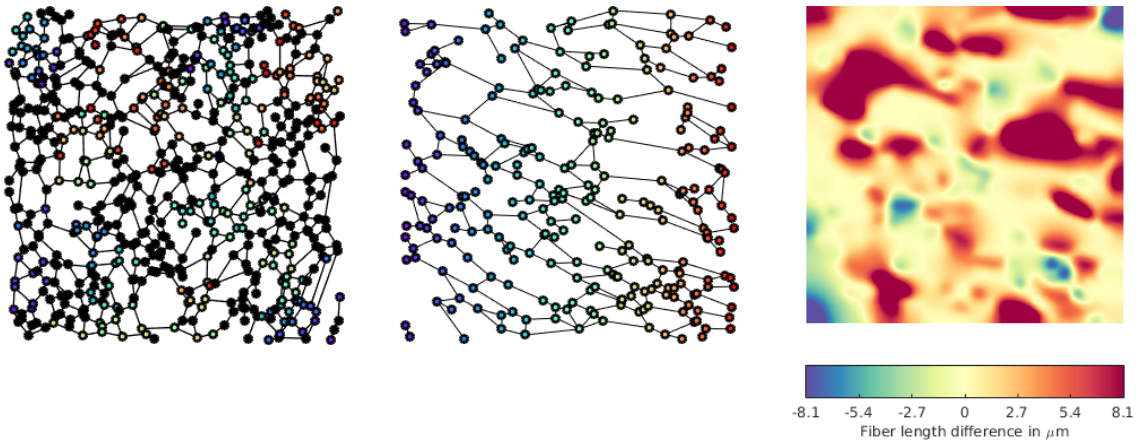


Figure 4. From left to right: graph network FN A+; graph network FN A+ ("tumour-like"); Deformation map between FN A+ and FN A+ (tumour-like)

Once we have derived a meaningful median graph based on graph-matching distances, we might be able to perform classification of the graph networks. Additionally, the fiber properties statistics inferred from the graph local properties, as well as Gabor filters parameters, can be of use to interpret the local differences within a specific class and among FN variants.

Anca Grapa's work is supported by the French Government (National Research Agency, ANR) through the "Investments for the Future" LABEX SIGNALIFE: program reference ANR-11-LABX-0028-01.

6.4. Classification of the Fibronectin Networks in Extracellular Matrices using CNN and DAG-SVM of confocal and coverslip scanner images

Participants: Ghosh Avrajit, Anca-Ioana Grapa, Laure Blanc-Féraud, Xavier Descombes.

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) networks, a major extracellular matrix (ECM) molecule, expressed in pathological states (fibrosis, cancer, etc).

Firstly, during one experiment, confocal images 3128×3128 pixels with a lateral resolution of 0.27μm/pixel were acquired with a Zeiss LSM710 confocal system $10X/0.45$ with the pinhole diameter set to its maximal value. Subsequently, images of FN variants in a different experiment were acquired using a coverslip scanner (Vectra Polaris Automated Quantitative Pathology Imaging System) based on fluorescence whole-slide scanning on a similar resolution to that of the confocal system.

For each of the experiments, 70 images (for every FN variant) corresponding to a representative region of 512×512 pixels were selected for feature extraction and classification.The set of 280 gray-scale images was classified with a DAG-SVM classifier using curvelet features using the parametrization from [24]. Additionally, it was classified with the GoogLeNet [30] pretrained Convolutional Neural Net (CNN) architecture using the MATLAB Deep Learning Toolbox and a 22-layer deep network trained on more than 1 million images for classification into 1000 object categories. A set of 196 images was used for the training of the algorithm, and the remaining 84 for testing it. The training image set was presented to the algorithm 25 times (epochs), in order to improve classification accuracy.

The results (Figures 5 , 7 , 6 , and 8) show that the information in the FN images is relevant enough in a CNN-based classification to distinguish FN variants better than curvelet-based features. Additionally, the coverslip scanner acquired samples are classified with a higher accuracy, underlining the potential benefit of using the scanner for future experiments.

Actual / Predicted	FN B-A+	FN B-A-	FN B+A-	FN B+A+
FN B-A+	85.7	0	28.5	14.3
FN B-A-	0	80.9	14.3	4.8
FN B+A-	0	9.5	90.5	0
FN B+A+	9.5	14.3	0	76.2

Figure 5. Confusion matrix in percentage form of the CNN classification of FN variant confocal images. General mean accuracy of classification is 83.3%.

6.5. Tumor cell tracking for automatic detection of cell death time, and classification of its type

Participants: Deborah Cottais, Eric Debreuve.

Actual/ Predicted	FN B-A+	FN B-A-	FB B+A-	FN B+A+
FN B-A+	64.3	2.9	25.7	7.1
FN B-A-	0	90	0	10
FN B+A-	25.7	4.3	45.7	24.3
FN B+A+	0	15.7	8.6	75.7

Figure 6. Confusion matrix in percentage form of the DAG-SVM classification of FN variants, using curvelets features. General mean accuracy of classification is 68.9%.

Actual/ Predicted	FN B-A+	FN B-A-	FB B+A-	FN B+A+
FN B-A+	95.2	0	0	4.7
FN B-A-	0	100	0	0
FN B+A-	0	4.7	62	33
FN B+A+	0	0	0	100

Figure 7. Confusion matrix in percentage form of the CNN classification of FN variant coverslip scanner images. General mean accuracy of classification is 89.3%.

Actual/ Predicted	FN B-A+	FN B-A-	FB B+A-	FN B+A+
FN B-A+	72.8	0	0	27.1
FN B-A-	0	85.7	8.5	5.7
FN B+A-	0	10	65.7	24.2
FN B+A+	15.7	0	11.43	72.8

Figure 8. Confusion matrix in percentage form of the DAG-SVM classification of FN variant coverslip scanner images (curvelet features).General mean accuracy of classification is 74.2%.

This work was made in collaboration with Jérémie Roux (IRCAN, Nice, France).

The available data were multi-channel videos acquired in fluorescence microscopy. We first performed cell segmentation on the channel in which the geometrical information was predominant. Then we performed tracking of the segmented cells. More precisely, we refer to tracking as the construction of cell trajectories along the video (see Fig. 9). By *transferring* this cell tracking onto the channel in which the radiometric information of the cells is the richest (mean intensity, variance, texture), we were able to extract characteristics for each cell, and study their temporal evolution to deduce the moment of cell death. Next, we are planning to develop a method of classification of the cell deaths into predefined types.

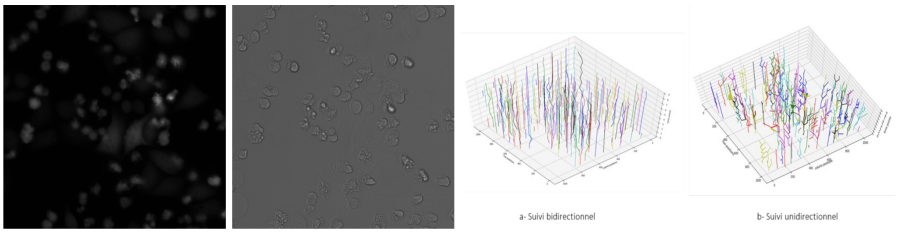


Figure 9. From left to right: two channels of a video frame and tracking trajectories of the segmented cells.

6.6. Cytoplasm segmentation from cells confocal microscopy images

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

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, ...).

The next step is to segment (i.e., extract automatically the region of) the cell cytoplasms on the GFP images. Indeed, the target RNP-IMP granules appear in that compartment of the cell and are visible through their GFP response. 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 cluster 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. An automatic algorithm based on the watershed transform has been chosen after evaluating and comparing different strategies (based on iterative clustering, minimum spanning tree, persistent edge selection ...).

The second step of the proposed approach performs superpixels merging to obtain the final segmentation. Starting from the previously detected nuclei to define cell seeds, the neighboring superpixels are merged iteratively if a radiometric similarity is detected. Ambiguities between neighboring cells are solved by combining radiometric and shape criteria. This cell growth process is considered layer by layer and performed in parallel.

6.7. Adaptative thresholding using persistent diagrams

Participants: Paul Emmanuel Ponsenard, Xavier Descombes.

In this project we have proposed a new algorithm for adaptative thresholding based on persistence diagrams. A common difficulty in binarizing biological images lies in the heterogeneity of the signal. This heterogeneity can be due to the sensor itself but also to variability in the cell response to a given marker. Therefore the binarization can not be adequately performed by using the same threshold on the whole image. In this context, adaptative approaches that estimate a local or regional proper threshold are needed. Last year, we have proposed a solution that embedded both a contrast term and a shape criterion to select the most relevant connected components among the different level sets of the image. In this work we focus on the connected component trajectories along the grey values defining the levels sets. More precisely, the persistent diagram studies the evolution of the different connected components of a binarized image for successive thresholds. The life time of a connected component is thus defined as the timelapse between its birth (gray level for which the component appears) and its death (gray level for which the component is merged with a neighboring one). As a final result for the binarization, we propose to keep the connected components with the longest lifetimes. A result on mitochondrial network binarization is shown on figure 10 .

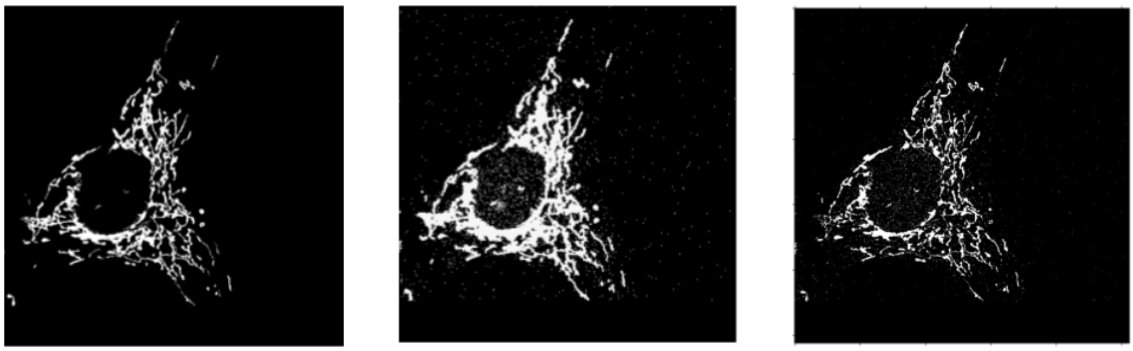


Figure 10. Image of Mitochondria from Bost team at C3M (left), Binarization obtained with a global threshold (middle) and with the persistence diagram approach. (right).

6.8. Graph matching and median graph through simulating annealing

Participants: Zhankeng Zhang, Xavier Descombes.

Graph matching when the number of nodes and edges differs is known as an NP-hard problem. Therefore, sub-optimal optimization algorithms have been proposed to solve this problem. In this work, we evaluate the possibility to reach, at least theoretically, the global optimum by using simulated annealing. We have developed an improved version of the simulating annealing scheme based on a Metropolis sampler. To solve the problem of dimension matching (different number of nodes) we have classically added dummy nodes in the smaller graph. Besides, we have shown that adding dummy nodes in both graphs provides more flexibility in the matching, thus improving the matching result. Finally, within this framework we were able to define and compute "median" graph as shown on figure 11 . The algorithm consists in aligning all the graphs in a first step. The median graph is then obtained by considering two types of move in the simulated annealing: adding/removing an edge and switching two nodes. To validate this work we have considered a classification scheme between graphs. The obtained results overcome those obtained with state of the art graph matching algorithms while the computational time remains reasonable.

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

Participants: Sarah Laroui, Eric Debreuve, Xavier Descombes.

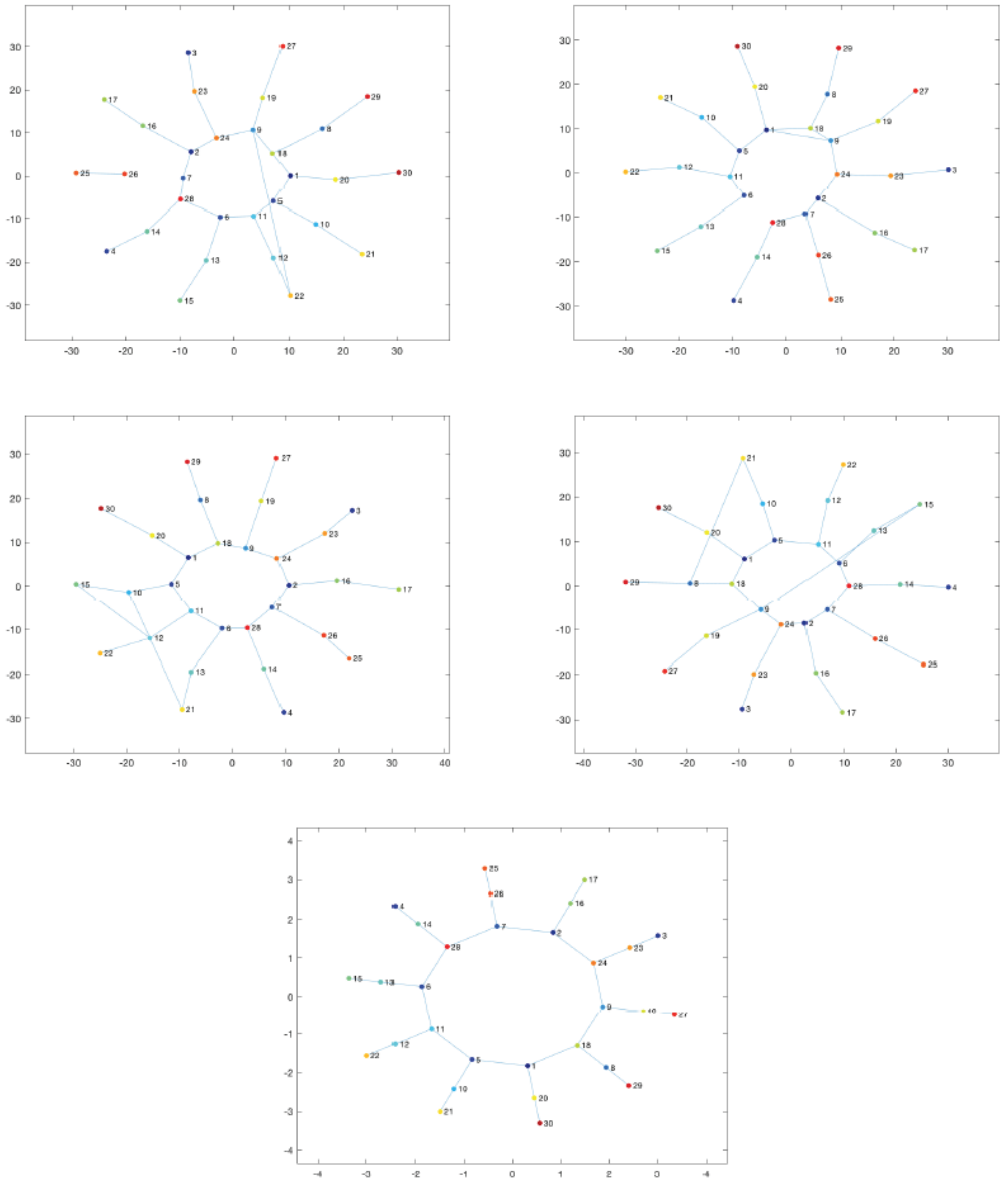


Figure 11. Four samples of noisy SUN graph and computed median graph

This work is made in collaboration with Aurelia Vernay and Florent Villiers (Bayer).

Botrytis cinerea is a reference model of filamentous phytopathogen fungi. Some chemical treatments can lead to characteristic morphological changes, or phenotypic signatures. These phenotypes could be associated with the treatment Mode of Action (Figure 12). In order to recognise and characterise different phenotypes and associate them with the different modes of action of the molecules (Figure 13), 24-hour images are taken by transmitted light microscopy. Because of the different dose-response effects, each given molecule is tested at ten concentrations.

We compared the results of classification of these images using two methods: random forests and convolutional neural networks (Deep Learning).

To learn the Random Forest classifier, we developed a robust image analysis and classification framework relying on morphometric and topological characteristics. A number of 16 features are extracted from three representations of the objects (binary mask, skeleton and graph). Some are calculated globally over all the objects of an image (ex: the skeleton length variance) while others are calculated on each object of an image (ex: the number of nodes of the graph). The second method uses a convolutional neural network. It has been implemented using Tensorflow, an open source library for Machine Learning, created by Google to develop applications in Deep Learning.

This method achieves better results than Random Forests, and it proved to be very robust to inter-experiment variations with an average classification accuracy of 88%. In addition, it does not require data pre-processing for feature extraction. However the explanatory aspect that exists with random forests is lost.

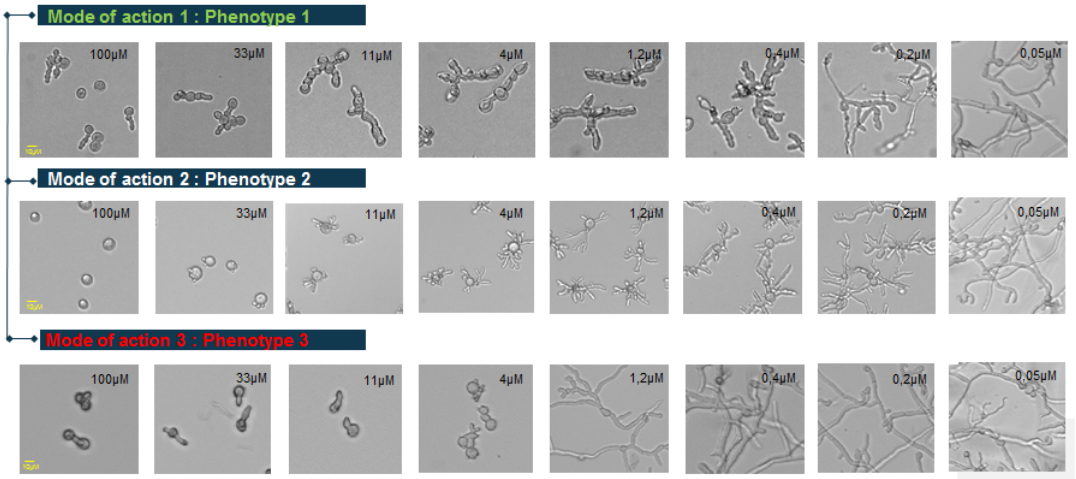


Figure 12. Characteristic phenotypic signatures for different chemical treatments at different concentrations (transmitted light microscopy, ImageXpress microscope, 10x lens).

6.10. Estimating the volume of a copepod from a single image with Deep Learning

Participants: Cédric Dubois, Eric Debreuve.

This work was made in collaboration with Jean-Olivier Irisson (Laboratoire d'Océanographie de Villefranche).

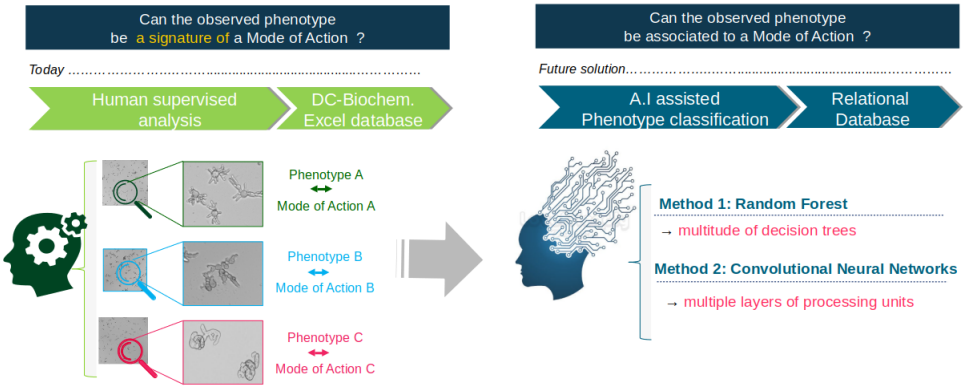


Figure 13. Example of phenotypic signatures obtained with molecules with three different mode of action. Strategy of automatic recognition and characterisation of different phenotypes and associate them with the different modes of action of the molecules.

Ecologists and biogeochemists are interested in estimating the volume of copepods (to then convert it into a biomass), a subclass of zooplankton, in order to estimate how much carbon it can store and how much it will store in the future. Those studies are made thanks to the online database EcoTaxa, which gives access to a large number of plankton images. The standard method used in ecology produces partially incorrect results due to geometric approximations and projection issues (from 3D to the 2D image plane). We first proposed a study of the error made by this method on the volume estimation of copepods. Then we proposed a new method based on the deep learning framework. Its performances have been analyzed on simulated data (Fig. 14) and preliminary tests have been made on a subset of the data of the *UVP5hd GreenEdge 2016* acquisition campaign available on EcoTaxa. Our work pointed out the limitations of both methods, indicating that a broader study is needed to improve the computation of copepod volumes.

This work formed the basis for Cédric Dubois's PhD which began on October 1st 2019 with a Ministère de la Recherche funding.

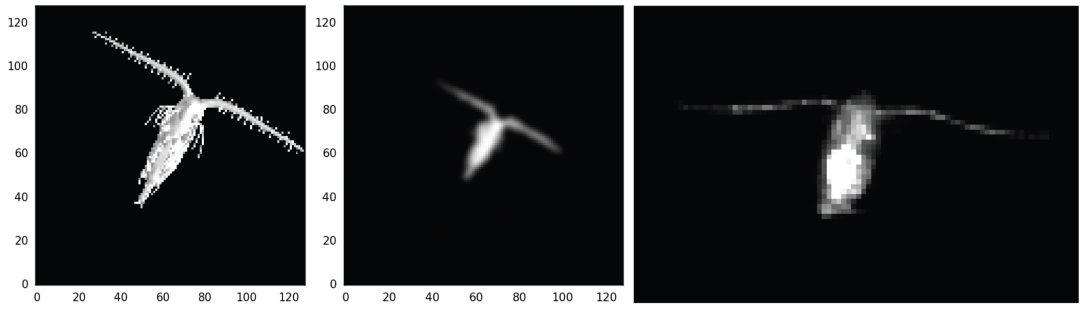


Figure 14. Left: synthetic 3D model of a typical copepod. Middle: a simulated 2D observation of the model. Right: a real observation.

6.11. Cell lineage calculation

Participants: Manuel Petit [Mosaic, Lyon], Christophe Godin [Mosaic, Lyon], Grégoire Malandain.

This work is made within the IPL Naviscope.

In recent years, techniques to image the development of biological organisms have made spectacular progresses. Researchers are now able to observe the trajectories corresponding to the development of 3D- plant tissues or animal embryo with cellular resolution. However, such observations yield a large amount of data, which, in turn, require fast and robust analysis tools to extract information while minimizing user interaction. The goal of M. Petit, which PhD thesis has begun november the 1st, is first to propose new lineage extraction schemes, and then analysis tools over a population of lineages.

6.12. Morphogenesis of the sea urchin embryo

Participants: Angie Moullet, Grégoire Malandain.

This work is made in collaboration with Barthélmy Delorme and Matteo Rauzi (iBV, Nice).

The goal of the project is to understand how biophysical forces are generated and how they work to produce exquisitely precise and controlled tissue shape changes in embryo development. Tissue morphogenesis is a process by which the embryo is reshaped into the final form of a developed animal. Tissues are constituted by cells that are interconnected one another: local changes of cell mechanical properties and shape drive consequent tissue shape change. Nevertheless, the knowledge per se of the mechanisms and mechanics at the cell level which drive cell shape changes is insufficient to explain how tissues change their shape. Emerging properties arise at higher scales resulting from the interaction of cells within tissues and of tissues coordinating and interacting with one another.

To study the embryo evolution at a cellular scale, temporal series will be acquired by a multi-view light-sheet microscope. We will use the Mediterranean sea urchin embryo species Paracentrotus lividus as a model system and focus on the process of tissue folding, that will process that is vital since folding defects can impair neurulation in vertebrates and gastrulation in all animals which are organized into the three germ layers. From the technological perspective, new tools are needed to be able to visualize cells and to provide quantifiable data at high temporal and spatial resolution over large regions and across the entire embryo.

The goal of A. Moullet's internship (that begins dec. the 1st) is to measure and study the archenteron length evolution over a population of sea urchin embryos.

6.13. 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 [27]. This approach is only applicable when the operator has avoided vessel superimposition over the vessel of interest. To further extend the concept, we explore the benefit of doing the deformable registration over the whole coronary tree. This benefit is illustrated in Fig. 15 and through tracking videos presented in https://3dvttracking.github.io/.

The proposed approach involves several algorithmic steps: a rigid registration of the tree to an iso-cardiac phase projection followed by a deformation of the tree represented as a tree-spline.

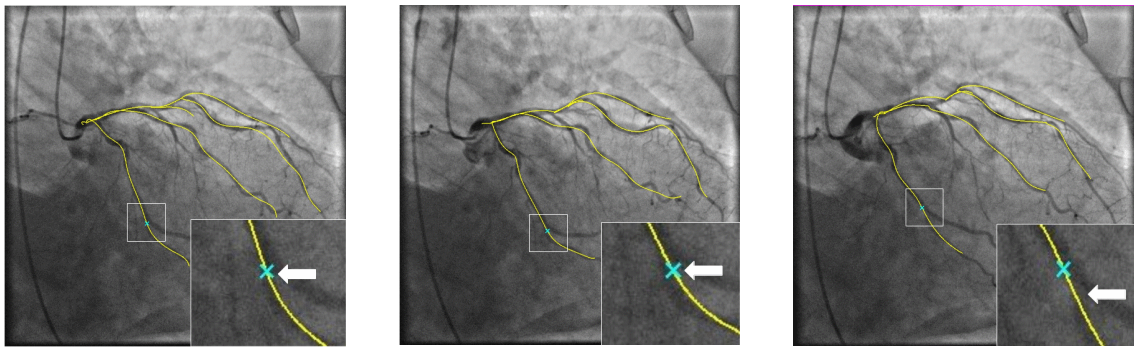


Figure 15. Tracking results for one patient over one cardiac cycle. The yellow curve represents the projected 3D vessel, the blue cross represents the point tracked as the bifurcation, and the white arrow points to the bifurcation. Those images come from a 15 frames sequence. This figure shows the frames 1, 6, 15, from left to right.

Indeed, a tree-spline i.e. a tree with a spline attached to each edge and shared control points between these points describes a 3D coronary tree and is able to represent its deformation along the time. We combine this description with a registration algorithm operating between the tree-spline and the angiographic projection of the coronary tree. It starts by the estimation of a rigid transformation for the iso cardiac phase time followed by a non-rigid deformation of the tree driven by the pairings formed between the projection of the edges of the tree-spline and the observed x-ray projection of the coronary arteries. The pairings are built taking into account the tree topology consistency. Anatomical constraints of length preservation is enforced when deforming the arteries.

This work has been published in FIMH [9].

MOSAIC Project-Team

6. New Results

6.1. Dynamical characterization of morphogenesis at cellular scale

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

- Related Research Axes: RA1 (Representation of biological organisms and their forms in silico) & RA3 (Plasticity & robustness of forms)
- Related 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:

Cell-scale atlases of development. One fundamental question linked to morphogenesis is at which level and timescale tissue or organ development is reproducible and stereotyped. To answer this question, variability must be quantitatively assessed. In the team we have created to this end 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.

Thanks to the invariant cellular lineage of early development of *P. mammillata* embryos and to 3D reconstruction of their development at cellular resolution, quantitative comparison of their properties from cell to tissue scale has been performed. After fluorescent membrane labelling, several embryos have been imaged for several hours by light-sheet microscopy. These images were then reconstructed through the segmentation pipeline ASTEC, which also automatically tracked each cell over several rounds of cell division. This large amount of data allowed us to create an atlas of geometrical and topological properties at cellular resolution, which gives unprecedented depth of information on the variability of ascidian development. In addition this atlas, coupled to previous knowledge on gene-expression dynamics from the ascidian genetic database (ANISEED), made it possible for us to develop a mathematical and computational model to explore the main drivers of early ascidian development, identified as area-of-contact-mediated cell-cell communications. This model was also validated by experimental manipulations and mutations induced in ascidian embryos. This work is currently under review [26].

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 in early phases, the development of most plant organs is under the influence of robust genetic patterns 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 Arabidopsis 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.

Robustness of ascidian embryonic development. The image segmentation pipeline ASTEC developed by the team in collaboration with the Inria Morpheme project-team in Sophia Antipolis and the CRBM team in Montpellier, allows the 3D reconstruction and tracking of each cell during early ascidian embryogenesis. This methods allowed us to reconstruct over 50 ascidian embryos, both wild-type and mutants. Exploiting this large database and the fixed cellular lineage of ascidian embryos, we extracted and compared geometrical and topological cellular properties. This allowed us to compare the intra-embryonic (left/right) to the inter-embryonic level of variability of several properties, including cell volume, cell-cell contacts and the structure of the tree seeded by each cell. This study demonstrated that the genetic-induced variability is comparable to the stochastic one, quantitatively showing that ascidian embryonic development is highly canalized, and that the high reproducibility of shapes observed during embryogenesis is rooted in the robustness of cellular geometry and topology. To look for the origin of this canalisation, we developed a mathematical model exploiting our quantitative geometric database and the previously-existing ascidian genetic database ANISEED. This model suggests that the main driver of ascidian development is the cell-cell communication mediated by direct physical contact, and hence dependent of the area-of-contact between neighbouring cells. This means that the robustness of cell topology and geometry is necessary for cell-cell biochemical interactions to give rise to the correct fate restriction events, which in turn we showed to be responsible for major changes of embryo geometry. We also tested and validated this feedback loop between cell contacts, fate restriction events and embryonic geometry predicted by the model by manipulations and mutations induced in ascidian embryos.These results are reported in a paper which is currently under review [26].

Robust extraction and characterization of cellular lineages. The quantification of temporal properties at cellular scale such as volumetric growth rate or strain patterns relies extensively on the identification of cellular lineages in time-lapse acquisitions of living tissues. In the case of plant tissues where the deformations between two consecutive time points can be very important in post-embryonic morphogenesis processes such as early flower development, it remains a real challenge to compute those lineages automatically, and manual user annotation is generally required to produce reliable results.

Building on the previous expertise of the team [25], [28] and on the state-of-the-art computational library for image analysis, timagetk, developed in collaboration with the Morpheme team, we currently develop a set of robust automatic cell lineaging methods for cases ranging from small to highly non-linear deformations. In the course of a M2 internship and the first months of a starting PhD work (Manuel Petit), a first so-called “naive” lineaging method has been implemented and validated on synthetic data with limited deformations. Methods involving optimal flow algorithms on graph structures and iterative image registration are being developed to provide robust results in the case of faster growing tissues. The output of these methods will allow to use the tools developed by the team for the analysis of spatio-temporal properties of growing cells at a much larger scale. This work is part of the Inria IPL Naviscope.

Reconstruction of Arabidopsis ovule development. 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. Given the technical difficulty of producing live-imaging acquisition sequences of ovules, we developed a method to perform a spatial registration of multiple individual ovules at various developmental states and in different global poses. Using the global cylindrical symmetry of the organ and the surface curvature as a key geometrical feature, we aligned individuals on their main axes and on their junction with the underlying placental tissue. Jointly with the 3D segmentation of cells in images, this will allow to evidence the invariant features of ovule development at cellular scale, and to study the robustness of the dynamics of the megaspore mother cell (MMC) across indivduals. This work was part of the Imago project.

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

Participants: Ayan Chaudhury, Christophe Godin, Jonathan Legrand, Katia Mirande.

- Related Research Axes: RA1 (Representations of forms in silico) & RA3 (Plasticity & robustness of forms)
- Related 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. In collaboration with the ROMI partners from Sony CSL, Paris, 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. We aim at releasing both hardware schematics and the developed software for image reconstruction to be used as cheap open-source solution to phenotype plants. In addition, 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:

- RGB images were generated from the model and used as input of the pipeline;
- a physical version of the model has been obtained using 3D printing techniques;

In both cases, knowing the generated phenotypic traits or the model shape allow to test the pipeline ability to reconstruct the plant and quantify its traits of interest

The developed reconstruction and quantification pipeline is not made from scratch but aggregate a number of available third party libraries and codes in addition to three active research topics: spectral clustering, skeleton extraction, and ML segmentation. 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.

6.3. Analysis of tree data

Participants: Romain Azaïs, Christophe Godin, Salah Eddine Habibeche [External Collaborator], Florian Ingels.

- 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 2019, 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 into 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 [11], 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. In [3], we present two approximation algorithms that are optimal but assume that the approximation can be obtained by only adding vertices to the initial data (or by only deleting vertices from the initial data). In [11], 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. Another algorithm based on the efficient simulation of conditioned random walks on the space of trees is currently under development. This work should result in the submission of a paper next year.

It should be noted that the aforementioned strategy and algorithms can only be applied to topological trees. In 2019, we also began a new project on approximation of trees with geometrical attributes on their vertices and with possibly a controlled loss of information during the compression.

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 [5], 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 [5]. We have introduced new estimators and stated their consistency. Our techniques improve the existing results both theoretically and numerically.

We continue to explore these questions for subcritical but surviving Galton-Watson trees. The conditioning is a source of bias that must be taken into account to build efficient estimators of the birth distribution. This work should be submitted to a journal next year.

Kernel methods for tree data. Standard statistical techniques – such as SVMs for supervised learning – are usually designed to process Euclidean data. However, trees are typically non-Euclidean, thus preventing using these methods. Kernel methods allow this problem to be overcome by mapping trees in Hilbert spaces. However, the choice of kernel determines the feature space obtained, and thus greatly influences the performance of the different statistical algorithms. Our work is therefore focused on the question of how to build a good kernel.

We first looked in [17] at a kernel of the literature, the subtree kernel, and showed that the choice of the weight function – arbitrarily fixed so far – was crucial for prediction problems. By proposing a new framework to calculate this kernel, based on the DAG compression of trees, we were able to propose a new weight, learned from the data. In particular, on 8 data sets, we have empirically shown that this new weight improves prediction error in 7 cases, and with a relative improvement of more than 50% in 4 of these cases. This work was presented at a national conference [15].

We then tried to generalize our framework by proposing a kernel that is no longer based on subtrees, but on more general structures. To this end, we have developed an algorithm for the exhaustive enumeration of such structures, namely the forest of subtrees with a uniform fringe. This work will be submitted for pre-publication early in the coming year.

6.4. Mechanics of tissue morphogenesis

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

- Related Research Works: RW2 (*Data-driven models*) & RW3 (*Plasticity & robustness of forms*)
- Related 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. Integrating cellular behaviors such as mechano-sensitivity, intercellular fluxes of materials and cell division into a macroscopic mechanical picture of morphogenesis is the topic of this section.

Flattening mechanism during organogenesis in plants. Many plant species have thin leaf blades and axisymmetric elongating organs, such as stems and roots. From a morphoelastic perspective, such complex shapes are currently believed to emerge from the coordination between strain-based growth and stress-based stiffening at the cellular level.

To study the plausibility of such an hypothesis, we conducted numerical simulations where both a stress-based stiffening mechanism of cell walls [29] and a strain-based growth mechanism [24] have been implemented. We performed such simulations on multicellular and multilayered ellipsoidal structures and track their aspect ratio as they developed under various parametrization sets. One key aspect we wanted to investigate was the effect of an heterogeneous stress-based stiffening mechanism on the overall dynamics: Starting from a given initial shape, can we get significantly different shapes by assuming the stress-based stiffening mechanism active only in specific parts of the structures?

Our results, in accordance with experimental measurements conducted simultaneously by biologist colleagues, showed that: (i) Stress-based stiffening was mandatory to grow flat and axisymmetric organs; (ii) in order to grow flat structures, stress-based stiffening should only be active on anticlinal inner walls.

This work was part of Jan Traas's ERC grant *Morphodynamics*. This work is currently under review, see preprint version [23].

Influence of cell division during flat organogenesis in plants. One key limitation of our 3D modeling approach of leaf-like organogenesis is the lack of cell division implementation. This can be seen as a major flaw in the mechanical understanding of flattening since cell divisions, by increasing the number of load bearing walls, impact significantly the redistribution of mechanical stresses within the tissue.

To alleviate this limitation, we developed a 2D modeling approach to complement the 3D one. This 2D model encompasses the same biophysical processes as the 3D one (described in the previous subsection): a stress-based stiffening and a strain-based growth mechanisms of cell walls; augmented with a cell division module. We used this 2D framework to investigate the flattening dynamics of structures mimicking ellipsoid cross sections of growing organs. Such cross section were described as vertex-based, multicellular and multilayered structures.

We first reproduced the results obtained with the 3D approach to ensure that both models agreed on similar situations, where no cell division was implemented. We tested then several rules of cell division orientation and check which one(s) produced the most efficient flattening process. We were able to show that heterogeneity in the division rule between the epidermis and the inner tissues led to the more efficient flattening process and that a stress-based division rule was the most efficient to produce flat structure.

This analysis is part of the manuscript currently under review and available online in a preprint version [23].

Influence of mechanical stress anisotropy on the orientation of cell divisions in animal tissues. Tight regulation of cell division orientation is fundamental for tissue development. Recently, a great effort has been put into biophysical understanding of the *long-axis* division rules (Hertwig's rule for animal cells, Errera's rule for plant cells) and the systematic deviations from these rules observed *in vivo*. In both plants and animals, such deviations often correlate with anisotropic tensions within the tissue. To what extend these deviations are regulated or simply the result of stochasticity?

To address these questions in animal cells, we modeled theoretically and numerically cell division as an active process in a many-body system. We showed that under isotropic tension a cell's long axis emerges as the energetically optimal division orientation and that anisotropic stresses biased the energetics, leading to systematic deviations form Hertwig's rule. These deviations, as reported experimentally, are correlated to the main direction of stress anisotropy.

Our model successfully predicted division orientation distributions within two experimental systems: epidermis of the ascidian *Phallusia mammillata* (where deviations from Hertwig's rule have been so far eluding explanation) and of the pupal epithelium of the dorsal thorax of *D. melanogaster*.

This work was part of the *Digem* project and was presented in two international conferences: *Mechanobiology and Physics of Life* (Lyon) and *Developmental and Cell Biology of the Future* (Paris); and at the yearly *InriaBio* meeting in Lyon. A paper is currently under review and a preprint is available on bioRxiv [22].

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, between experiments and modeling, address the influence of turgor pressure heterogeneities on relative growth rate between cells. We showed that the coupling between fluxes and mechanics allows to predict observed morphological heterogeneities without any *ad hoc* assumption.

This work was part of the Agropolis fundation project *MecaFruit3D* and Arezki Boudaoud's ERC *PhyMorph*. It resulted in a publication in PLoS Computational Biology [7] that introduces the theoretical model and studies some of its properties. Another paper [27] presents the comparisons with experiments and is currently under review.

Development of *de novo* finite element (F.E.) library dedicated to mechanical simulations performed on complex cellularized structures. In order to compute accurately the mechanical stress field borne by multicellular pressurized 3D structures (such as plant tissues), we needed to update our existing library (*tissueMeca*, see [24]). Three key aspects had to be upgraded (i) the control over the F.E. solver, (ii) tracking of its precision and (iii) integration of the F.E. framework with the rest of our pipeline.

To that end, we decided to switch from *Sofa* to *FEniCS* (https://fenicsproject.org/) as the core F.E. framework used within our simulation pipeline. We started to develop a dedicated library, called *CellFem*, to solve F.E. problems on *PropertyTopomesh* instances (the data structure we developed within the team to describe multicellular plant tissues). *CellFem* provides a high level API to define and resolve variational problems to solve linear as well as non-linear elastic and elasto-plastic problems related to plant tissue morphogenesis.

In parallel, we also started the development of a meshing library (based on the GMSH library (http://gmsh.info/)) called *CellMesh* and dedicated to the triangulation of simplicial complexes. This work is currently under development.

6.5. Signaling and transport for tissue patterning

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

- Related Research Axes: RA1 (Representations of forms in silico) & RA2 (Data-driven models)
- Related Key Modeling 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 the spatial precision of the targeting. 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 govern the rhythmic patterning of organs, and the consequent emergence of phyllotaxis.

Using *Arabidopsis thaliana* as a model system, we develop 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 an original method to quantify in 3D the polarization of auxin transport for every anticlinal wall of the first layer of cells in confocal images. The developed method [13] was thoroughly evaluated against super-resolution acquisitions of the same tissue obtained using radial fluctuations (SRRF), and show to provide highly consistent results (less than 10% incorrect polarities, 80% of cells with a polarity vector error lesser than 30°). The digitally reconstructed networks evidenced an overall stable convergence of PIN1 polarities towards the center of the meristem, with a local convergence and divergence pattern that could explain the dynamics of auxin distributions in the meristem [19].

Landmark-based registration for the averaging of meristem patterning. To perform statistics of meristem patterning at the scale of a population, we developed a series of tools to compute a rigid 3D transformation that registers any individual meristem into a common cylindrical reference frame in which point-wise comparison is meaningful. The orignal method relies on the identification of biological landmarks (apex and main symmetry axis of the meristematic dome, position of the lastly emerged organ primordium and direction of the phyllotactic spiral) to compute this transform. These landmarks can be extracted from image acquisitions of meristems carrying the right fluorescent bio-markers (*CLV3* central zone marker for the apex, *DIIV* auxin bio-sensor for the organ primordia) using an original method that relies on the computation of 2D continuous maps of epidermal signal from discrete point clouds. The use of this registration method allowed to evidence key features of the transcriptional response of mersitematic cells to auxin [19].

In a second time, we aim to generalize the method to images without specific bio-markers, using only the geometry of the tissue to identify the relevant landmarks. To do so, machine learning approaches making use of the data processed for [19] are being developed and evaluated. This new landmark-based registration method would drastically improve the ability of comparing different individual meristems, open the way to spatial statistics over of multiple genetic and molecular signals, and contribute to an integrated tissue-level view of meristem patterning.

Computational models of integrated transport and signaling. Guided by new discoveries on auxin patterning dynamics in the shoot apical meristem (SAM) of *A. thaliana*, we developed a theoretical model of active and passive auxin transport. This model, built on existing view of auxin active transport [30], [31], naturally integrates the role of deeper cellular layers in the SAM and the mutual feedbacks between different components of the auxin-transport machinery. Through numerical simulation, the consequences of competing theories on PIN polarisation mechanism on auxin dynamics were explored. These results will serve, in quantitative comparisons with *in vivo* observation, to validate hypotheses on molecular mechanisms of auxin transport and to provide information on the role of memory effects and information fluxes during patterning.

These works were part of the *BioSensors* HFSP project and are carried out in the *Phyllo* ENS-Lyon project. These works gave rise to a journal article which is currently under review and have been partly presented at the *International Worskhop on Image Analysis Methods for the Plant Sciences* in Bron in July 2019.

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 activation/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 [6] and [10] for a 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.

6.7. Miscellaneous

Participants: Romain Azaïs, Christophe Godin, Bruno Leggio.

Measurements and nonlocal correlations in quantum mechanics. Based on a long standing collaboration between Christophe Godin and Przemyslaw Prusinkiewicz from the University of Calgary on the analysis of connections between computer simulation paradigms and quantum mechanics, we theoretically investigated with the quantum mechanics expertize of Bruno Leggio in the team effects of measurements on quantum systems, mostly in connection with quantum non-locality and entanglement. At the same time, we exploit formal and conceptual analogies between quantum theory and biologically-inspired structures to study the latter under new paradigms.

One fruitful line of research deals with the inherent non-locality of correlations between measurement outcomes, characterizing the quantum world. These phenomena are described by the celebrated Bell inequalities. We study ways to generalize such inequalities to better capture non-local correlations, at the same time shedding light on the origin of the discrepancy between quantum and classical stochasticity. In parallel, we develop and profit from formal analogies between the theory of non-locality and the exploration of fractal structures in the context of simulation of arborescent systems.

Another research line sees the application of parameter-estimation techniques for piecewise deterministic Markovian processes (PDMP), developed by members of the team, to the special case of quantum dynamics: under certain conditions, the evolution of an open quantum system can be described as a PDMP, with a specific and non-trivial structure marking its departure from classical behaviour. We show [21] that approaches to appraise parameter values of the evolving systems, developed in the context of classical dynamics, can be successfully applied to the specific case of quantum systems.

Finally, a third research topic consists of the study of the structure of typical quantum correlations, called entanglement, and its relation to thermal noise induced in a quantum system by its unavoidable interaction with its surrounding environment. We show [9] that the quantitative amount of noise represents a tight upper bound on the amount of bipartite quantum correlation two systems can establish between them.

Statistical analysis and stochastic modelling of penguin diving. The activity at sea of penguins can be reconstructed from measurement devices equipped on the animals during their trips. We study the relative behavior of the time under water with respect to the time spent at the surface from a dataset of about 100 thousands dives of little penguins. We show that dives that form a bout in which the penguin explores a patch of preys show a type of stationarity. We have built a mathematical model of sequences of dives that can be optimized in terms of number of preys catched by the animal under physiological constraints. This reproduces the stationary behavior observed in the data.

PLEIADE Project-Team

7. New Results

7.1. Genetic Determinisms of Aborted Fermentation in Winemaking

This study contributes to the understanding of the mechanisms leading to stuck fermentation in winemaking, an economic prejudice [7]. A number of factors can trigger stuck or aborted fermentation such as high temperature, high ethanol concentration, low pH. The biodiversity of natural yeast strains used in winemaking starters has as a consequence that some of them are more prone to abort fermentation than others, indicating a genetic determinism. Crosses between strains called "sensitive" or "resistant" to stuck fermentation occurrence, followed by back-crosses with the "sensitive" parent while selecting for the "resistant" phenotype, allowed us to reduce the amount of genetic material inherited from the "resistant" parent in the progeny, ending to 3 small introgression areas after 4 generations. Quantitative Trait Locus (QTL) detection in this progeny (77 strains) involved characterization of SNP inheritence (circa 1200 validated SNPs) from either parent, through micro-array hybridization, mapping of the SNP on the reference genome and phenotypic measurements on the progeny. This analysis made it possible to detect two genes which, when inactivated by naturally occurring mutations, act as major perturbators of several fermentation parameters in winemaking physiological conditions. Consequently, our industrial partner incorporated into its catalogue of winemaking starters, strains carrying the functional forms of these genes.

7.2. Characterization of Molecular Biodiversity

In 2019 PLEIADE developed new methods for characterizing molecular biodiversity (see [4], [5] for applications). This point itself has been developed with two approaches in 2019, each with the beginning of a PhD.

- Building OTUs from a pairwise distance matrix typically is an unsupervised clustering issue. In this new development, the PhD student (PLEIADE) tests whether SBM (Stochastic Block Model) approach yields relevant results for a global characterization of biodiversity beyond a summary by a scalar index. This is done in collaboration with MIAT INRAE research unit in Toulouse and EPC HiePACS. It represents a connection between metabarcoding and statistical modeling, a topic which deserves investigation and is expanding. (Figure 5 from [8])
- A major goal of PLEIADE is to develop a geometric view on biodiversity. The tool selected up to now is to associate a point cloud to a dataset (pairwise distances between sequences) and study its shape. In 2019, PLEIADE has been associated in a collaboration with HiePACS to begin a new topic: comparison between point clouds, each cloud being associated to a data set. Indeed, the development of metabarcoding leads to the new issue of comparison between OTUs built from different dataset. This approach is part of the issues raised in a PhD supervized by HiePACS, in collaboration with PLEIADE.

7.3. Scaling Metabarcoding Programs

Metabarcoding is a series of technical procedures to build molecular based inventories from large datasets of amplicons. We derived new methods and tools to scale metabarcoding programs in collaboration with EPC HiePACS. This has been realized through following participation in research projects:

- Contribution to ADT Gordon project in Inria BSO. The objective of this project (partners: Tadaam (coordinator), STORM, HiePACS, PLEIADE) is to integrate SVD as an available tool in Chameleon, starPU and new Madeleine. The contribution of Pleiade is to bring metabarcoding as a use case, and random projection ([6]) as a method for scaling Multidimensional Scaling (which requires an SVD) in collaboration with HiePACS with a template implemented in Diodon. In 2019, PLEIADE has bought to the project a series of 55 matrices with size about 10^5 rows/column which, assembled, yield a full pairwise distance matrix between one million of sequences. The objective is to reach the million in 2020.

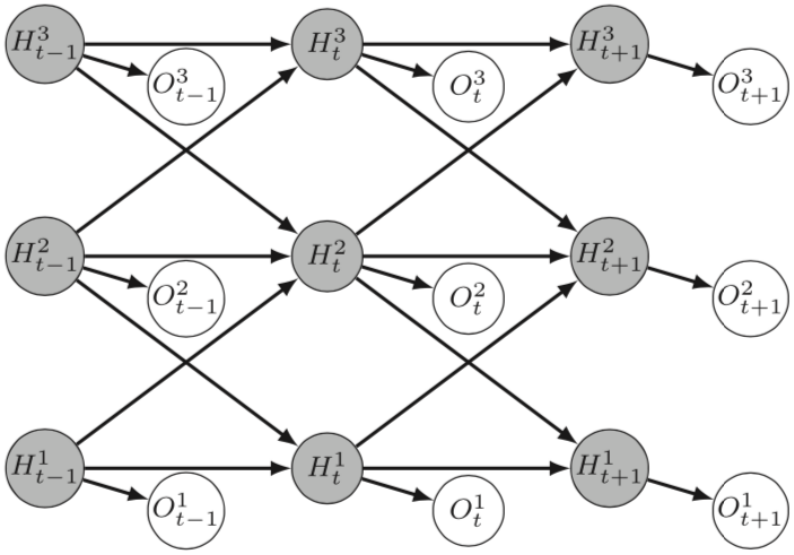


Figure 5. Graphical representation of a coupled HMM with three hidden chains (from [8])

- Contribution to Region Nouvelle Aquitaine project "HPC Scalable Ecosystem." This project is chaired by HiePACS. In collaboration with this EPC, PLEIADE is involved in developing a new approach for comparing OTUs built from different datasets.

7.4. Linking Homology and Function for Algal Desaturases

Polyunsaturated fatty acids (PUFA) such as Omega-3 that are essential for human health cannot be synthesized by the human body and must be acquired through the consumption of certains foods, such as oily fish, that are becoming increasing difficult to produce sustainably. However, fish do not produce them either: their role is to concentrate PUFA through the food chain. There is consequently considerable interest in producing these essential nutriments directly, through the cultivation of domesticated strains of naturally occurring or of engineered strains of green algae.

Ultimately, polyunsaturated fatty acids are produced by molecular machines called **desaturases**. While desaturases are abundant in all branches of life, the link between gene sequence and the precise activity of the corresponding enzyme is poorly understood. The particular challenge is that, while the catalytic active site is well conserved, the features that recognize the substrate and that determine the regiospecificity of the enzyme are not. In order to produce specific PUFA at industrial scale, it is necessary to develop efficient tools for high-throughput identification of candidate genes in algal species, and precise models for designing desaturases through synthetic biology.

In collaboration with the LBM (UMR 5200 CNRS) and with the support of the Inria Project Lab *In silico algae*, we used a core collection of thirteen desaturases from *Osteococcus tauri* to explore the link between homology and function in 23 species ([9]). The study reinforced our understanding of the evolutionary conservation of desaturases and confirmed the identification of substrate and regio- specificity through graph neighborhoods. We were further able to extend the identification of PPR motifs correlated with specificity. This work is ongoing. Since most of the pertinent desaturases are membrane bound, the prediction of protein structure has proved perilous, but we are hopeful that future work will allow us to use structure-inspired prediction to narrow in on the sites responsible for specificity despite their poor sequence conservation.

SERPICO Project-Team

7. New Results

7.1. Empirical SURE-guided microscopy super-resolution image reconstruction from confocal multi-array detectors

Participants: Sylvain Prigent, Charles Kervrann.

Recent confocal microscopes use an array detector instead of single point detector to take multiple views of the same sample. The microscope output is then an array of images, one image per detector. The array of images is then processed to build a single image with higher signal-to-noise ratio and higher resolution than a classical confocal microscope image. In the literature, several methods have been recently proposed to reconstruct the single high resolution image: i/ the ISM method combines array registration and Wiener deconvolution; ii/ the IFED method estimates a high resolution image by subtracting the outer detectors of the array to the inner detectors; iii/ the ISFED consists in subtracting the outer registered detectors and the outer raw images. In that context, we proposed a SURE-guided (Stein's unbiased risk estimation) estimation method to automatically select the parameter ϵ controlling the IFED and ISFED reconstruction algorithms (see Figure 4). We showed on real data that the proposed method is capable to achieve a resolution close to 100 nm without any deconvolution method.

Software: AiryscanJ (see Section 6.4).
Collaborator: S. Dutertre (IGDR – Institute Genetics & Development of Rennes).

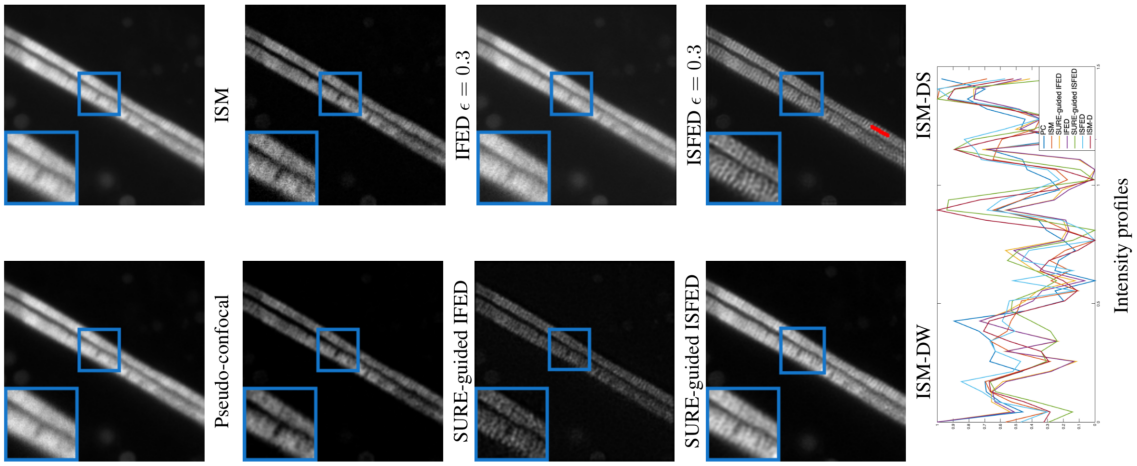


Figure 4. Results obtained on a c.elegans sample and intensity profiles for all the tested methods. The profile line is shown with a red line on the ISM-DS image (top right).

7.2. Dense mapping of intracellular diffusion and drift from single-particle tracking data analysis

Participants: Antoine Salomon, Cesar Augusto Valades Cruz, Charles Kervrann.

It is of primary interest for biologists to be able to locally estimate diffusion and drift inside a cell. In our framework, we assumed that particle motion is governed by the following Langevin equation: $\mathbf{dx} = \mathbf{b}(\mathbf{x})dt + \sigma(\mathbf{x})\mathbf{dw}$ where $\mathbf{x}$ denotes the 2D or 3D spatial coordinates of the particle, $\mathbf{b}$ the drift vector, σ the diffusion coefficient, and $\mathbf{w}$ the standard Gaussian white noise. In that context, we proposed a new mapping method inspired from [50] that developed a method providing results in the form of matrices by scanning the data by blocks, and from the framework in [5], dedicated to both classifying particle motion types and detecting potential motion switches along a trajectory. To avoid the calculation of both drift and diffusion in cell coordinates where no data is available, we replaced the scanning movement of an averaging window by a Gaussian window centered on trajectory points. Each drift vector and each diffusion coefficient are calculated at coordinates corresponding exactly to the coordinates given by the preliminary particle tracking, which provides more details. A nonparametric three-decision test enables to label trajectories or sub-trajectories [5]. This information is then used. to calculate drift and diffusion coefficient (or Kramers-Moyal coefficients) maps separately on each class of motion with the most appropriate diffusion models: confined motion (the particle is bound to a specific point), Brownian motion (the particle moves randomly), and directed motion (the particle moves in a given direction) (see Figure 5).

Software: THOTH and CPAnalysis (see Sections 6.2 and 6.6).

Collaborators: V. Briane (UNSW Sydney, School of Medical Sciences, Australia),
L. Leconte and J. Salamero (CNRS-UMR 144, Institut Curie, PSL Research University),
L. Johannes (U1143 INSERM / CNRS-UMR 3666, Institut Curie, PSL Research University),
E. Derivery (MRC laboratory of Molecular Biology, Cambridge, UK),
L. Muresan (Cambridge Advanced Imaging Centre, UK).

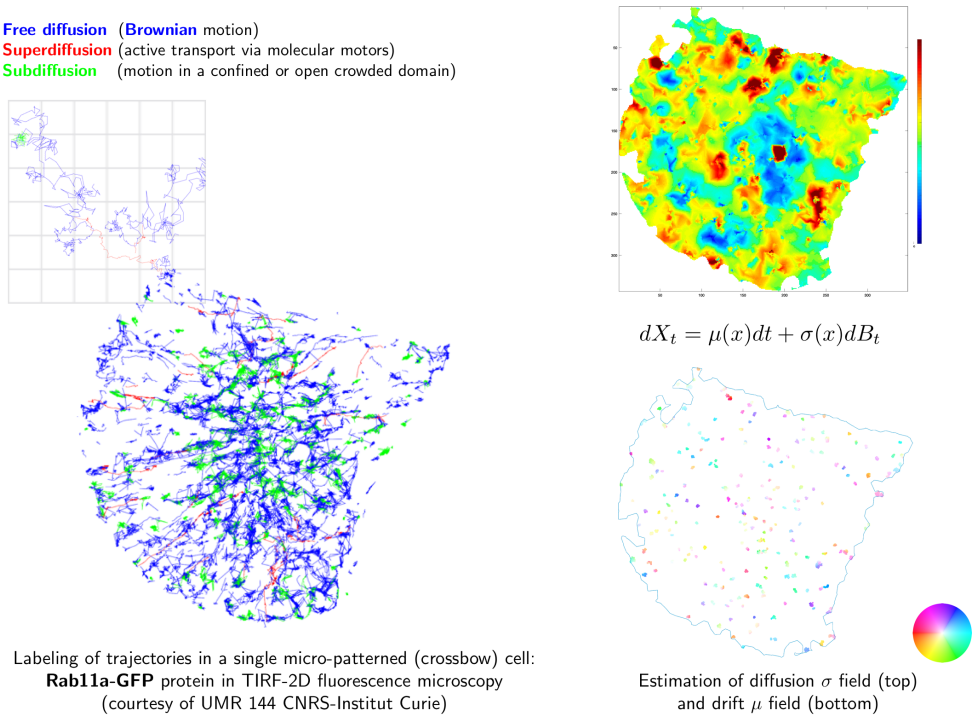


Figure 5. Intracellular diffusion and drift maps estimated from simulated tracking data (FluoSIM).

7.3. 3D tracking of endocytic and exocytic events using lattice light sheet microscopy

Participants: Cesar Augusto Valades Cruz, Ludovic Leconte, Jean Salamero, Charles Kervrann.

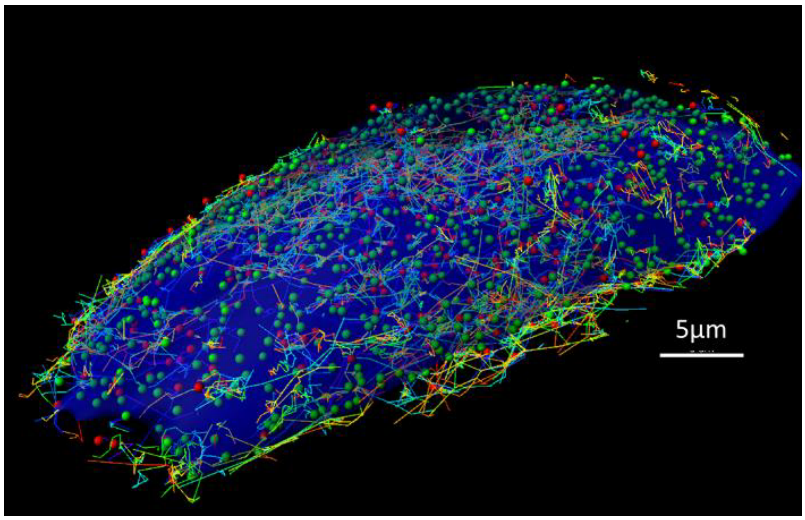


Figure 6. 3D tracking of Gal3-Atto647n (red) vs AP2-eGFP (green) adaptor protein in SUM159 cell.

The study of the whole cell dynamics of endocytic/exocytic-recycling events has proven difficult until recently because of lack of sensitivity, limited speed, photobleaching and phototoxicity associated with conventional imaging modalities. The Lattice Light Sheet Microscope (LLSM) allows to overcome these difficulties, yet reaching high spatial resolution. 3D images are captured for several minutes at a high acquisition frequency, and enables the study of signaling, transport, and stochastic self-assembly in complex environments. In addition, this imaging technique and 3D-tracking allow us to characterize the molecular machineries involved in the exocytosis and endocytosis mechanisms. We have the opportunity to observe a series of sequential events corresponding to the fusion with the plasma membrane (exocytosis) and the formation of endocytic carriers, including the trafficking of vesicles throughout the entire membrane system. We have got preliminary results of the coordination of vesicle recycling from the endosomal recycling compartment up to the plasma membrane using LLSM imaging and 3D tracking. In addition, we introduced a quantitative analysis of endocytosis dynamics of AP2 adaptor complex, Galectin-3 (see Figure 6) and Transferrin using single particle tracking analysis of 3D+time data. These case studies clearly demonstrated the advantage of lattice light sheet microscopy for imaging endocytic/exocytic events in single cells.

Software: THOTH and CPAnalysis (see Sections 6.2 and 6.6).

Collaborators: C. Wunder and L. Johannes (Institut Curie, PSL Research University, Cellular and Chemical Biology, U1143 INSERM / UMR 3666 CNRS).

7.4. 3D flow estimation in 3D fluorescence microscopy image sequences

Participants: Sandeep Manandhar, Patrick Bouthemy, Charles Kervrann.

We have proposed a variational approach for 3D optical flow computation from a pair of fluorescence microscopy volumes. This computational method has been extensively evaluated on appropriate simulated data. To simulate a volume pair with a ground truth flow field, we extended the Horn-and-Schunck optical flow method to 3D (3DHS). The computed flow field by 3DHS between two input images is then used to generate a new pair volume. The new source and target volumes along with the corresponding 3DHS flow fields serve as the dataset with the ground truth. The latter was used for the parameterization of our method and for the comparative study with the state-of-the art method proposed by Amat et al., 2012 [44].

Meanwhile, we proposed a novel error measure named SAE (for Structural Angular Error) for 3D optical flow, in absence of any ground truth flow field (see Figure 7). SAE measures the angular difference between the principal orientations of the structures present in the backward warped target and the source volumes at each voxel. We found out that the average of SAE (ASAE) and average of the end-point-error (a standard optical flow error in presence of ground truth) behave similarly.

We also integrated L_1 regularization in our variational approach. In contrast to our previous L_2 regularization approach, this method preserves discontinuities of the flow field. However, both of our methods are time demanding and not parallelizable in implementation. Then, we integrated our Census Signature-based data term with total variation regularization that also produces discontinuous flow fields. Consequently, we took the splitting approach for optimization. A gain of four times was obtained in the calculation speed with 12-core implementation of the new method, compared to our previous two methods, for still similar ASAE score.

We have also proposed two new methods for the visualization of the 3D flow fields, named 3DHSV and MIP-flow respectively. The 3DHSV method color codes the 2D projections of the flow field in the Hue and the Saturation color spaces, while mapping the off-the-plane motion to the Value space. MIP-flow also encodes the 2D projections to the Hue and the Saturation spaces. However, it only considers encoding the 2D projections of the 3D vectors corresponding to the maximum intensity points in the direction perpendicular to the projection plane. This work is carried out in collaboration with UTSW Dallas in the frame of the Inria associated team CytoDI.

Software: FlowScope (see Section 6.7).
Collaborators: P. Roudot, E. Welf and G. Danuser (UTSW Medical Center, Dallas, USA).

7.5. Probabilistic overall reconstruction of membrane-associated molecular dynamics from partial observations in rod-shaped bacteria

Participants: Yunjiao Lu, Charles Kervrann.

Understanding the mechanisms that maintain the structure of rod-shaped bacteria is a challenging problem in cell biological research. Thanks to progress in molecular biology and microscopy (e.g Total Internal Reflection Fluorescence (TIRF) microscopy), we have the opportunity to observe the dynamics of the cell wall construction workers, that is the membrane-associated molecular machines (MMs). Due to the cylindrical form of the bacteria and the 2D selective visualization in TIRF microscopy, only around one third of the perimeter can be observed at a given time. Nevertheless, from the partial observed bacteria surface images, earlier studies showed that a fraction of the MMs performs directed motion, across the image field quasi-orthogonally to the cylinder axis.

Accordingly, we addressed the problem of the connection of motion segments on a cylindrical surface, assuming that one MM may re-enters into the observed region (OR), a certain period of time after having left the field of view. The directed MM motions are assumed as Brownian motion with drift. The birth and death events of the MMs are supposed to happen independently and uniformly on the surface. Given a set of observed segments entering and exiting the OR, we proposed a probabilistic framework to calculate the probabilities of the events of birth, death and re-entry, based on speed and diffusion of the motion and the time of exit and entry. Even though two third of the surface is hidden as shown in Figure 8 , this framework allows us to derive a computational procedure aiming at connecting segments belonging to the same trajectory, and then recovering directed MMs dynamics on the whole surface. The performance of the method has been demonstrated on appropriate simulation data that mimics MMs dynamics observed in TIRF microscopy.

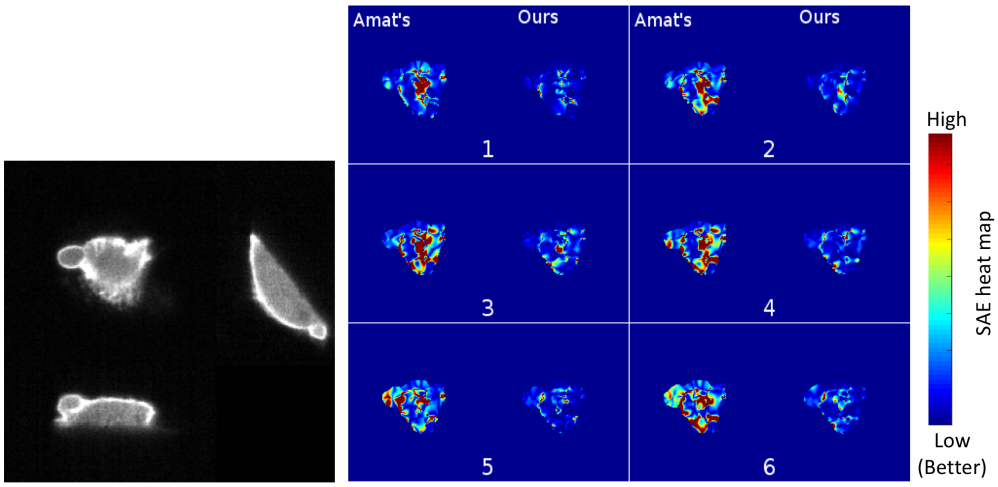


Figure 7. SAE maps (right) to compare our variational method to the Amat's method [44] applied to a 3D image pair (left) depicting a MV3 melanoma cell undergoing blebbing on a coverslip observed with Diagonally scanned Light-Sheet Microscopy (2.86 Hz sampling frequency, $300 \times 300 \times 83/50 \times 50 \times 30\mu m^3$*) (input images by courtesy of Danuser lab, UTSW Dallas, USA).*

Collaborators: A. Trubuil and P. Hodara (INRA MaIAGE unit, Jouy-en-Josas),
R. Carballido-López and C. Billaudeau (INRA, UR MICALIS, Jouy-en-Josas).

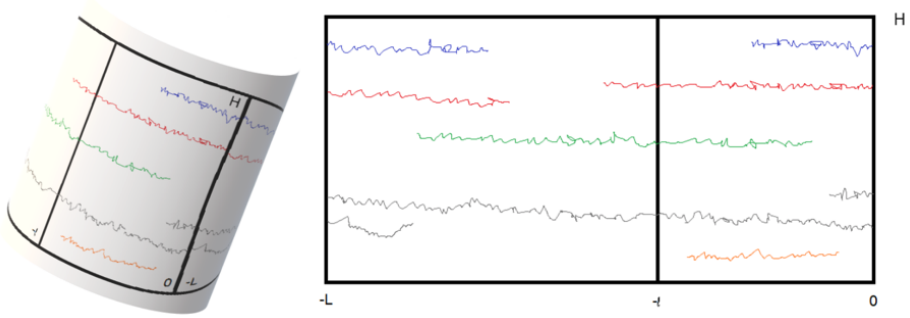


Figure 8. Trajectories on the cylinder and its 2D representation. The unobserved region is $(-L, -l] \times [0, H]$ *and the observed region is* $(l, 0] \times [0, H]$.

7.6. Data assimilation and modeling of cell division mechanism

Participants: Anca-Georgiana Caranfil, Charles Kervrann.

Asymmetric cell division is a complex process that is not yet fully understood. A very well-known example of such a division is the first division of C.elegans embryo. To improve our understanding of this process, we used mathematical modeling to study the first division of C.elegans embryo, both on wild type cells and under a wide range of genetic perturbations. Asymmetry is clearly visible at the end of the anaphase, as the mitotic spindle is off-center. The study of the mitotic spindle dynamics is, thus, a useful tool to gain insights into the general mechanics of the system used by the cell to correctly achieve asymmetric division. The overall spindle behavior is led by the spindle poles behavior. We proposed a new dynamic model for the posterior spindle pole that explains the oscillatory behavior during anaphase and confirms 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. By theoretically analyzing our model, 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 number 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 bio-physical 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. This model will be instrumental in probing the function of yet undocumented genes involved in controlling cell division dynamics.

Collaborators: Y. Le Cunff and J. Pécréaux (IGDR – Institute of Genetics & Development of Rennes).

7.7. Convolutional Neural Networks algorithms for calcium signal segmentation in astrocytes in 3D lattice light sheet microscopy

Participants: Anais Badoual, Charles Kervrann.

Astrocytes, glial cells of the central nervous system, are detectors and regulators of neuronal information processing. It is established that neuronal synapses are physical sites of intercellular contact that transmit and transform information in a very rapid and flexible way, playing a pivotal role for learning and memory formation as well as neurological diseases of the mammalian brain. Astrocytes are thought to integrate neuronal inputs and modulate information transfer between neurons. In particular, cytoplasmic calcium signaling in astrocytes is believed to be crucial for astrocyte-neuron communication. However, quantification of intracellular calcium signals in astrocytes is hindered by the complexity of their cell shape, that consists of a cell body sprouting a highly ramified set of large to very fine protrusions called processes. Until recently, the quantification of intracellular propagation of calcium signal in astrocytes with fluorescent calcium indicators has been restricted to two dimensions, either 2D cell cultures or 2D slicing of a 3D setup. However it is not clear what amount of information is lost by ignoring the 3rd dimension in these experiments. The emergent 3D Lattice Light Sheet Microscopy (LLSM) is a powerful and promising technology (voxel size: 250nm x 250nm x 700nm; acquisition time: 200 frames per second) to give a much more complete and refined view of the dynamic behavior of calcium signaling in astrocytes inside living brain slices and in the intact mouse brain in vivo. Unfortunately, we lack image analysis tools to locate, segment, track and quantify the propagation of those 3D calcium signals in very ramified cell shapes.

In this context, we have started to develop an image processing tool for neurobiologists that 1) detects and segments calcium signals in 3D+time LLSM images, and 2) classifies these signals based on their 3D space-time morphological characterization. To do so, we focus on 3D convolutional network and machine learning techniques.

Collaborators: V. Nägerl and M. Arizono (Interdisciplinary Institute for Neuroscience, Bordeaux), H. Berry and A. Denisot (EPC BEAGLE, Inria Rhone-Alpes).

7.8. Geo-colocalization and coorientation in fluorescence super-resolution microscopy

Participants: Frédéric Lavancier, Reda Alami Chantoufi, Aymeric Lechevranton, Antoine Salomon, Charles Kervrann.

Colocalization aims at characterizing spatial associations between two fluorescently-tagged biomolecules by quantifying the co-occurrence and correlation between the two channels acquired in fluorescence microscopy. This problem remains an open issue in diffraction-limited microscopy and raises new challenges with the emergence of super-resolution imaging. In [19], we proposed an original method (GcoPS) that exploits the random sets structure of the tagged molecules to provide an explicit testing procedure. GcoPS requires the adjustment of a p-value that guarantees more reproducibility and more objective interpretation and takes as inputs two 2D or 3D binary segmented images. This year, we extended this approach to the estimation of local co-localization. This amounts to applying the statistical test on windows randomly drawn in the whole image. A multiple testing procedure allows us to compute a global partial colocalization score. Meanwhile, the excursion sets of colocalization score map estimated by Gaussian smoothing are very helpful to detect regions of interest corresponding to significant colocalization and anti-colocalization sites. This approach has been evaluated on STORM (Stochastic Optical Reconstruction Microscopy) images which provides several hundreds thousands of super-localized positions of individual molecules with an average accuracy of 10-20 nanometers (see Figure 9). Finally, the method has been successively extended to the geo-coorientation (or geo-coalignment) of 2D-3D vectors (optical flow, tensors) and trajectories to analyze the molecular interactions.

Software: GcoPS (see Section 6.1).

Collaborators: J. Salamero (CNRS-UMR 144, Institut Curie, PSL Research University),
G. Bertolin (IGDR – Institute of Genetics & Development of Rennes),
M. Lelek and C. Zimmer (Institut Pasteur, Paris).

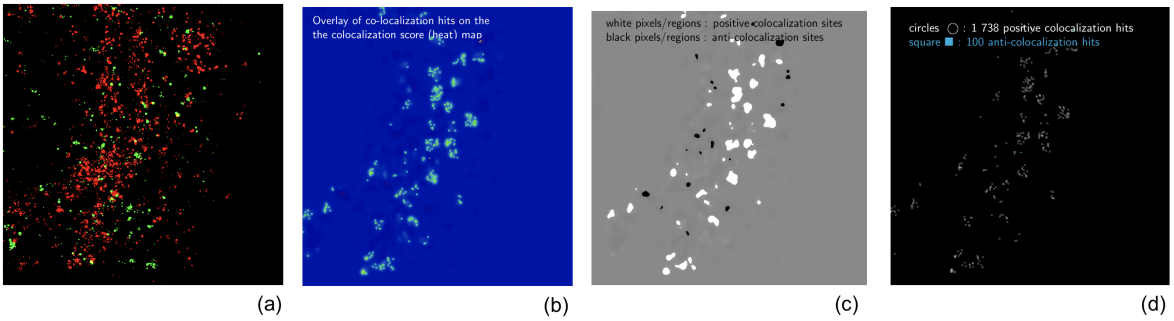


Figure 9. Illustration of geo-colocalization of two molecules (red/green channels) in STORM super-resolution microscopy (original image size: 4576 × 3564 pixels; pixel size: 3 nanometers). (a) Overlay of two channels (sub-region of the original pair); (b) colocalization hits overlaid on the score (heat) map; (c) excursion sets of detected colocalization (white) and anti-colocalization (black) sites overlaid on the score map; (d) detected co-localization (circles) and anti-co-localization (squares) hits.

7.9. Immersive and interactive visualization of 3D temporal data using a space time hypercube

Participants: Gwendal Fouché, Charles Kervrann.

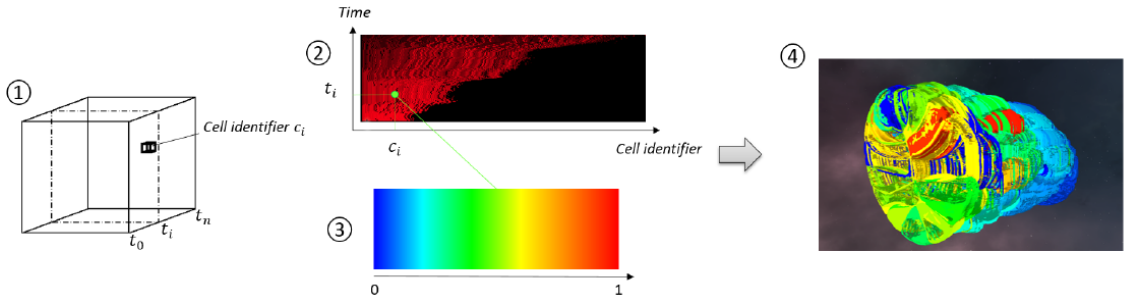


Figure 10. Flow diagram of the STC generation. In step 1, the user places the interactive clipping plan to get the desired cross-section. In step 2, camera parameters are automatically set in order to render the cross section at each time point. In step 3, the top image presents the output of the rendering operation, using the RGB channel to save cell identifiers; the bottom image is the result of an edge detection filter that will be useful for display. In step 4, the rendered images are stacked into a 3D texture. Each voxel contains a cell identifier, and its position in terms of depth indicates a time point t_i.

The analysis of multidimensional time-varying datasets, whose size grows as recording and simulating techniques advance, faces challenges on the representation and visualization of dense data, as well as on the study of temporal variations. In this context, we proposed an extension of the well-known Space-Time Cube (STC) visualization technique in order to visualize time-varying 3D spatial data acquired in 3D fluorescence microscopy, taking advantage of the interaction capabilities of Virtual Reality. The extended STC is based on a user-driven projection of the spatial and temporal information modeled as a 4D Space-Time Hypercube (STH). This projection yields a 3D STC visualization, which can also encode non-spatial quantitative data. Moreover, we proposed a set of tools allowing the user to manipulate the 3D STC that benefits from the visualization, exploration and interaction possibilities offered by immersive environments (see Figure 10). Finally, the extended STC has been integrated in a VR application for visualization of spatiotemporal biological data, illustrating the usage of the proposed visualization method for the morphogenesis analysis.

Collaborators: F. Argelaguet (EPC HYBRID, Inria Rennes),
E. Faure (Laboratory of Computer Science, Robotics and Microelectronics of Montpellier).

7.10. Unsupervised motion saliency map estimation based on optical flow inpainting

Participants: Léo Maczyta, Patrick Bouthemy.

We have addressed the problem of motion saliency in videos. Salient moving regions are regions that exhibit motion departing from their spatial context in the image, that is, different from the surrounding motion. In contrast to video saliency approaches, we estimate dynamic saliency based on motion information only. We propose a new unsupervised paradigm to compute motion saliency maps. The key ingredient is the flow inpainting stage. We have to compare the flow field in a given area, likely to be a salient moving element, with the flow field that would have been induced in the very same area with the surrounding motion. The former can be computed by any optical flow method. The latter is not directly available, since it is not observed. Yet, it can be predicted by a flow inpainting method. This is precisely the originality of our motion saliency approach.

Our method is then two-fold. First, we extract candidate salient regions from the optical flow boundaries. Secondly, we estimate the inpainted flow using an extension of a diffusion-based method for image inpainting, and we compare the inpainted flow to the original optical flow in these regions. We interpret the possible discrepancy (or residual flow) between the two flows as an indicator of motion saliency. In addition, we combine a backward and forward processing of the video sequence. The method is flexible and general enough, by relying on motion information only. Experimental results on the DAVIS 2016 benchmark demonstrate that the method compares favorably with state-of-the-art video saliency methods. Additionally, by estimating the residual flow, we provide additional information regarding motion saliency that could be further exploited (see Figure 11).

Collaborators: O. Le Meur (Percept team, IRISA, Rennes).

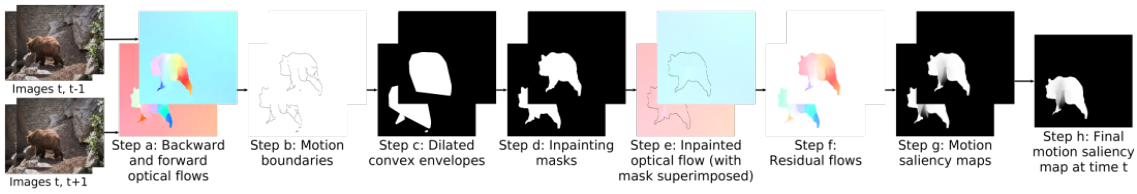


Figure 11. Workflow of the motion saliency estimation method.

ARAMIS Project-Team

7. New Results

7.1. Predicting PET-derived Demyelination from Multimodal MRI using Sketcher-Refiner Adversarial Training for Multiple Sclerosis

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

Multiple sclerosis (MS) is the most common demyelinating disease. In MS, demyelination occurs in the white matter of the brain and in the spinal cord. It is thus essential to measure the tissue myelin content to understand the physiopathology of MS, track progression and assess treatment efficacy. Positron emission tomography (PET) with [11C]PIB is a reliable method to measure myelin content in vivo. However, the availability of PET in clinical centers is limited. Moreover, it is expensive to acquire and invasive due to the injection of a radioactive tracer. By contrast, MR imaging is non-invasive, less expensive and widely available, but conventional MRI sequences cannot provide a direct and reliable measure of myelin. In this work, we therefore propose, to the best of our knowledge for the first time, a method to predict the PET-derived myelin content map from multimodal MRI. To that purpose, we introduce a new approach called Sketcher-Refiner generative adversarial networks (GANs) with specifically designed adversarial loss functions. The first network (Sketcher) generates global anatomical and physiological information. The second network (Refiner) refines and generates the tissue myelin content. A visual attention saliency map is also proposed to interpret the attention of neural networks. Our approach is shown to outperform the state-of-the-art methods in terms of image quality and myelin content prediction. Particularly, our prediction results show similar results to the PET-derived gold standard at both global and voxel-wise levels indicating the potential for clinical management of patients with MS.

More details in [25].

7.2. Reproducible evaluation of methods for predicting progression to Alzheimer's disease from clinical and neuroimaging data

Participants: Jorge Samper-González, Ninon Burgos, Simona Bottani, Marie-Odile Habert, Stéphane Epelbaum, Theodoros Evgeniou, Olivier Colliot [Correspondant].

Various machine learning methods have been proposed for predicting progression of patients with mild cognitive impairment (MCI) to Alzheimer's disease (AD) using neuroimaging data. Even though the vast majority of these works use the public dataset ADNI, reproducing their results is complicated because they often do not make available elements that are essential for reproducibility, such as selected participants and input data, image preprocessing and cross-validation procedures. Comparability is also an issue. Specially, the influence of different components like preprocessing, feature extraction or classification algorithms on the performance is difficult to evaluate. Finally, these studies rarely compare their results to models built from clinical data only, a critical aspect to demonstrate the utility of neuroimaging. In our previous work, 1, 2 we presented a framework for reproducible and objective classification experiments in AD, that included automatic conversion of ADNI database into the BIDS community standard, image preprocessing pipelines and machine learning evaluation. We applied this framework to perform unimodal classifications of T1 MRI and FDG-PET images. In the present paper, we extend this work to the combination of multimodal clinical and neuroimaging data. All experiments are based on standard approaches (namely SVM and random forests). In particular, we assess the added value of neuroimaging over using only clinical data. We first demonstrate that using only demographic and clinical data (gender, education level, MMSE, CDR sum of boxes, ADASCog) results in a balanced accuracy of 75% (AUC of 0.84). This performance is higher than that of standard

neuroimaging-based classifiers. We then propose a simple trick to improve the performance of neuroimaging-based classifiers: training from AD patients and controls (rather than from MCI patients) improves the performance of FDG-PET classification by 5 percent points, reaching the level of the clinical classifier. Finally, combining clinical and neuroimaging data, prediction results further improved to 80% balanced accuracy and an AUC of 0.88). These prediction accuracies, obtained in a reproducible way, provide a base to develop on top of it and, to compare against, more sophisticated methods. All the code of the framework and the experiments is publicly available at https://github.com/aramis-lab/AD-ML.

More details in [34].

7.3. Disrupted core-periphery structure of multimodal brain networks in Alzheimer's disease

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

In Alzheimer's disease (AD), the progressive atrophy leads to aberrant network reconfigurations both at structural and functional levels. In such network reorganization, the core and peripheral nodes appear to be crucial for the prediction of clinical outcome because of their ability to influence large-scale functional integration. However, the role of the different types of brain connectivity in such prediction still remains unclear. Using a multiplex network approach we integrated information from DWI, fMRI, and MEG brain connectivity to extract an enriched description of the core-periphery structure in a group of AD patients and age-matched controls. Globally, the regional coreness-that is, the probability of a region to be in the multiplex core-significantly decreased in AD patients as result of a random disconnection process initiated by the neurodegeneration. Locally, the most impacted areas were in the core of the network-including temporal, parietal, and occipital areas-while we reported compensatory increments for the peripheral regions in the sensorimotor system. Furthermore, these network changes significantly predicted the cognitive and memory impairment of patients. Taken together these results indicate that a more accurate description of neurodegenerative diseases can be obtained from the multimodal integration of neuroimaging-derived network data.

More details in [20]

7.4. Network neuroscience for optimizing brain–computer interfaces

Participants: Fabrizio de Vico Fallani [Correspondant], Danielle Bassett.

Human-machine interactions are being increasingly explored to create alternative ways of communication and to improve our daily life. Based on a classification of the user's intention from the user's underlying neural activity, brain-computer interfaces (BCIs) allow direct interactions with the external environment while bypassing the traditional effector of the musculoskeletal system. Despite the enormous potential of BCIs, there are still a number of challenges that limit their societal impact, ranging from the correct decoding of a human's thoughts, to the application of effective learning strategies. Despite several important engineering advances, the basic neuroscience behind these challenges remains poorly explored. Indeed, BCIs involve complex dynamic changes related to neural plasticity at a diverse range of spatiotemporal scales. One promising antidote to this complexity lies in network science, which provides a natural language in which to model the organizational principles of brain architecture and function as manifest in its interconnectivity. Here, we briefly review the main limitations currently affecting BCIs, and we offer our perspective on how they can be addressed by means of network theoretic approaches. We posit that the emerging field of network neuroscience will prove to be an effective tool to unlock human-machine interactions.

More details in [13]

7.5. Quality Assessment of Single-Channel EEG for Wearable Devices

Participants: Fanny Grosselin, Xavier Navarro-Sune, Alessia Vozzi, Katerina Pandremmenou, Fabrizio de Vico Fallani, Yohan Attal, Mario Chavez [Correspondant].

The recent embedding of electroencephalographic (EEG) electrodes in wearable devices raises the problem of the quality of the data recorded in such uncontrolled environments. These recordings are often obtained with dry single-channel EEG devices, and may be contaminated by many sources of noise which can compromise the detection and characterization of the brain state studied. In this paper, we propose a classification-based approach to effectively quantify artefact contamination in EEG segments, and discriminate muscular artefacts. The performance of our method were assessed on different databases containing either artificially contaminated or real artefacts recorded with different type of sensors, including wet and dry EEG electrodes. Furthermore, the quality of unlabelled databases was evaluated. For all the studied databases, the proposed method is able to rapidly assess the quality of the EEG signals with an accuracy higher than 90

More details in [19]

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

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

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[7]

7.7. Learning low-dimensional representations of shape data sets with diffeomorphic autoencoders

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

Contemporary deformation-based morphometry offers parametric classes of diffeomorphisms that can be searched to compute the optimal transformation that warps a shape into another, thus defining a similarity metric for shape objects. Extending such classes to capture the geometrical variability in always more varied statistical situations represents an active research topic. This quest for genericity however leads to computationally-intensive estimation problems. Instead, we propose in this work to learn the best-adapted class of diffeomorphisms along with its parametrization, for a shape data set of interest. Optimization is carried out with an auto-encoding variational inference approach, offering in turn a coherent model-estimator pair that we name diffeomorphic auto-encoder. The main contributions are: (i) an original network-based method to construct diffeomorphisms, (ii) a current-splatting layer that allows neural network architectures to process meshes, (iii) illustrations on simulated and real data sets that show differences in the learned statistical distributions of shapes when compared to a standard approach.

More details in[30]

7.8. Learning disease progression models with longitudinal data and missing values

Participants: Raphaël Couronné [correspondant], Marie Vidailhet, Jean-Christophe Corvol, Stéphane Lehéricy, Stanley Durrleman.

Statistical methods have been developed for the analysis of longitudinal data in neurodegenerative diseases. To cope with the lack of temporal markers- i.e. to account for subject-specific disease progression in regard to age- a common strategy consists in realigning the individual sequence data in time. Patient's specific trajectories can indeed be seen as spatiotemporal perturbations of the same normative disease trajectory. However, these models do not easily allow one to account for multimodal data, which more than often include missing values. Indeed, it is rare that imaging and clinical examinations for instance are performed at the same frequency in clinical protocols. Multimodal models also need to allow a different profile of progression for data with different structure and representation. We propose to use a generative mixed effect model that considers the progression trajectories as curves on a Riemannian Manifold. We use the concept of product manifold to handle multimodal data, and leverage the generative aspect of our model to handle missing values. We assess the robuste-ness of our methods toward missing values frequency on both synthetic and real data. Finally we apply our model on a real-world dataset to model Parkinson's disease progression from data derived from clinical examination and imaging.

More details in[31]

7.9. Learning the clustering of longitudinal shape data sets into a mixture of independent or branching trajectories

Participants: Vianney Debavelaere [correspondant], Stéphanie Allassonnière, Stanley Durrleman.

Given repeated observations of several subjects over time, i.e. a longitudinal data set, this work introduces a new model to learn a classification of the shapes progression in an unsupervised setting: we automatically cluster a longitudinal data set in different classes without labels. Our method learns for each cluster an average shape trajectory (or representative curve) and its variance in space and time. Representative trajectories are built as the combination of pieces of curves. This mixture model is flexible enough to handle independent trajectories for each cluster as well as fork and merge scenarios. The estimation of such non linear mixture models in high dimension is known to be difficult because of the trapping states effect that hampers the optimisation of cluster assignments during training. We address this issue by using a tempered version of the stochastic EM algorithm. Finally, we apply our algorithm on different data sets. First, synthetic data are used to show that a tempered scheme achieves better convergence. We then apply our method to different real data sets: 1D RECIST score used to monitor tumors growth, 3D facial expressions and meshes of the hippocampus. In particular, we show how the method can be used to test different scenarios of hippocampus atrophy in ageing by using an heteregenous population of normal ageing individuals and mild cognitive impaired subjects.

More details in[32]

7.10. Auto-encoding meshes of any topology with the current-splatting and exponentiation layers

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

Deep learning has met key applications in image computing, but still lacks processing paradigms for meshes, i.e. collections of elementary geometrical parts such as points, segments or triangles. Meshes are both a powerful representation for geometrical objects, and a challenge for network architectures because of their inherent irregular structure. This work contributes to adapt classical deep learning paradigms to this particular type of data in three ways. First, we introduce the current-splatting layer which embeds meshes in a metric space, allowing the downstream network to process them without any assumption on their topology: they may be composed of varied numbers of elements or connected components, contain holes, or bear high

levels of geometrical noise. Second, we adapt to meshes the exponentiation layer which, from an upstream image array, generates shapes with a diffeomorphic control over their topology. Third, we take advantage of those layers to devise a variational auto-encoding architecture, which we interpret as a generative statistical model that learns adapted low-dimensional representations for mesh data sets. An explicit norm-control layer ensures the correspondence between the latent-space Euclidean metric and the shape-space log-Euclidean one. We illustrate this method on simulated and real data sets, and show the practical relevance of the learned representation for visualization, classification and mesh synthesis.

More details in[29]

7.11. Riemannian Geometry Learning for Disease Progression Modelling

Participants: Maxime Louis, Raphael Couronne, Igor Koval, Benjamin Charlier, Stanley Durrleman.

The analysis of longitudinal trajectories is a longstanding problem in medical imaging which is often tackled in the context of Riemannian geometry: the set of observations is assumed to lie on an a priori known Riemannian manifold. When dealing with high-dimensional or complex data, it is in general not possible to design a Riemannian geometry of relevance. In this work, we perform Riemannian manifold learning in association with the statistical task of longitudinal trajectory analysis. After inference, we obtain both a submanifold of observations and a Riemannian metric so that the observed progressions are geodesics. This is achieved using a deep generative network, which maps trajectories in a low-dimensional Euclidean space to the observation space.

More details in[33]

7.12. How many patients are eligible for disease-modifying treatment in Alzheimer's disease? A French national observational study over 5 years.

Participants: Stéphane Epelbaum [Correspondant], Claire Paquet, Jacques Hugon, Julien Dumurgier, David Wallon, Didier Hannequin, Thérèse Jonveaux, Annick Besozzi, Stéphane Pouponneau, Caroline Hommet, Frédéric Blanc, Laetitia Berly, Adrien Julian, Marc Paccalin, Florence Pasquier, Julie Bellet, Claire Boutoleau-Bretonniere, Tiphaine Charriau, Olivier Rouaud, Olivier Madec, Aurélie Mouton, Renaud David, Samir Bekadar, Roxanne Fabre, Emmanuelle Liegey, Walter Deberdt, Philippe Robert, Bruno Dubois.

We aimed to study the epidemiology of the prodromal and mild stages of Alzheimer's disease (AD) patients who are eligible for clinical trials with disease-modifying therapies. We analyzed two large complementary databases to study the incidence and characteristics of this population on a nationwide scope in France from 2014 to 2018. The National Alzheimer Database contains data from 357 memory centres and 90 private neurologists. Data from 2014 to 2018 have been analyzed. Patients, 50–85 years old, diagnosed with AD who had an Mini-Mental State Exam (MMSE) score greater or equal to 20 were included. We excluded patients with mixed and non-AD neurocognitive disorders. Descriptive statistics of the population of interest was the primary measure. Results In the National Alzheimer Database, 550,198 patients were assessed. Among them, 72,174 (13.1%) were diagnosed with AD and had an MMSE greater or equal to 20. Using corrections for specificity of clinical diagnosis of AD, we estimated that about 50,000 (9.1%) had a prodromal or mild AD. In the combined electronic clinical records database of 11 French expert memory centres, a diagnosis of prodromal or mild AD, certified by the use of cerebrospinal fluid AD biomarkers, could be established in 195 (1.3%) out of 14 596 patients. AD was not frequently diagnosed at a prodromal or mild dementia stage in France in 2014 to 2018. Diagnosis rarely relied on a pathophysiological marker even in expert memory centres. National databases will be valuable to monitor early stage AD diagnosis efficacy in memory centres when a disease-modifying treatment becomes available. More details in [15]

7.13. EEG evidence of compensatory mechanisms in preclinical Alzheimer's disease

Participants: Sinead Gaubert, Federico Raimondo, Marion Houot, Marie-Constance Corsi, Jacobo Diego Sitt, Bertrand Hermann, Delphine Oudiette, Geoffroy Gagliardi, Marie Odile Habert, Bruno Dubois, Fabrizio de Vico Fallani, Hovagim Bakardjian, Stéphane Epelbaum [Correspondant].

Early biomarkers are needed to identify individuals at high risk of preclinical Alzheimer's disease and to better understand the pathophysiological processes of disease progression. Preclinical Alzheimer's disease EEG changes would be non-invasive and cheap screening tools and could also help to predict future progression to clinical Alzheimer's disease. However, the impact of amyloid-beta deposition and neurodegeneration on EEG biomarkers needs to be elucidated. We included participants from the INSIGHT-preAD cohort, which is an ongoing single-centre multimodal observational study that was designed to identify risk factors and markers of progression to clinical Alzheimer's disease in 318 cognitively normal individuals aged 70-85 years with a subjective memory complaint. We divided the subjects into four groups, according to their amyloid status (based on 18F-florbetapir PET) and neurodegeneration status (evidenced by 18F-fluorodeoxyglucose PET brain metabolism in Alzheimer's disease signature regions). The first group was amyloid-positive and neurodegeneration-positive, which corresponds to stage 2 of preclinical Alzheimer's disease. The second group was amyloid-positive and neurodegeneration-negative, which corresponds to stage 1 of preclinical Alzheimer's disease. The third group was amyloid-negative and neurodegeneration-positive, which corresponds to 'suspected non-Alzheimer's pathophysiology'. The last group was the control group, defined by amyloid-negative and neurodegeneration-negative subjects. We analysed 314 baseline 256-channel high-density eyes closed 1-min resting state EEG recordings. EEG biomarkers included spectral measures, algorithmic complexity and functional connectivity assessed with a novel information-theoretic measure, weighted symbolic mutual information. The most prominent effects of neurodegeneration on EEG metrics were localized in frontocentral regions with an increase in high frequency oscillations (higher beta and gamma power) and a decrease in low frequency oscillations (lower delta power), higher spectral entropy, higher complexity and increased functional connectivity measured by weighted symbolic mutual information in theta band. Neurodegeneration was associated with a widespread increase of median spectral frequency. We found a non-linear relationship between amyloid burden and EEG metrics in neurodegeneration-positive subjects, either following a U-shape curve for delta power or an inverted U-shape curve for the other metrics, meaning that EEG patterns are modulated differently depending on the degree of amyloid burden. This finding suggests initial compensatory mechanisms that are overwhelmed for the highest amyloid load. Together, these results indicate that EEG metrics are useful biomarkers for the preclinical stage of Alzheimer's disease.

More details in [17]

7.14. Latent class analysis identifies functional decline with Amsterdam IADL in preclinical Alzheimer's disease

Participants: Sarah-Christine Villeneuve, Marion Houot, Federica Cacciamani, Merike Verrijp, Marie Odile Habert, Bruno Dubois, Sietske Sikkes, Stéphane Epelbaum [Correspondant].

Trials in Alzheimer's disease (AD) now include participants at the earliest stages to prevent further decline. However, the lack of tools sensitive to subtle functional changes in early-stage AD hinders the development of new therapies as it is difficult to prove their clinical relevance. We assessed functional changes over three years in 289 elderly memory complainers from the Investigation of Alzheimer's Predictors in subjective memory complainers cohort using the Am-sterdam Instrumental-Activities-of-Daily-Living questionnaire (A-IADL-Q). No overall functional decline related to AD imaging markers was evidenced. However, five distinct classes of A-IADL-Q trajectories were identified. The largest class (212 [73.4%]) had stable A-IADL-Q scores over 3 years. A second group (23 [8.0%]) showed a persistent functional decline, higher amyloid load (P<.0005), and lower education (P<.0392). The A-IADL-Q identified a subtle functional decline in asymptomatic at-risk AD individuals. This could have important implications in the field of early intervention in AD

More details in [24]

ATHENA Project-Team

7. New Results

7.1. Computational Diffusion MRI

7.1.1. Coarse-Grained Spatiotemporal Acquisition Design for Diffusion MRI

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], Demian Wassermann [Inria Parietal], Rachid Deriche.

Acquisition protocols that allow to capture time-dependent changes in diffusion signal require long imaging time. We address this issue through an optimized subsampling scheme that maximizes accuracy of the spatiotemporal diffusion signal representation, qτ-dMRI, for given time constraints. Our proposed coarse-grained variant of the problem reduces the space of feasible acquisition parameters compared to the fine-grained approach causing no significant deterioration of a reconstruction accuracy in most of the studied cases.

This work has been published in [25].

7.1.2. A Computational Framework For Generating Rotation Invariant Features And Its Application In Diffusion MRI

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

In this work, we present a novel computational framework for analytically generating a complete set of algebraically independent Rotation Invariant Features (RIF) given the Laplace-series expansion of a spherical function. Our computational framework provides a closed-form solution for these new invariants, which are the natural expansion of the well known spherical mean, power-spectrum and bispectrum invariants. We highlight the maximal number of algebraically independent invariants which can be obtained from a truncated Spherical Harmonic (SH) representation of a spherical function and show that most of these new invariants can be linked to statistical and geometrical measures of spherical functions, such as the mean, the variance and the volume of the spherical signal. Moreover, we demonstrate their application to dMRI signal modeling including the Apparent Diffusion Coefficient (ADC), the diffusion signal and the fiber Orientation Distribution Function (fODF). In addition, using both synthetic and real data, we test the ability of our invariants to estimate brain tissue microstructure in healthy subjects and show that our framework provides more flexibility and open up new opportunities for innovative development in the domain of microstructure recovery from diffusion MRI.

This work has been published in [20].

7.1.3. A Novel Characterization of Traumatic Brain Injury in White Matter with Diffusion MRI Spherical-Harmonics Rotation Invariants

Participants: Mauro Zucchelli, Samuel Deslauriers-Gauthier, Drew Parker [Penn Applied Connectomics and Imaging Group, Philadelphia], Junghoon John Kim [Department of Molecular, Cellular & Biomedical Sciences, New York], Ragini Verma [Penn Applied Connectomics and Imaging Group, Philadelphia], Rachid Deriche.

The current DTI-based markers of traumatic brain injury are able to capture affected WM in the brain, but missthe areas of crossing fibers and complex WM due to the simplicity of the model. In this work, we use a novelset of spherical-harmonics rotation invariant indices, recently proposed in the literature. We demonstrate thatthese 12 invariants capture all the information provided by DTI. But in addition, they capture differences incomplex WM, beyond DTI measures. This combined with the clinical feasibility of the method, paves the wayfor them to be used as better markers of brain injury.

This work has been published in [31].

7.1.4. The Dmipy Toolbox: Diffusion MRI Multi-Compartment Modeling and Microstructure Recovery Made Easy

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

Non-invasive estimation of brain microstructure features using diffusion MRI (dMRI)—known as Microstructure Imaging—has become an increasingly diverse and complicated field over the last decades. Multi-compartment (MC)-models, representing the measured diffusion signal as a linear combination of signal models of distinct tissue types, have been developed in many forms to estimate these features. However, a generalized implementation of MC-modeling as a whole, providing deeper insights in its capabilities, remains missing. To address this fact, we present Diffusion Microstructure Imaging in Python (Dmipy), an open-source toolbox implementing PGSE-based MC-modeling in its most general form. Dmipy allows on-the-fly implementation, signal modeling, and optimization of any user-defined MC-model, for any PGSE acquisition scheme. Dmipy follows a "building block"-based philosophy to Microstructure Imaging, meaning MC-models are modularly constructed to include any number and type of tissue models, allowing simultaneous representation of a tissue's diffusivity, orientation, volume fractions, axon orientation dispersion, and axon diameter distribution. In particular, Dmipy is geared toward facilitating reproducible, reliable MC-modeling pipelines, often allowing the whole process from model construction to parameter map recovery in fewer than 10 lines of code. To demonstrate Dmipy's ease of use and potential, we implement a wide range of well-known MC-models, including IVIM, AxCaliber, NODDI(x), Bingham-NODDI, the spherical mean-based SMT and MC-MDI, and spherical convolution-based single- and multi-tissue CSD. By allowing parameter cascading between MC-models, Dmipy also facilitates implementation of advanced approaches like CSD with voxel-varying kernels and single-shell 3-tissue CSD. By providing a well-tested, user-friendly toolbox that simplifies the interaction with the otherwise complicated field of dMRI-based Microstructure Imaging, Dmipy contributes to more reproducible, high-quality research.

This work has been published in [12].

7.1.5. Effects of tractography filtering on the topology and interpretability of connectomes.

Participants: Matteo Frigo, Samuel Deslauriers-Gauthier, Drew Parker [Penn Applied Connectomics and Imaging Group, Philadelphia], Abdol Aziz Ould Ismail [Penn Applied Connectomics and Imaging Group, Philadelphia], Junghoon John Kim [Department of Molecular, Cellular & Biomedical Sciences, New York], Ragini Verma [Penn Applied Connectomics and Imaging Group, Philadelphia], Rachid Deriche.

The analysis of connectomes and their associated network metrics forms an important part of clinical studies. These connectomes are based on tractography algorithms to estimate the structural connectivity between brain regions. However, tractography algorithms, are prone to false positive connections and this affects the quality of the connectomes. Several tractography filtering techniques (TFTs) have been proposed to alleviate this issue in studies, but their effect on connectomic analyses of pathology has not been investigated. The aim of this work is to investigate how TFTs affect network metrics and their interpretation in the context of clinical studies.

This work has been published in [29]

7.1.6. Spherical convolutional neural network for fiber orientation distribution function and micro-structure parameter estimation from dMRI

Participants: Sara Sedlar, Samuel Deslauriers-Gauthier, Théodore Papadopoulo, Rachid Deriche.

Convolutional neural networks (CNNs) are proven to be a powerful tool for many computer vision problems where the data is acquired on a regular grid in Euclidean space. As the dMRI signals used in our experiments are acquired on spheres, we have investigated spherical CNN model (S2-CNN). In regular CNNs, during convolution, kernels are translated over the input feature maps with equidistant steps. In S2-CNN, both kernels and feature maps are represented in the 3D rotation group - SO(3) manifold. A rotation in SO(3) is analogous to a translation in Euclidean space. However, there is no regular equidistant grid in SO(3). As a consequence,

the convolution is performed in the rotational harmonics (Fourier) domain. In this work, we investigate how the S2-CNN can be adapted to properties of dMRI data, such as antipodal symmetry, the presence of Rician noise, multiple sampling shells, etc.

This work currently in progress.

7.1.7. Adaptive phase correction of diffusion-weighted images

Participants: Marco Pizzolato [Signal Processing Lab (LTS5), EPFL, Lausanne], Guillaume Gilbertb [MR Clinical Science, Philips Healthcare Canada, Markham, ON], Jean-Philippe Thiran [Signal Processing Lab (LTS5), EPFL, Lausanne], Maxime Descoteaux [Université de Sherbrooke, Sherbrooke], Rachid Deriche.

Phase correction (PC) is a preprocessing technique that exploits the phase of images acquired in Magnetic Resonance Imaging (MRI) to obtain real-valued images containing tissue contrast with additive Gaussian noise, as opposed to magnitude images which follow a non-Gaussian distribution, e.g. Rician. PC finds its natural application to diffusion-weighted images (DWIs) due to their inherent low signal-to-noise ratio and consequent non-Gaussianity that induces a signal overestimation bias that propagates to the calculated diffusion indices. PC effectiveness depends upon the quality of the phase estimation, which is often performed via a regularization procedure. We show that a suboptimal regularization can produce alterations of the true image contrast in the real-valued phase-corrected images. We propose adaptive phase correction (APC), a method where the phase is estimated by using MRI noise information to perform a complex-valued image regularization that accounts for the local variance of the noise. We show, on synthetic and acquired data, that APC leads to phase-corrected real-valued DWIs that present a reduced number of alterations and a reduced bias. The substantial absence of parameters for which human input is required favors a straightforward integration of APC in MRI processing pipelines.

This work has been published in [17].

7.1.8. Towards validation of diffusion MRI tractography: bridging the resolution gap with 3D Polarized Light Imaging

Participants: Abib Olushola Yessouffou Alimi, Samuel Deslauriers-Gauthier, Rachid Deriche.

Three-dimensional Polarized Light Imaging (3D-PLI) is an optical approach presented as a good candidate for validation of diffusion Magnetic Resonance Imaging (dMRI) results such as orientation estimates (fiber Orientation Distribution Functions) and tractography. We developed an anlytical approach to reconstruct fiber ODFs from 3D-PLI datasets. From these fODFs, here we compute brain fiber tracts via dMRI-based probabilistic tractography algorithm. Reconstructed fODFs at different scales proves the ability to bridge the resolution gap between 3D-PLI and dMRI, demonstrating, therefore, a great promise to validate diffusion MRI tractography thanks to multi-scale fiber tracking based on 3D-PLI.

This work has been published in [21].

7.1.9. Analytical Fiber ODF Reconstruction in 3D Polarized Light Imaging: Performance Assessment

Participants: Abib Olushola Yessouffou Alimi, Samuel Deslauriers-Gauthier, Felix Matuschke [INM-1 - Institute of Neuroscience and Medicine, Jülich], Daniel Schmitz [INM-1 - Institute of Neuroscience and Medicine, Jülich], Markus Axer [INM-1 - Institute of Neuroscience and Medicine, Jülich], Rachid Deriche.

Three dimensional Polarized Light Imaging (3D-PLI) allows to map the spatial fiber structure of postmortem tissue at a sub-millimeter resolution, thanks to its birefringence property. Different methods have been recently proposed to reconstruct the fiber orientation distribution function (fODF) from high-resolution vector data provided by 3D-PLI. Here, we focus on the analytical fODF computation approach, which uses the spherical harmonics to represent the fODF and analytically computes the spherical harmonics coefficients via the spherical Fourier transform. This work deals with the assessment of the performance of this approach on rich synthetic data which simulates the geometry of the neuronal fibers and on human brain dataset. A computational complexity and robustness to noise analysis demonstrate the interest and great potential of the approach.

This work has been published in [22].

7.2. Unveiling brain activity using M/EEG

7.2.1. Fast Approximation of EEG Forward Problem and Application to Tissue Conductivity Estimation

Participants: Kostiantyn Maksymenko, Maureen Clerc, 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 fine 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 published in [15].

7.2.2. 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 has been published in [8].

7.2.3. Convolutional autoencoder for waveform learning

Participants: Sara Sedlar, Maureen Clerc, Rachid Deriche, Théodore Papadopoulo.

Electro- or Magneto-encephalographic (M/EEG) signals measured on the scalp can be modeled as a linear combination of source signals occurring in different cortical regions. Analysis of specific recurrent waveforms from measurements can help in the evaluation of several neurological disorders such as epilepsy, Alzheimer's disease, and narcolepsy. In addition, detection of the neural events evoked by certain stimuli is crucial for brain-computer interfaces. Such M/EEG signals are quite faint and inherently affected by an important noise, generated by irrelevant brain activities, by other organs, by external ambient noise or imperfections of the measuring devices. In addition, there are intra- and inter-subject variabilities, meaning that the relevant waveforms vary in terms of amplitudes, shapes, and time delays. This makes waveform learning on such signals a quite complex task. In order to address these problems, a number of dictionary (here waveforms) learning based approaches has been proposed. The common framework behind those approaches is an

alternative estimation of data-driven waveforms and their corresponding activations in terms of amplitudes and positions over time. Motivated by the success of these methods and the advances in deep learning, we propose a method based on a convolutional auto-encoder that aims at improving more traditional approaches. Auto-encoders are unsupervised neural network models that have been successfully used for data compression, feature learning, denoising and clustering. Auto-encoders are composed of an encoder which creates a code also known as bottle-neck and decoder that is supposed to reconstruct input signal given the code. By penalizing reconstruction loss function with certain constraints we can guide the auto-encoder to perform compression, denoising, clustering etc. For the moment, the properties of the model are investigated on single-channel synthetic data imitating three types of neurological activities (spikes, short oscillatory and low frequency saw-tooth waveforms) mixed using a realistic leadfield matrix (source space to sensor space transform).

This work is in current progress.

7.2.4. Automatic detection of epileptic seizures by video-EEG

Participants: Mamoudou Sano, Hugo Cadis [IPMC], Fabrice Duprat [IPMC], Massimo Mantegazza [IPMC], Maureen Clerc, Théodore Papadopoulo.

Epilepsy is a serious condition that affects almost 50 million people worldwide. Despite several generations of antiepileptic treatments, the rate of drug-resistant patients remains around 30% and the discovery of new pharmacological targets is therefore a crucial issue.

In order to find pharmacological targets, several animal models make it possible to study the mechanisms of establishment of epileptic disease, or epileptogenesis, and the consequences of repeated spontaneous attacks which characterize epilepsy. Recording an electroencephalogram (EEG) remains the best way to understand these mechanisms. However, the placement of electrodes on small animals such as mice is difficult or even impossible depending on the age of the animal or other used protocols. The use of video recordings over several days, weeks or months makes it possible to observe the animals with a minimum of disturbances and to assess the severity of the crises on a behavioral scale. In both cases, the visual analysis of hundreds of hours of video and/or EEG recordings is very long and error-prone.

The goal of this joint IPMC, ATHENA work was to improve acquisition techniques and develop software tools to automate both EEG and video analysis. EEG analysis was based on the "Adaptive Waveform Learning" that was developed in the group a few years ago [61]. This is work in progress.

7.3. Combined fMRI, M/EEG and dMRI

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

Participants: Samuel Deslauriers-Gauthier, Jean-Marc Lina [ETS - Ecole de Technologie Supérieure, Montréal], Russel Butler [Université de Sherbrooke, Sherbrooke], Kevin Whittingstall [Université de Sherbrooke, Sherbrooke], Pierre-Michel Bernier [Université de Sherbrooke, Sherbrooke], Maxime Descoteaux [Université de Sherbrooke, Sherbrooke], 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 has been published in [11].

7.3.2. Structural connectivity to reconstruct brain activation and effective connectivity between brain regions

Participants: Brahim Belaoucha, Théodore Papadopoulo.

Understanding how brain regions interact to perform a specific task is very challenging. EEG and MEG are two non-invasive imaging modalities that allow the measurement of brain activation with high temporal resolution. Several works in EEG/MEG source reconstruction show that estimating brain activation can be improved by considering spatio-temporal constraints but only few of them use structural information to do so. In this work, we present a source reconstruction algorithm that uses brain structural connectivity, estimated from diffusion MRI (dMRI), to constrain the EEG/MEG source reconstruction. Contrarily to most source reconstruction methods which reconstruct activation for each time instant, the proposed method estimates an initial reconstruction for the first time instants and a multivariate autoregressive model that explains the data in further time instants. This autoregressive model can be thought as an estimation of the effective connectivity between brain regions. We called this algorithm iterative Source and Dynamics reconstruction (iSDR). This paper presents the overall iSDR approach and how the proposed model is optimized to obtain both brain activation and brain region interactions. The accuracy of our method is demonstrated using synthetic data in which it shows a good capability to reconstruct both activation and connectivity. iSDR is also tested with real data (face recognition task). The results are in phase with other works published with the same data and others that used different imaging modalities with the same task showing that the choice of using an autoregressive model gives relevant results.

This work has been submitted to the non-invasive brain imaging special issue of Journal of Neural Engineering.

7.3.3. Estimation of Axon Conduction Delay, Conduction Speed, and Diameter from Information Flow using Diffusion MRI and MEG.

Participants: Samuel Deslauriers-Gauthier, Rachid Deriche.

The different lengths and conduction velocities of axons connecting cortical regions of the brain yield information transmission delays which are believed to be fundamental to brain dynamics. While early work on axon conduction velocity was based on ex vivo measurements , more recent work makes use of a combination of diffusion Magnetic Resonance Imaging (MRI) tractography and electroencephalography (EEG) to estimate axon conduction velocity in vivo. An essential intermediary step in this later strategy is to estimate the inter hemispheric transfer time (IHTT) using EEG. The IHTT is estimated by measuring the latency between the peaks or by computing the lag to maximum correlation on contra lateral electrodes. These approaches do not take the subjects anatomy into account and, due to the limited number of electrodes used, only partially leverage the information provided by EEG. In our previous work, we proposed a method, named Connectivity Informed Maximum Entropy on the Mean (CIMEM), to estimate information flow in the white matter of the brain. CIMEM is built around a Bayesian network which represents the cortical regions of the brain and their connections, observed using diffusion MRI tractography. This Bayesian network is used to constrain the EEG inverse problem and estimate which white matter connections are used to transfer information between cortical regions. In our previous work, CIMEM was used to infer the information flow in the white matter by assuming a constant conduction velocity for all connections. In this context, the conduction speed, and thus the delays, were inputs used to help constrain the problem. Here, we instead assume that the connection used

to transfer information across the hemispheres is known, due the design of the acquisition paradigm, but that its conduction velocity must be estimated.

This work has been published in [23].

7.3.4. Estimation of Axonal Conduction Speed and the Inter Hemispheric Transfer Time using Connectivity Informed Maximum Entropy on the Mean

Participants: Samuel Deslauriers-Gauthier, Rachid Deriche.

The different lengths and conduction velocities of axons connecting cortical regions of the brain yield information transmission delays which are believed to be fundamental to brain dynamics. A critical step in the estimation of axon conduction speed in vivo is the estimation of the inter hemispheric transfer time (IHTT). The IHTT is estimated using electroencephalography (EEG) by measuring the latency between the peaks of specific electrodes or by computing the lag to maximum correlation on contra lateral electrodes. These approaches do not take the subject's anatomy into account and, due to the limited number of electrodes used, only partially leverage the information provided by EEG. Using the previous published Connectivity Informed Maximum Entropy on the Mean (CIMEM) method, we propose a new approach to estimate the IHTT. In CIMEM, a Bayesian network is built using the structural connectivity information between cortical regions. EEG signals are then used as evidence into this network to compute the posterior probability of a connection being active at a particular time. Here, we propose a new quantity which measures how much of the EEG signals are supported by connections, which is maximized when the correct conduction delays are used. Using simulations, we show that CIMEM provides a more accurate estimation of the IHTT compared to the peak latency and lag to maximum correlation methods.

This work has been published in [24].

7.3.5. A Unified Model for Structure–function Mapping Based on Eigenmodes

Participants: Samuel Deslauriers-Gauthier, Rachid Deriche.

Characterizing the connection between brain structure and brain function is essential for understanding how behaviour emerges from the underlying anatomy. To this end, a common representation of the brain is that of a network, where nodes represent cortical and sub–cortical gray matter volumes and edges represent the strength of structural or functional connectivity. A convenient representation of this network is that of a matrix, where entries represent the strength of the structural connectivity (SC) or functional connectivity (FC) between nodes. A number of studies have shown that the network structure of the white matter shapes functional connectivity, leading to the idea that it should be possible to predict the function given the structure. A strategy is to learn a direct mapping from the SC matrix to the FC matrix. In this work, we show that the mappings currently proposed in the literature can be generalized to a single model and that this model can be used to generate new structure-function mappings. We tested our general model on 40 subjects of the Human Connectome Project and demonstrated that for specific choices of parameters, our model reduces to previously proposed models and yields comparable results. However, by allowing to choose the eigenvalue and eigenvector mapping independently, our models can also produce novel mapping that improve the prediction of FC from SC.

This work is currently under submission to OHBM.

7.3.6. Connectivity-informed spatio-temporal MEG source reconstruction: Simulation results using a MAR model

Participants: Ivana Kojcic, Théodore Papadopoulo, Samuel Deslauriers-Gauthier, Rachid Deriche.

Recovering brain activity from M/EEG measurements is an ill-posed problem and prior constraints need to be introduced in order to obtain unique solution. The majority of the methods use spatial and/or temporal constraints, without taking account of long-range connectivity. In this work, we propose a new connectivity-informed spatio-temporal approach to constrain the inverse problem using supplementary information coming from diffusion MRI. We present results based on simulated brain activity using a Multivariate Autoregressive Model, with realistic subject anatomy obtained from Human Connectome Project dataset.

This work has been published in [35].

7.3.7. Connectivity-informed solution for spatio-temporal M/EEG source reconstruction

Participants: Ivana Kojcic, Théodore Papadopoulo, Samuel Deslauriers-Gauthier, Rachid Deriche.

Recovering brain activity from M/EEG measurements is an ill–posed problem and prior constraints need to be introduced in order to obtain unique solution. The majority of the methods use spatial and/or temporal constraints, without taking account of long–range connectivity. In this work, we propose a new connectivity–informed spatio–temporal approach to constrain the inverse problem using supplementary information coming from difusion MRI. We present results based on simulated brain activity obtained with realistic subject anatomy from Human Connectome Project dataset.

This work has been published in [34].

7.3.8. Deconvolution of fMRI Data using a Paradigm Free Iterative Approach based on Partial Differential Equations

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

Functional magnetic resonance imaging (fMRI) is a technique which indirectly measures neural activations via the blood oxygenated level dependent (BOLD) signal. So far, few approaches have been proposed to regularize the fMRI data, while recovering the underlying activations at the voxel level. In particular, for task fMRI, voxels time courses are fitted on a given experimental paradigm. To avoid the necessity of a priori information on the pattern, supposing the brain works with blocks of constant activation, Farouj et al. has developed a deconvolution approach which solves the optimizations problem by splitting it into two regularization problems, i.e. spatial and temporal. Starting from this idea, we propose a paradigm-free iterative algorithm based on partial differential equations (PDEs) which minimizes the image variations, while preserving sharp transitions (i.e. brain activations), in the space and the time dimensions at once.

This work has been published in [27].

7.3.9. Novel 4-D Algorithm for Functional MRI Image Regularization using Partial Differential Equations

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

State-of-the-art techniques for denoising functional MRI (fMRI) images consider the problems of spatial and temporal regularization as decoupled tasks. In this work we propose a partial differential equations (PDEs)-based algorithm that acts directly on the 4-D fMRI image. Our approach is based on the idea that large image variations should be preserved as they occur during brain activation, but small variations should be smoothed to remove noise. Starting from this principle, by means of PDEs we were able to smooth the fMRI image with an anisotropic regularization, thus recovering the location of the brain activations in space and their timing and duration.

This work has been published in [28].

7.3.10. Spatially Varying Monte Carlo Sure for the Regularization of Biomedical Images

Participants: Marco Pizzolato [Signal Processing Lab (LTS5), EPFL, Lausanne], Erick Jorge Canales-Rodríguez [Radiology Department CHUV, Lausanne], Jean-Philippe Thiran [Signal Processing Lab (LTS5), EPFL, Lausanne], Rachid Deriche.

Regularization, filtering, and denoising of biomedical images requires the use of appropriate filters and the adoption of efficient regularization criteria. It has been shown that the Stein's Unbiased Risk Estimate (SURE) can be used as a proxy for the mean squared error (MSE), thus giving an effective criterion for choosing the regularization amount as to that minimizing SURE. Often, due to the complexity of the adopted filters and solvers, this proxy must be calculated with a Monte Carlo method. In practical biomedical applications, however, images are affected by spatially-varying noise distributions, which must be taken into account. We propose a modification to the Monte Carlo method, called svSURE, that accounts for the spatial variability of the noise variance, and show that it correctly estimates the MSE in such cases.

This work has been published in [30].

7.3.11. The visual word form area (VWFA) is part of both language and attention circuitry

Participants: Lang Chen, Demian Wasserman, Daniel Abrams, John Kochalka, Guillermo Gallardo-Diez, Vinod Menon.

While predominant models of visual word form area (VWFA) function argue for its specific role in decoding written language, other accounts propose a more general role of VWFA in complex visual processing. However, a comprehensive examination of structural and functional VWFA circuits and their relationship to behavior has been missing. Here, using high-resolution multimodal imaging data from a large Human Connectome Project cohort (N = 313), we demonstrate robust patterns of VWFA connectivity with both canonical language and attentional networks. Brain-behavior relationships revealed a striking pattern of double dissociation: structural connectivity of VWFA with lateral temporal language network predicted language, but not visuo-spatial attention abilities, while VWFA connectivity with dorsal fronto-parietal attention network predicted visuo-spatial attention, but not language abilities. Our findings support a multiplex model of VWFA function characterized by distinct circuits for integrating language and attention, and point to connectivity-constrained cognition as a key principle of human brain organization.

This work has been published in [10].

7.4. Brain Computer Interfaces

7.4.1. Augmenting Motor Imagery Learning for Brain–Computer Interfacing Using Electrical Stimulation as Feedback

Participants: Saugat Bhattacharyya [School of Bio-Science and Engineering, Calcutta], Mitsuhiro Hayashibe [Tohoku University, Sendai], Maureen Clerc.

Brain-computer Interfaces (BCI) and Functional electrical stimulation (FES) contribute significantly to induce cortical learning and to elicit peripheral neuronal activation processes and thus, are highly effective to promote motor recovery. This study aims at understanding the effect of FES as a neural feedback and its influence on the learning process for motor imagery tasks while comparing its performance with a classical visual feedback protocol. The participants were randomly separated into two groups: one group was provided with visual feedback (VIS) while the other received electrical stimulation (FES) as feedback. Both groups performed various motor imagery tasks while feedback was provided in form of a bi-directional bar for VIS group and targeted electrical stimulation on the upper and lower limbs for FES group. The results shown in this paper suggest that the FES based feedback is more intuitive to the participants, hence, the superior results as compared to the visual feedback. The results suggest that the convergence of BCI with FES modality could improve the learning of the patients both in terms of accuracy and speed and provide a practical solution to the BCI learning process in rehabilitation.

This work, obtained in the context of the BCI-LIFT IPL, has been published in [9].

7.4.2. Adaptive parameter setting in a code modulated visual evoked potentials BCI

Participants: Federica Turi, Maureen Clerc.

Code-modulated visual evoked potentials (c-VEPs) BCI are designed for high-speed communication. The setting of stimulus parameters is fundamental for this type of BCI, because stimulus parameters have an influence on the performance of the system. In this work we design a c-VEP BCI for word spelling, in which it is possible to find the optimal stimulus presentation rate per each subject thanks to an adaptive setting parameter phase. This phase takes place at the beginning of each session and allows to define the stimulus parameters that are used during the spelling phase. The different stimuli are modulated by a binary m-sequence circular-shifted by a different time lag and a template matching method is applied for the target detection. We acquired data from 4 subjects in two sessions. The results obtained for the offline spelling show the variability between subjects and therefore the importance of subject-dependent adaptation of c-VEP BCI.

This work has been published in [32].

7.4.3. Participation to the Cybathlon BCI Series

Participants: Karine Leclerc [Centre René Labreuille, Le Cannet], Magali Mambrucchi [Centre René Labreuille, Le Cannet], Amandine Audino, Pierre Giacalone, Federica Turi, Maureen Clerc, Théodore Papadopoulo.

The CYBATHLON is a unique championship in which people with physical disabilities compete against each other to complete everyday tasks using state-of-the-art technical assistance systems. Athena participated in the CYBATHLON BCI Series that took place on September 8th, 2019 as a satellite event of the Graz Brain–Computer Interface Conference. Athena was part of a bigger Inria team which encompassed also the Inria Bordeaux Sud-Ouest Potioc team (participants from Bordeaux are not listed). For both Inria sub-teams, it was a first participation to such a competition : we learned a lot about the practical issues of working with people with physical disabilities and on all the practical issues that can encounter a BCI user out of the lab. The actual competition consisted of driving a car on a track by issuing three types of commands (Left, Right, Lights) using mental imagery. Even though our pilot finished last, she was for each run leading the race till a few seconds before its end. A great satisfaction was to see that the software that we built worked reliably out of the lab (many teams have had troubles in issuing commands and had to redo a race). Yet, this required a lot of last minute work to integrate smoothly in the competition system: we learned a lot in this respect. The poster [36] summarises this effort.

7.4.4. BCI Performance prediction

Participants: Maureen Clerc, Nathalie Gayraud, Laurent Bougrain [NeuroSys Project-Team], Sébastien Rimbert [NeuroSys Project-Team], Stéphanie Fleck [Perseus].

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). Our results showed 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. This work is an outcome of the BCI-LIFT IPL and was published in [18]. Another joint publication [19] investigated median nerve stimulation as a new approach to detect intraoperative awareness during General Anesthesia.

7.4.5. EEG Classification of Auditory Attention

Participants: Joan Belo, Johann Benerradi, Maureen Clerc, Michel Pascal [Nice Music Conservatory], Daniele Schön [Institut de Neurosciences des Systèmes].

In a Master's thesis [33] in collaboration with Nice Music Conservatory and Institut de Neurosciences des Systèmes, we focused on analyzing auditory attention of human participants who are presented two auditory streams, simultaneously on left and right. By analyzing the EEG signals measured, the problem is to detect to which stream the participant is attending. Auditory Attention is also the topic of the PhD thesis of Joan Belo, funded by a CIFRE with Oticon Medical.

7.4.6. Innovative Brain-Computer Interface based on motor cortex activity to detect accidental awareness during general anesthesia

Participants: Sébastien Rimbert, Philippe Guerci, Nathalie Gayraud, Claude Meistelman, Laurent Bougrain.

Accidental Awareness during General Anesthesia (AAGA) occurs in 1-2% of high-risk practice patients and is responsible for severe psychological trauma, termed post-traumatic stress disorder (PTSD). Currently, monitoring techniques have limited accuracy in predicting or detecting AAGA. Since the first reflex of a patient experiencing AAGA is to move, a passive Brain-Computer Interface (BCI) based on the detection of an intention of movement would be conceivable to alert the anesthetist and prevent this phenomenon. However, the way in which the propofol (an anesthetic drug commonly used for inducing and maintaining general anesthesia) affects the motor brain activity and is reflected by the electroencephalo-graphic (EEG) signal has been poorly investigated and is not clearly understood. The goal of this forward-looking study is to investigate the motor activity behavior with step-wise increase of propofol doses in 4 healthy subjects and provide a proof of concept for such an innovative BCI.

This work has been published in [26].

BIOVISION Project-Team

7. New Results

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

7.1.1. Multilayered Analysis of Newspaper Structure and Design

Participants: Hui-Yin Wu, Pierre Kornprobst.

The understanding of newspaper document structure can help in the adaptation of text and visual content for different devices and media [53], as well as, in the context of low vision, to enhance accessibility by combining magnification and text-to-speech aids. However, automated segmentation of complex document structures like newspapers remains an ongoing challenge due its dense layout with numerous visual and textual design elements [38], [44].

To address this challenge, we propose a multi-layered analysis of structure and design, presented in [27]. Taking images of newspaper front pages as input, our approach uses a combination of computer vision techniques to segment newspapers with complex layouts into meaningful blocks of varying degrees of granularity, and convolutional neural network (CNN) to classify each block. The final output presents a visualization of the various design elements present in the newspaper such as in Figure 2 . Compared to previous approaches, our method introduces a much larger set of design-related labels (23 labels against less than 10 before) resulting in a very fine description of the pages, with high accuracy (83%), as shown in Figure 3 .

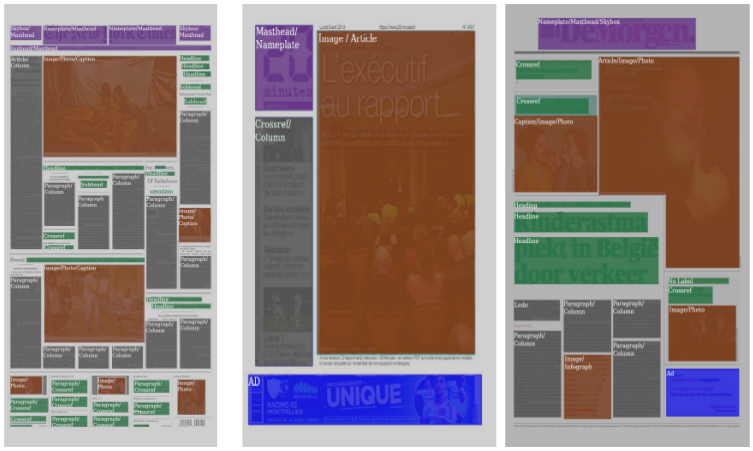


Figure 2. Visualization of the classification results on three different newspapers in our test set. Colors indicate primary categories as masthead elements (purple), text column (gray), ads (blue), images (brown) and minor text elements (green). Original images copyright of (from left to right) New York Times, 20 Minutes, and DeMorgan, courtesy of Newseum.

This work is presented in [27].

7.1.2. Towards accessible news reading design in virtual reality for low vision

Participants: Hui-Yin Wu, Aurélie Calabrèse, Pierre Kornprobst.

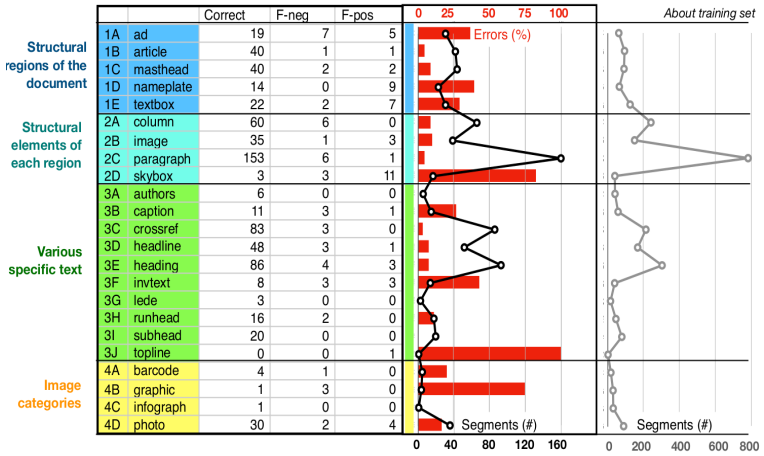


Figure 3. Classification result per categories. Results are presented in the table, showing the number of design elements that are correctly classified, false-negative (i.e. missing label), and false-positive (i.e. wrongly assigned label). Then a red chart shows the errors together with the number of segments (black line) used for the test dataset. To the right, the grey curve indicates the number of segments which were available in the training set.

Low-vision conditions resulting in partial loss of the visual field strongly affect patients' daily tasks and routines, and none more prominently than the ability to access text. Though vision aids such as magnifiers, digital screens, and text-to-speech devices can improve overall accessibility to text, news media, which is non-linear and has complex and volatile formatting, bars low-vision patients from easy access to essential news content [54].

Our aim is to position virtual reality as the next step towards accessible and enjoyable news reading for the low vision. Our ongoing work, which we present in [26], consists of an extensive review into existing research on low-vision reading technologies and accessibility for modern news media. From previous research and studies, we then conduct an analysis into the advantages of virtual reality for low-vision reading and propose comprehensive guidelines for visual accessibility design in virtual reality, with a focus on reading. This is coupled with a hands-on study of eight reading applications in virtual reality to evaluate how accessibility design is currently implemented in existing products. Finally, we present a framework that integrates the design principles resulting from our analysis and study, and implement a proof-of-concept for this framework using browser-based graphics (Figure 4 and 5) to demonstrate the feasibility of our proposal with modern virtual reality technology.

This work is presented in [26].

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

7.2.1. From micro- to macroscopic description of the retina

7.2.1.1. Retinal Waves

Participants: Dora Matzakos-Karvouniari [Laboratoire Jean-Alexandre Dieudonné, (LJAD), Nice, France], Bruno Cessac, Lionel Gil [Institut de Physique de Nice (InPhyNi), France].

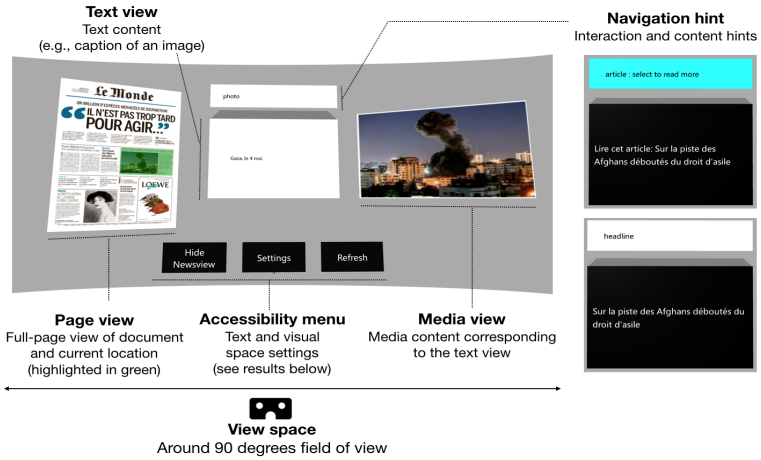


Figure 4. Application prototype: The global overview of the newspaper page is shown side-by-side with the enlarged text and images of the highlighted region. Navigation hints above the card show what type of content is displayed (e.g. photo, heading, paragraph) and whether the card can be selected (i.e. highlighted in light blue) to reveal further content. Text and images of the newspaper are purely for demonstrating a proof-of-concept. Excerpted from 7 May 2019 issue of ©Le Monde.

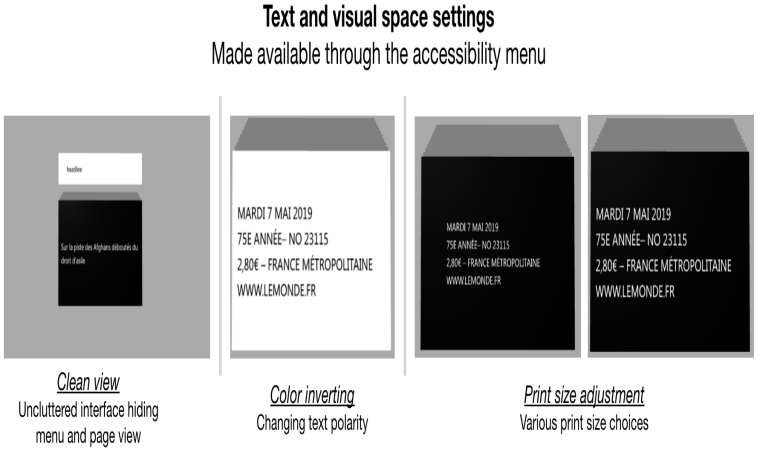


Figure 5. The accessibility menu provides a number of functions including (1) showing/hiding the page and menu view to personalize the view space, (2) invert foreground and background color only for text content, and (3) change the print size. Text on the cards excerpted from 7 May 2019 issue of ©Le Monde

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 [63], [47], [58], [48]. 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 [46]. 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 [64].

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 [14], [34]. 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.

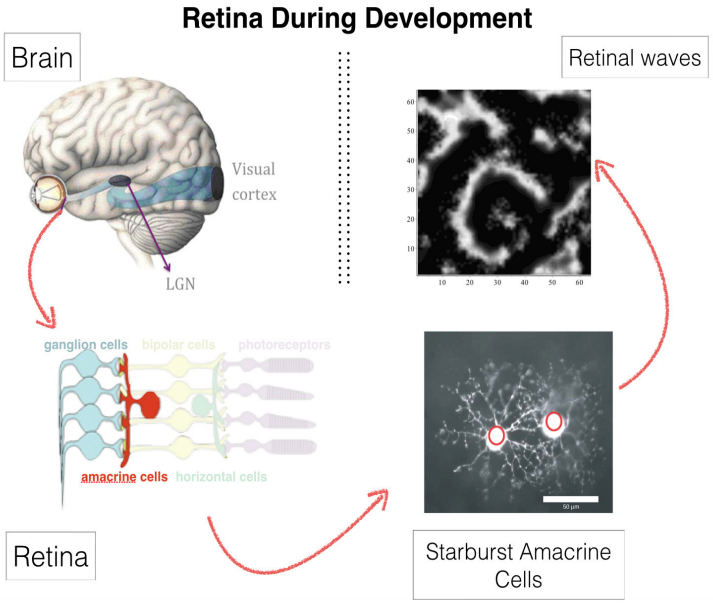


Figure 6. Top left. View of the human visual system. Bottom left. Right after birth the retina is not fully developed (shadowed parts). The retinal waves (top right) contributes to this development. They are mediated by specific cells (Starburst Amacrine Cells in stage II, bottom right).

Based on this biophysical model we have analysed here the dynamics of retinal waves and their statistics. We show that, despite the acetylcholine coupling intensity has been experimentally observed to change during development, SACs retinal waves can nevertheless stay in a regime with power law distributions, reminiscent of a critical regime. Thus, this regime occurs on a range of coupling parameters instead of a single point as in usual phase transitions. We explain this phenomenon thanks to a coherence-resonance mechanism, where noise is responsible for the broadening of the critical coupling strength range. This work has been presented in [14], [16], [25]

7.2.1.2. Anticipation in the retina and the visual cortex V1

Participants: Bruno Cessac, Frédéric Chavane [Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France)], Alain Destexhe [Institute de Neuroscience Paris-Saclay (UNIC)], Sandrine

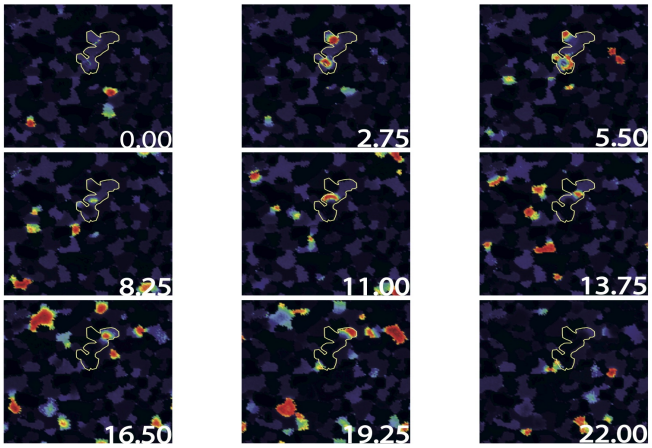


Figure 7. Example of the two dimensional time evolution of the calcium concentration C. Dark regions correspond to low calcium concentration while red corresponds to high concentration (wave). The time (in s) is displayed in the bottom right corner. The same thin white line in the center of each image, delimits a closed domain where wave propagates almost periodically. Hence, after the sequence shown above which correspond to a single period, a new one takes place a time later with a new wave following almost the same trajectory. The domain delimited by the white line is circled with high sAHP regions and therefore slowly evolves with time. A numerical movie is available on the website : https://www.youtube.com/watch?v=shMR3NMCBDE

Chemla [Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France)], Selma Souihel, Matteo Di Volo [Institute de Neuroscience Paris-Saclay (UNIC)].

This work has been done in the context of the ANR Trajectory and Selma Souihel Thesis [11].

Vision is initiated in the retina, where light is converted into electrical signals by photoreceptors, sent to bipolar cells then ganglion cells, generating spike trains. Visual information is then transmitted to the thalamus via the optic nerve which in turn transmits it to the visual cortex. The retinal processing alone takes time, up to 150 ms, not to mention the time lags introduced by synaptic transmissions between the three processing units. This shows that the existence of compensatory mechanisms to reduce processing delays is absolutely essential. These compensatory mechanisms are known as anticipation. Anticipation first occurs at the level of the retina and is further carried out by the primary visual cortex. In its first occurrence, anticipation is either characterized by a shift in the the peak response, or a short range wave of activation. In the second case, it is characterized by a wider range wave of activation.

The first contribution of this work is the development of a generalized 2D model of the retina, mimicking three types of ganglion cells : Fast OFF cells with gain control, direction selective cells with gap junction connectivity, and differential motion cells connected through an upstream amacrine circuit, able of anticipating different kinds of moving stimuli. The second contribution is to use our retina model as an input to a mean field cortical model to reproduce motion anticipation as observed in voltage sensitive dye imaging recordings. Throughout our work, we will study the effect of non linear phenomena involved in anticipation, as well as connectivity, both at the level of the retina and the primary visual cortex. The integrated retino-cortical model allowed us to study the effects of anticipation on two-dimensional stimuli, and to highlight the collaborative aspect of anticipation mechanisms in the retina and the cortex.

This work has been presented in [17], [22], [20], [34], [21], [11]

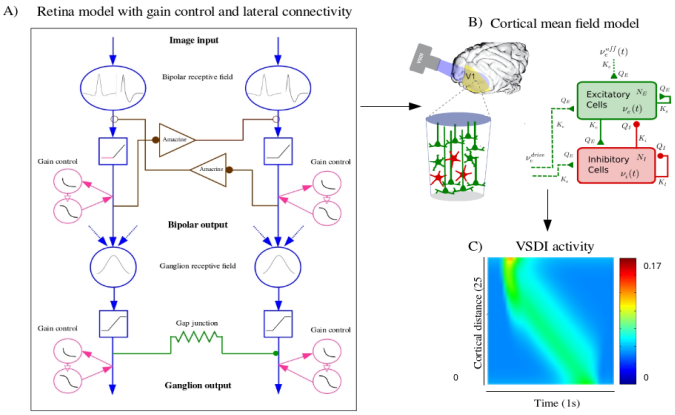


Figure 8. Schematic of our retino cortical model for anticipation. Left. Structure of the retinal model with 3 pathways: Blue: Gain control. Green: Gap junctions laterally connecting Ganglion Cells. Brown. Amacrine cells lateral connectivity. Right top. Cortical model (from Destexhe-Boustani 2009). Bottom. Simulation of VSDI activity in response to a moving bar.

7.2.1.3. Dynamical synapse in the retina

Participants: Bruno Cessac, Simone Ebert, Olivier Marre [Institut de la Vision (IdV), Paris, France], Romain Veltz [MathNeuro].

A very sophisticated example of the computations within the retina is observed when the visual system is exposed to a periodic stimulus, such as a regular series of flashes. If the retina would simply respond proportionally to the stimulus input, one might expect that ganglion cells would become entrained into aperiodic activity, responding to each flash. When the stimulus sequence ends,the activity would end as well. However, ganglion cells can exhibit various different kinds of response patterns to this form of stimulation shown in Figure 3. At the beginning of the stimulus, cells typically respond to the first flash of this new stimulus with a peak of activity, but then very rapidly decay in the amplitude of their response to the following flashes. Most remarkably, when the flash sequence ends ganglion cells do not just stop to respond, but in fact may generate a pulse of activity signalling the missing stimulus. This property of indicating a deviation from an expected pattern has been termed the Omitted Stimulus Response (OSR) (Schwartz, Harris, Shrom, Berry, 2007). The aim of this study was to implement and compare the two existing models of Omitted Stimulus Response in the retina, as well as exploring potential mechanisms that may be involved in generating it. Especially synaptic mechanisms may provide an explanation here, but the integration of such a mechanism into an OSR model has not been explored yet. A potential synaptic property that provides an interesting candidate to test here would be short-term plasticity (STP), which modulates synaptic efficacy depending on the previous activity in a short time interval (Blitz, Foster, Regehr, 2004). STP thus modulates signal transmission and a consecutive spike pattern and has been found to take place within the retina (Dunn, Rieke, 2008). Examining the models' underlying mechanisms, advantages and disadvantages as well as similarities and differences will provide a good foundation to modify existing models by adding potential mechanisms and exploring their effect on a ganglion cell response to a periodic stimulus. Ultimately, this may help shedding light on cellular properties in a neuronal circuit as of as few as 3 cells can contribute to already interpreting information from the environment.

This work has been presented in [32]. It has lead to experiments done in the Institut de la Vision by S. Ebert (internship Biovision) and O. Marre (Institut de la Vision (IdV), Paris, France) in November 2019.

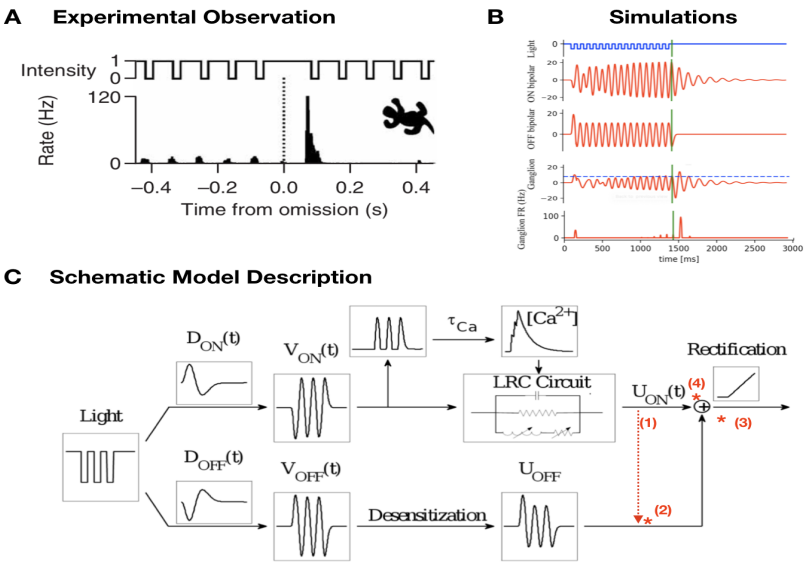


Figure 9. A. Experimentally observed Omitted Stimulus Response (OSR) to a periodic flash sequence. B. Simulations performed with an existing Model from Gao et al., 2009. C. Schematic description of the used 'Calcium-tuned Oscillator' Model from Gao et al., 2009. It is based on a feedforward circuit consisting of two different pathways with different intrinsic processing steps. Both pathways are combined to represent the synaptic input ganglion cells, who's activity with generate an Omitted Stimulus Response. Planned modifications planned are marked in red. (1) a presynaptic connection before summation of both pathways. (2),(3),(4) are synapses where short term synaptic plasticity could occur.

7.2.2. Numerical modelling of the retina in normal and pathological conditions

7.2.2.1. Probing retinal function with a multi-layered simulator

Participants: Bruno Cessac, Gerrit Hilgen [Institute of Neuroscience (ION), Newcastle, UK], Evgenia Kartsaki, Evelyne Sernagor [Institute of Neuroscience (ION), Newcastle, UK].

Our brain can recreate images from interpreting a stream of information emitted by one million parallel channels in the retina. This ability is partly due to the astonishing functional and anatomical diversity of the retinal ganglion cells (RGCs), each interpreting a different feature of the visual scene. How precisely this complexity is encoded in the spike trains produced by the population of RGCs is, however, largely unknown. Adding to the complexity, RGCs "speak" to each other during complex tasks (especially motion handling), via amacrine cells (ACs - lateral connectivity). To decipher their role, we study an experimental setting where neurons co-express the genes Grik4 or Scnn1a and excitatory or inhibitory DREADDs (Designer Receptors Exclusively Activated by Designer Drugs), activated by the designer drug CNO. Switching on or off RGCs and/or ACs cells may not only impact the RGCs individual response but also their concerted activity to different stimuli, thus allowing us to understand how they contribute to the encoding of complex visual scenes. However, it is difficult to distinguish on pure experimental grounds the effect of CNO when both cell types express DREADDs, as these cells "antagonise" each other. Contrarily, numerical simulation can afford it. Here, we propose a novel simulation platform that can reflect normal and impaired retinal function (from single-cell to large-scale level). It is able to handle different visual processing circuits and allows us to visualise responses to visual scenes (movies). In addition, the platform allows simulation of retinal responses when DREADD-expressing cell subclasses are either silenced or excited with CNO. To demonstrate how our simulator works, we deploy a circuit that handles motion on a large-scale level and study how the retina responds to visual scenes by visualising retinal processing at each level. The simulator also provides a tunable parameter to control the CNO effect (excitation or inhibition). Consequently, it facilitates the disentanglement of the effect of CNO on ACs and RGCs. Nevertheless, simulations and experiments are widely complementary. Experiments are necessary to constrain the numerical model and check its validity (especially, its predictions), while the computational approach affords to explore aspects that cannot be easily achieved experimentally.

This work has been presented in [24], [19], [33]

7.2.2.2. Simulating the cortical activity evoked by artificial retinal implants

Participants: Teva Andréoletti, Bruno Cessac, Frédéric Chavane [Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France)], Sébastien Roux [Institut de Neurosciences de la Timone (CNRS and Aix-Marseille Université, France)].

Recent advances in neuroscience and microelectronics opens up the possibility of partially restoring vision to blind patients using retinal prostheses. These are devices capturing the light of a visual scene and converting it to electric impulses sent by a matrix of electrodes chirurgically fixed on the retina. The simulation of an electrode elicits an activation in the visual cortex that evokes a percept similar to a light spot called phosphene. The joint stimulation of electrodes allows to reproduce simple shapes (letters, objects, stairs) and to restore a low resolution vision to blind people (see Fig 1). This domain of research is however at an early stage compared to cochlear implants. Especially, the way an electric stimulation activates the visual cortex is still poorly understood. The group of F. Chavane (NeOpTo team at INT Marseille) has used mesoscopic recordings of cortical activity (optical imaging) to better understand the activity evoked by stimulation of the retina with implanted multi electrodes arrays (Roux et al 2016 eLife). Their results show that local stimulation of the retina evoked a cortical activity that is up to 10 times larger than what is expected based on the activity evoked by visual stimuli. This result is in line with known poor resolutions of percepts evoked by stimulation of artificial retinas implanted in blind patients. This observed spread of evoked cortical activity is now better understood. An important effect, evidenced by Roux et al (2017) https://elifesciences.org/articles/12687 is the asymmetrical spread of electric activity induced by the direct activation of retinal cells axons away from their somata.

This effect can be modelled at the level of a single electrode with a significant match to experimental measurement. Retinal prostheses integrate hundreds of electrodes and this model can be used to anticipate the simultaneous activation of several electrodes reproducing the shape of an object (Fig 1). This figure has been produced by a retina simulator, called Macular, developed by the Biovision team at Inria, and aiming at reproducing the retina response to stimulation in normal (stimulation by light) and pathological conditions (electric stimulation by prostheses) https://team.inria.fr/biovision/macular-software/ . In a previous work https://hal.inria.fr/hal-02292831 [28], [29] we have been able to numerically model the effect of the static joint stimulation of electrodes in retina prostheses on the primary visual cortex (V1) and to compare it to normal vision.

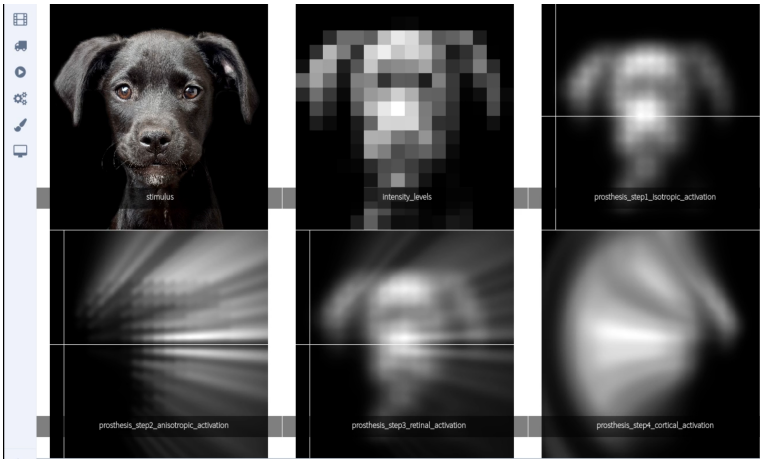


Figure 10. Simulation of the retinal and cortical response to a prosthesis simulation. An image (up-left) is digitalized into small squares (up-left). Each square corresponds to the degree of activation of a corresponding electrode in retinal implant (up-right). The electric stimulation activates neurones en passant of retinal cells leading to non linear diffusion (left bottom) and an effective stimulation pattern (down-middle) which is blurred in comparison to the expected stimulation pattern (up-right). The induced cortical representation is shown in the bottom-right figure.

This work has been presented in [28], [29]

7.3. Neuronal modelling

7.3.1. Linear response in neuronal networks: from neurons dynamics to collective responses

Participant: Bruno Cessac.

We have reviewed two examples where the linear response of a neuronal network submitted to an external stimulus can be derived explicitly, including network parameters dependence. This is done in a statistical physics-like approach where one associates to the spontaneous dynamics of the model a natural notion of Gibbs distribution inherited from ergodic theory or stochastic processes. These two examples are the Amari-Wilson-Cowan mode and a conductance based Integrate and Fire model

This work has been published in [13], [31], [18].

7.3.2. On the role of Nav1.7 sodium channels in chronic pain: an experimental and computational study

Participants: Lyle Armstrong [Institute of Neuroscience (ION), Newcastle, UK], Alberto Capurro [Institute of Neuroscience (ION), Newcastle, UK], Bruno Cessac, Jack Thornton [Institute of Neuroscience (ION), Newcastle, UK], Evelyne Sernagor [Institute of Neuroscience (ION), Newcastle, UK].

Chronic pain is a global healthcare problem with a huge societal impact. Its management remains generally unsatisfactory, with no single treatment clinically approved in most cases. In this study we use an in vitro model of erythromelalgia consisting of dorsal root ganglion neurons derived from human induced pluripotent stem cells obtained from a patient (carrying the mutation F1449V) and a control subject. We combine neurophysiology and computational modelling to focus on the Nav1.7 voltage gated sodium channel, which acts as an amplifier of the receptor potential in nociceptive neurons and plays a critical role in erythromelalgia due to gain of function mutations causing the channel to open with smaller depolarisations. Using extracellular recordings, we found that the scorpion toxin OD1 (a Nav1.7 channel opener) increases dorsal root ganglion cell excitability in cultures obtained from the control donor, evidenced by an increase in spontaneous discharges, firing rate and spike amplitude. In addition, we confirmed previous reports of voltage clamp experiments concerning an increase in spontaneous discharge in the patient cell cultures and the analgesic effects of the Nav1.7 blocker PF-05089771. Our findings are explained with a conductance-based model of the dorsal root ganglion neuron, exploring its behaviour for different values of half activation voltage and inactivation removal rate of the Nav1.7 current. Erythromelalgia was simulated through a decrease of the Nav1.7 half activation voltage, turning previously subthreshold stimuli to pain-inducing, and successfully counteracted with the channel blocker. The painful effects of OD1 were simulated through a quicker removal of Nav1.7 inactivation that reproduced the effects of the toxin not only on the spike frequency but also on its amplitude. This work has been submitted to J. Neuroscience. [30].

7.3.3. Ghost attractors in spontaneous brain activity: wandering in a repertoire of functionally relevant BOLD phaselocking solutions

Participants: Joana Cabral [Department of Psychiatry, Medical Sciences Division, University of Oxford,UK], Bruno Cessac, Gustavo Deco [Catalan Institute for Research and Advance Studies (ICREA), Spain], Morten L. Kringelbach [University of Oxford, UK], Jakube Vohryzek [Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Denmark].

Functionally relevant network patterns form transiently in brain activity during rest, where a given subset of brain areas exhibits temporally synchronized BOLD signals. To adequately assess the biophysical mechanisms governing intrinsic brain activity, a detailed characterization of the dynamical features of functional networks is needed from the experimental side to inform theoretical models. In this work, we use an open-source fMRI dataset from 100 unrelated participants from the Human Connectome Project and analyse whole-brain activity using Leading Eigenvector Dynamics Analysis, which focuses on the detection of recurrent phase-locking patterns in the BOLD signal. Borrowing tools from dynamical systems theory, we characterise spontaneous brain activity in the form of trajectories within a low-dimensional phase space. Decomposing the phase space into Voronoi-like cells using k-means clustering algorithm, we demonstrate that the cluster centroids (representing recurrent BOLD phase-locking patterns) closely overlap with previously identified resting-state networks. We further demonstrate that the metric associated with the phase-locking patterns shows moderate reliability across recordings indicating potential existence of subject specific dynamical landscapes. Our results point to the hypothesis that functional brain networks behave as ghost attractor states in a low-dimensional phase space, providing insights into the evolutionary rules governing brain activity in the spontaneous state and reinforcing the importance of addressing brain function within the framework of dynamical systems theory. This work has been submitted to Frontiers in Systems Neuroscience.

CAMIN Project-Team

6. New Results

6.1. Selectivity of implanted neural electrical stimulation

Participants: Lucie William, David Guiraud, Charles Fattal, Christine Azevedo, Arthur Hiairrassary.

In the context of using a multi-contact cuff electrode positioned around a trunk nerve to activate selectively the fascicles leading to selective movements, a pre-clinical study was performed on the sciatic nerve of four rabbits (Lab. Chirurgie Experimentale, Institut de Biologie, University of Montpellier). The purpose was to compare and classify six different currents configuration (current ratios) (Fig.8) with a 12- contact cuff electrode using selectivity, robustness (i.e. ability to maintain selectivity within a range of current amplitudes) and efficiency (i.e. electrical consumption of the considered multipolar configuration *versus* the electrical consumption of the reference whole-ring configuration) indexes.

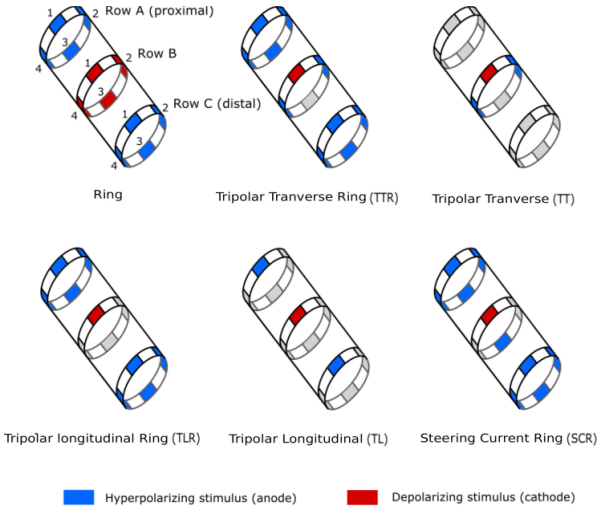


Figure 8. Six different configurations of the 12-contact electrode were tested: Ring, Tripolar Transverse Ring (TTR), Tripolar Transverse (TT), Tripolar Longitudinal Ring (TLR), Tripolar Longitudinal (TL), Steering Current Ring (SCR)

Results indicated that the optimal configuration depends on the weights applied to selectivity robustness and efficiency criteria. Tripolar transverse is the most robust configuration and the less efficient, whereas tripolar longitudinal ring is efficient but not robust. New configurations issued from a previous theoretical study we carried out such as steering current ring appears as good compromise between the 3 criteria [18].
The PhD of Lucie William (started in October 2019) will be the continuation of this work in the context of neural electrical stimulation of complete quadriplegic human participants (AGILIS project).

6.2. Selective Neural Electrical Stimulation to restores Hand and Forearm Movements in Individuals with Complete Tetraplegia

Participants: David Guiraud, Charles Fattal, Christine Azevedo, Mélissa Dali, Jacques Teissier [Beau Soleil clinic, Montpellier], Anthony Gélys [Propara Rehab. Center, Montpellier].

Selective neural electrical stimulation of radial and median nerves enables the activation of functional movements in the paralyzed hand of individuals with tetraplegia. In eight participants (Clinique Beau Soleil and Propara Rehabilitation Center, Montpellier) with complete tetraplegia, during a programmed surgery and under complete anesthesia, we demonstrated that selective stimulation based on multicontact cuff electrodes and optimized current spreading over the active contacts provided isolated, compound, functional and strong movements. Several configurations were needed to target different areas within the nerve to obtain all the envisioned movements. We further confirmed that the upper limb nerves have muscle specific fascicles, which makes possible to activate isolated movements. The future goal is to provide patients with functional restoration of object grasping and releasing with a minimally invasive solution: only two cuff electrodes above the elbow. This will be the objective of AGILIS project supported by EIT Health.

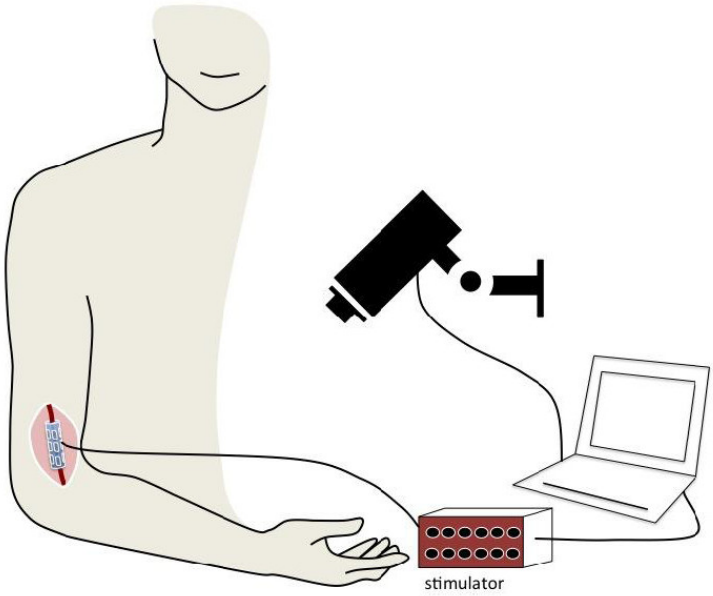


Figure 9. Neural electrical stimulation of radial or median nerve using a multi-contact cuff electrode allows to elicit different individual or grouped muscle contractions inducing different fingers and wrist movements.

6.3. Assisted Grasping in Individuals with Tetraplegia: Improving Control through Residual Muscle Contraction and Movement

Participants: Lucas Fonseca [UnB, Brazil], David Guiraud, Charles Fattal, Christine Azevedo, Arthur Hiairrassary, Camilo Silva, Anthony Gélys [Propara Rehab. Center, Montpellier].

Individuals who sustained a spinal cord injury often lose important motor skills, and cannot perform basic daily living activities. Several assistive technologies, including robotic assistance and functional electrical stimulation, have been developed to restore lost functions. However, designing reliable interfaces to control assistive devices for individuals with C4–C8 complete tetraplegia remains challenging. Although with limited grasping ability, they can often control upper arm movements via residual muscle contraction. We have explored the feasibility of drawing upon these residual functions to pilot two devices, a robotic hand and an electrical stimulator. We studied two modalities, supra-lesional electromyography (EMG), and upper arm inertial sensors (IMU). We interpreted the muscle activity or arm movements of subjects with tetraplegia

attempting to control the opening/closing of a robotic hand, and the extension/flexion of their own contralateral hand muscles activated by electrical stimulation. Two groups of participants with quadriplegia were recruited (Clinique Propara, Montpellier): eight subjects issued EMG-based commands; nine other subjects issued IMU-based commands. For each participant, we selected at least two muscles or gestures detectable by our algorithms. Despite little training, all participants could control the robot's gestures or electrical stimulation of their own arm via muscle contraction or limb motion [20].

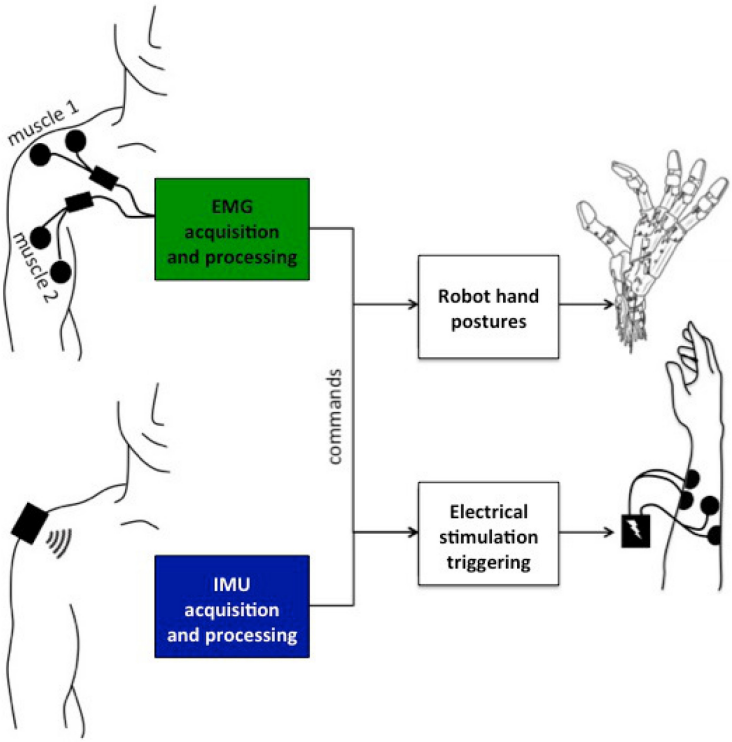


Figure 10. Protocol principle. EMG or IMU signals are converted into commands for the robotic hand or the electrical stimulator. The robotic hand has three possible gestures: at-rest, open and close. The electrical stimulator can receive three commands: no stimulation, stimulate channel 1 (wrist flexion) or stimulate channel 2 (wrist extension). Users are able to observe the outcome of their input and use it as biofeedback.

In the AGILIS project supported by EIT Health, we intend to extend this approach to participants with 2 implanted electrodes on median and radial nerves participating in a 30-days clinical study (APHP, Paris and Clinique La Châtaigneraie, Menucourt).

We are currently working on the software that is responsible for acquiring sensor data and controlling the stimulator (§5.2.4). The previous algorithms are being implemented in a single platform focusing on the 30-days clinical study. The residual motion based system was improved based on the results published in [20]. It is also faster and more efficient. The inertial sensors now have higher frequency, which leads to higher accuracy of movement classification, particularly with faster movements.

6.4. Modeling and simulation of a human hand

Participants: Daniel Simon, Ahmed Farek.

The AGILIS stimulation system is intended to generate grasping action on some objects such as balls and cans. A high-fidelity hand model and associated simulation software was developed to anticipate real experiments and help for the system identification and tuning [30]. The hand model uses 23 degrees of freedom for the wrist and fingers. 28 muscles are considered, including the 12 muscles which are expected to be activated using electrical stimulation of the median and radial nerves. Others are also considered in the model as, even if not stimulated, they contribute to the hand and fingers movements through passive forces when extended. Several actuation models are investigated to allow for the identification of muscles-to-movements relations.

The active forces provided by the stimulated muscles are computed thanks to the original model developed over the past years by the team, where the inputs are currents injected to muscles or nerves. The fingers are assumed to interact with the grasped object through elastic contacts and limited friction.

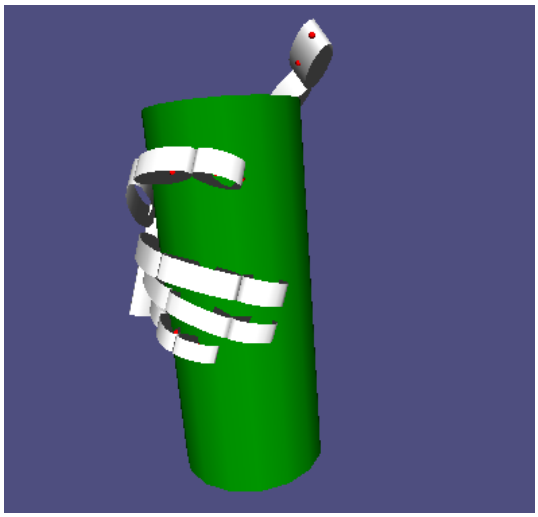


Figure 11. Simulation of a stimulated hand grasping a can

6.5. Functional impact of a self-triggered grasping neuroprosthesis in post-stroke subjects

Participants: David Gasq, Christine Azevedo, Ronan Le Guillou, Jérôme Froger [CHU Nîmes, France].

The improvement of the grasp abilities remains a challenge in the 50% of post-stroke subjects who have not recovered functional grasping due to paralysis of the finger's extensor muscles. The ePrehension-Stroke is a prospective, bicentric (promoted by the CHU de Nîmes), multi-crossover, blinded evaluation study which assesses the functional impact of a self-triggered grasping neuroprosthesis. We have developed a specific software, NeuroPrehens, which controls external electrical stimulations applied over finger's extensor muscles and was triggered by voluntary head movements or electromyography activity of leg muscles. The main objective is to assess the impact of the self-triggered grasping neuroprosthesis on the ability to perform a standardized task of grasping, moving and releasing either a glass (palmar grasp) or a spoon (key pinch), compared to the absence of neuroprosthesis use. Secondary objectives are to assess (1) the preferential modes of neuroprosthesis control, (2) the impact of the neuroprosthesis on a standardized unimanual grip scale (Action Arm Research Test), (3) the psycho-social impacts (Psychosocial Impact of Assistive Devices Scale questionnaire) and the subject's satisfaction and tolerance (Quebec User Assessment of Satisfaction with Assistive Technology questionnaire) related to neuroprosthesis use. Over 20 subjects planned to include until

June 2020, we have included 8 subjects since July 2019. The prospects of this pilot study are to develop a fully wearable and self-piloted neuroprosthesis that can be used in daily life by the largest number of post-stroke subjects who have not recovered active grasping abilities.

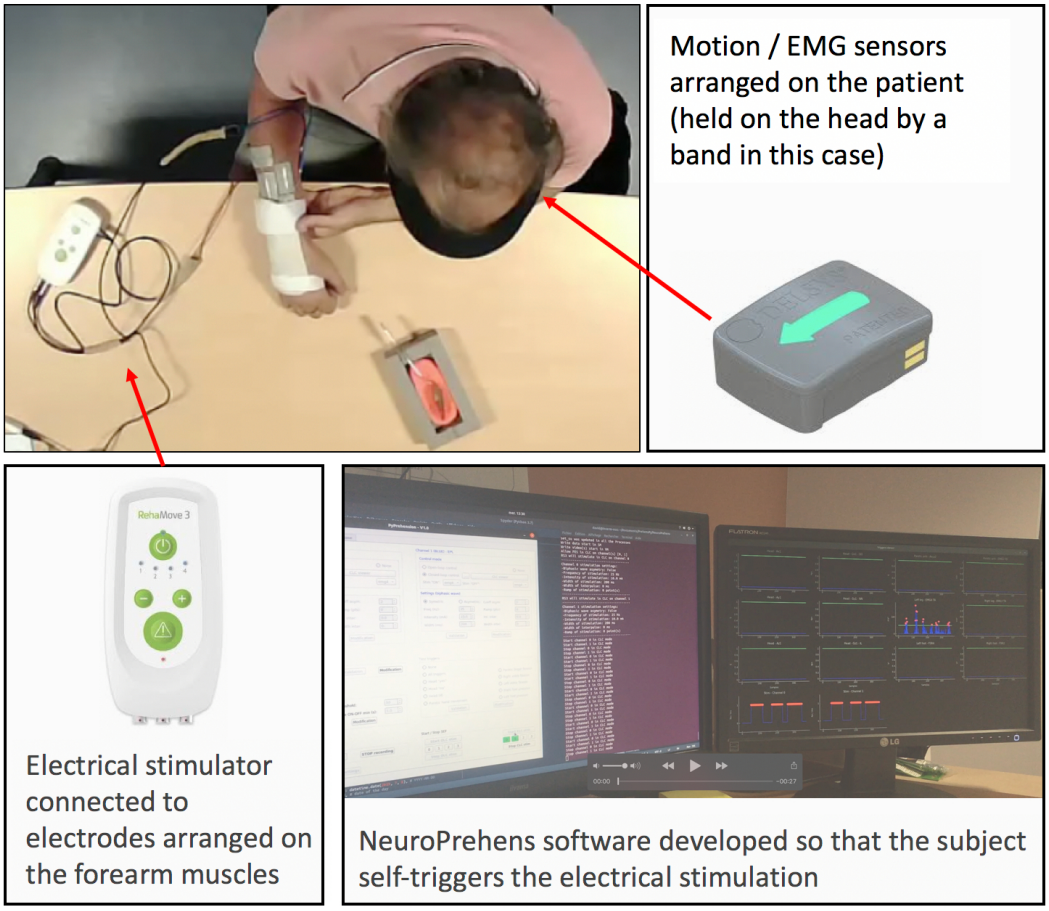


Figure 12. Experimental device constituting the self-triggered grasping neuroprosthesis.

6.6. Near-infrared spectroscopy time course under hypercapnia

Participants: Victor Vagné, David Guiraud, Vincent Costalat [CHU Montpellier], Emmanuelle Le-Bars, Stephane Perrey.

Partial arterial pressure of carbon dioxide (CO2) modulates cerebral blood flow through vasoreactivity mechanism. Near infrared spectroscopy (NIRS) can be used to record these changes in cerebral hemodynamics. However, no laterality comparison of the NIRS signal has been performed despite being a prerequisite for the use of such method in a vasoreactivity monitoring context. We propose to investigate laterality of NIRS signal in response to a CO2-inhalation-based hypercapnia paradigm in healthy volunteers.

Methods: Eleven healthy volunteers (6 women, 5 men, mean age: 31 $\pm$ 11) underwent a 3-block-design inhalation paradigm: normoxia (5min, “baseline”) – hypercapnia (2min, “stimulation”) – normoxia (5min, “post-stimulation”). NIRS signal was measured using a two-channel oximeter (INVOS 5100C, Medtronic,

USA) with sensors placed symmetrically on both the left and right sides on each subject's forehead. Additional heart rate (HR) monitoring was performed simultaneously. Based on the NIRS mean signal pattern, an a priori model of parametric identification was applied for each channel to quantify parameters of interest (amplitude, time delay, excitation and relaxation time) for each inhalation block.

Results: HR increased significantly during the stimulation block. The quality of the model was satisfactory: mean absolute error between modeled and experimental signals were lower than the resolution of the device. No significant lateralization were found between left and right values of most of the parameters.

Conclusion: Due to the lack of lateralization, this parametric identification of NIRS responses to hypercapnia could bring light to a potential asymmetry and be used as a biomarker in patients with cerebrovascular diseases.

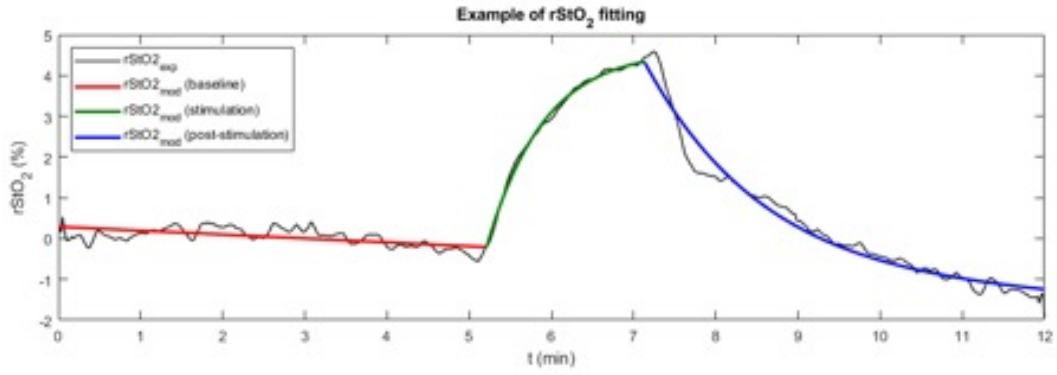


Figure 13. Example of curve fitting on NIRS signal with a compartmental first order model in response to a 2-min hypercapnic stimulus.

6.7. EPIONE

Participants: David Andreu, David Guiraud, Arthur Hiairrassary.

The project was completed in 2017 but major publications were issued in 2018 and 2019 reporting the most important results of both stimulation of the upper and lower limbs in amputees to restore sensations using 4 TIME electrodes. We developed original algorithms that convert signals acquired from sensors of the artificial lower limb, namely the prosthetic limb, into stimulus to the afferent branches of the sciatic nerve. This pioneering work shows that not only the gait performances were greatly enhanced but also the phantom pain relief was effective with a long lasting after stopping the therapy [22] [21]. These results follows the previous ones pbtained on the upper limb with similar results [23].

6.8. FES-assisted cycling

Participants: Benoît Sijobert, Ronan Le Guillou, Charles Fattal, Christine Azevedo, Martin Schmoll, Emerson Fachin-Martins [UnB, Brazil], Henrique Resende [UFMG, Brazil], David Lobato [UnB, Brazil].

Our team is working for several years on FES-assisted cycling for individuals with spinal cord injury. We intend to improve cycling accessibility to a larger population in order to propose exercising and leisure activity to improve quality of life and self esteem. On this context we have been working on three aspects this year: 1) improving training to be able to propose patients in rehabilitation centers a simplified and acceptable protocol to prepare muscles to cycling, 2) improving usability in a rehabilitation context to ease and simplify the access to FES-cycling, 3) better understanding fatigue phenomena to improve cycling performances.

A funding (EDF Foundation) has been obtained by our clinical partner "CRF La Châtaigneraie" to perform a clinical protocol to follow up the physical preparation of individuals with spinal cord injury to manage overground active pedaling after 4 months of 3 sessions per week training at home. The protocol has been approved by an ethical committee (§3.3). One of the participants will be involved in Cybathlon 2020 event. The inclusions began in September 2019. A longitudinal follow-up will allow to precisely assess the performances progress along the training period. After the 4-months at home training the participants will be using the overground cycling platform that has been developed by our team (§5.2.2).

Figure 14. Muscular preparation for overground cycling training. Left: Participant executing strengthening program with conventional multichannel stimulator (CEFAR); Right: Participant performing endurance training on FES-ergocycle.

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. Our latest study on this subject introduces a novel approach of a Functional Electrical Stimulation (FES) controller intended for FES-induced cycling based on inertial measurement units (IMUs). This study aimed at simplifying the design of electrical stimulation timing patterns while providing a method adapted to different users and devices. In most of the different studies and commercial devices, the crank angle is used as an input to trigger stimulation onset. We propose to use instead thigh inclination as the reference information to build stimulation timing patterns. The tilting angles of both thighs are estimated from one inertial sensor located above each of the knees. An IF-THEN rules algorithm detects online and automatically the thigh peak angles in order to start and stop the stimulation of quadriceps muscles depending on these events. One participant with complete paraplegia was included and was able to propel a recumbent trike using the proposed approach after a very short setting time. This new modality opens the way to a

simpler and user-friendly method to automatically design FES-induced cycling stimulation patterns, adapted to a clinical use, to multiple bike geometries and user morphologies. Using the online peak knee flexion algorithm developed in the study presented in last years section 6.2 to continuously detect this event, we validated a novel approach in order to trigger the quadriceps stimulation at the beginning of the pushing phase. These results can be seen in Figure 15 . Enabling this method 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 [24].

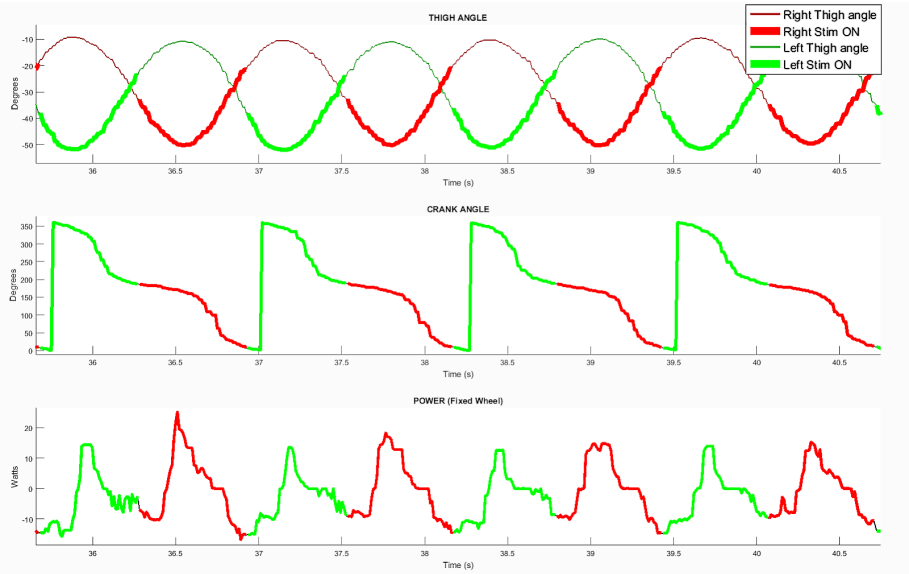


Figure 15. Data sample illustrating the results over four pedalling cycles in home trainer condition. TOP: Left (green) and right (red) thigh tilting angles - MIDDLE: crank angle - BOTTOM: developed power. The two stimulation channels activation are highlighted.

Another important aspect concerning FES-cycling is the pilots ability to resist fatigue for a prolonged time. Muscular activation as a result of electrical nerve stimulation is known to introduce a rather quick onset of fatigue. Therefore different approaches have been tested in literature to reduce the effective stimulation frequency received by individual motor-units. Several studies were able to show improvements using distributed multichannel stimulation against conventional single channel stimulation. A direct comparison between the different techniques is difficult as all studies use different methods of quantifying muscular fatigue. Further most studies fail to mention measured absolute values during a contraction at maximum strength. Therefore our team was designing a new testing protocol in collaboration with the University of Brasilia (CACAO Associate team) with the aim to assess muscular fatigue of currently published and new electrode positions against conventional single channel stimulation (baseline) in a more practical setting. The fatigue testing protocol was tailored to mimic 10 min FES-cycling at 50 RPM using an isokinetic dynamometer (Biodex System 4). Assuming a torque-production of 40 percent of the maximal torque-production-capacity of a well-trained quadriceps muscle to be sufficient for FES-cycling. The active torque produced at this starting level was measured in a series of contractions, tracking the decline of torque. The study was conducted in Brasilia on 3 individuals expressing a complete spinal cord injury. All participants were enrolled in a FES-training program for about 14months. All participants were highly motivated and fulfilled the demanded inclusion criteria ensuring a safe execution of the protocol. For every subject both legs were measured individually leading to an overall sample size of n=6. Currently the data-analysis is still in progress. The results of this study should lead to optimized stimulation techniques to prolong the onset of fatigue during FES-cycling.

Figure 16. Overground cycling. The participant was able to propel the recumbent trike over a 40 meter corridor.

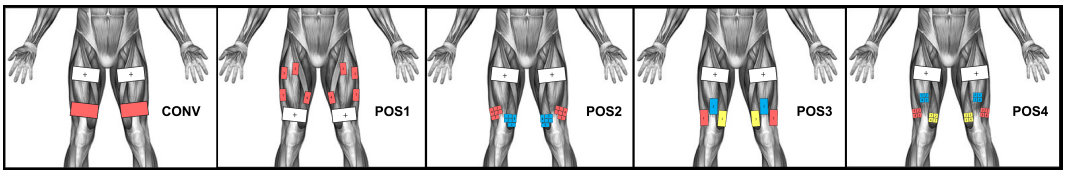


Figure 17. Electrode configurations examined. Electrodes marked with "+" were the Anodes (reference electrodes). CONV: Standard electrode configuration 40 Hz delivered to one pair of electrodes; POS1: One channel with 40 Hz distributed over 4 electrodes – common anode; POS2: Two channels with each 40 Hz distributed via 4 electrodes – common anode; POS3: Three channels of 40 Hz each delivered to one electrode – common anode: POS4: Three channels with each 40 Hz distributed via 4 electrodes – common anode

6.9. Breathing detection via tracheal sounds

Participants: Xinyue Lu, David Guiraud, Christine Azevedo, Serge Renaux [Neuroresp], Thomas Similowski [Hosp. LA Salpêtrière, Paris].

Individuals with a respiratory paralysis are essentially supplied by mechanical ventilation. However, severe drawbacks of mechanical ventilation were reported: low autonomy, high health costs, infection risk, etc. If patients' phrenic nerves and diaphragms are still functional, implanted diaphragm pacing can provide them a more natural respiration. Compared to classic mechanical ventilation, implanted diaphragm pacing can cancel some of the disadvantages mentioned above, and can also help to significantly improve speech and recover some olfactory sensation.

But existing implanted diaphragm pacing systems can not monitor patient's induced respiration and they stimulate at constant intensity and frequency - they work in open-loop. It means that stimulation intensity,

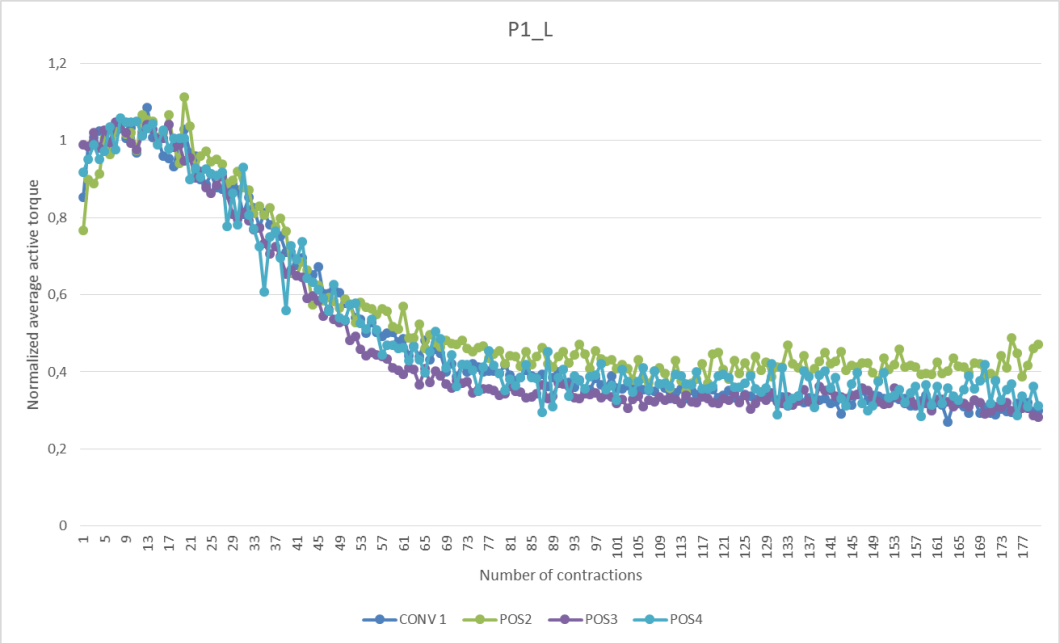


Figure 18. Preliminary data retrieved of the left leg of patient 1. The average active torque of the first 20 contractions was used to normalize fatigue curves. POS1 was excluded from the figure due to exaggerated fluctuations in torque.

pulse width and frequency are fixed at the installation of the implant, updated at each control visit, but do not adapt to patient's continuous situation evolution because of the absence of respiratory monitoring. To close the loop, an ambulatory respiratory monitoring solution needs to be developed. Adding adaptive abilities to existing systems would improve the efficiency of the delivered stimulation.

The gold standard for apnea/hypoventilation evaluation is the polygraph, which includes an pulse oximeter and at least one respiratory flow sensor. In a clinical use, flow sensors could be nasal cannula, pneumotachograph, thermistor or plethysmograph. But these sensors need to be placed over the face or are sensitive to patient's movements. They are therefore not compatible with an implanted diaphragm pacing system which is portable and for a daily living use. With this in mind, this study investigated an acoustic method. The proposed tracheal sounds recording requires only one tiny microphone fixed on the neck with a support, which is the only physical contact with the patient.

Figure 19. The position of microphone to record tracheal sounds.

Many previous studies have shown some positive results on respiration analysis from tracheal sounds in sleep apnea, especially for obstructive sleep apnea. But only few methods are developed for real-time applications (processing delay within seconds) with robustness requirements, indeed, all these studies have been carried out in quiet and controlled acoustic environments with stable sources of noises, and with limited movements of the subjects (during sleep).

In collaboration with NEURORESP company and La Salpêtrière Gospital (Paris) we are investigating the possibility to perform a real-time and continuous breathing detection (day and night), even during wakefulness in noisy environments. We proposed thea method with tracheal sounds recorded on the neck at suprasternal notch (Fig.19). This method is noninvasive and easy to apply. And the recorded tracheal sounds contain not only respiratory sounds, but also heart beats sounds (as phonocardiogram: PCG) so that some basic cardiac information, as cardiac rhythm, could be calculated. Furthermore, inspired by ECG-derived respiration, the similar method could also be applied on obtained PCG to get respiratory information.

The proposed method has been tested on 30 recordings from 15 healthy subjects with different respiratory condition, one example is shown in Fig.20 . We proposed a new algorithm to detect respiration phases, by combining the signal processing both in the temporal (envelope and PCG-derived respiration) and the frequency domains. We assessed the performances of the algorithm in emulated noisy environments. The accuracy, sensibility and specificity of system are all superior to 90%. The result is good enough to show a proof of such a conception. Furthermore, a tracheal sounds recording from a patient under implanted phrenic nerve stimulation has shown that the recording system has the possibility to capt an image of stimulation impulse from the wireless transmission. Getting the synchronization with respiratory sounds and stimulation signals can help to verify and even to adjust patient's stimulation parameters.

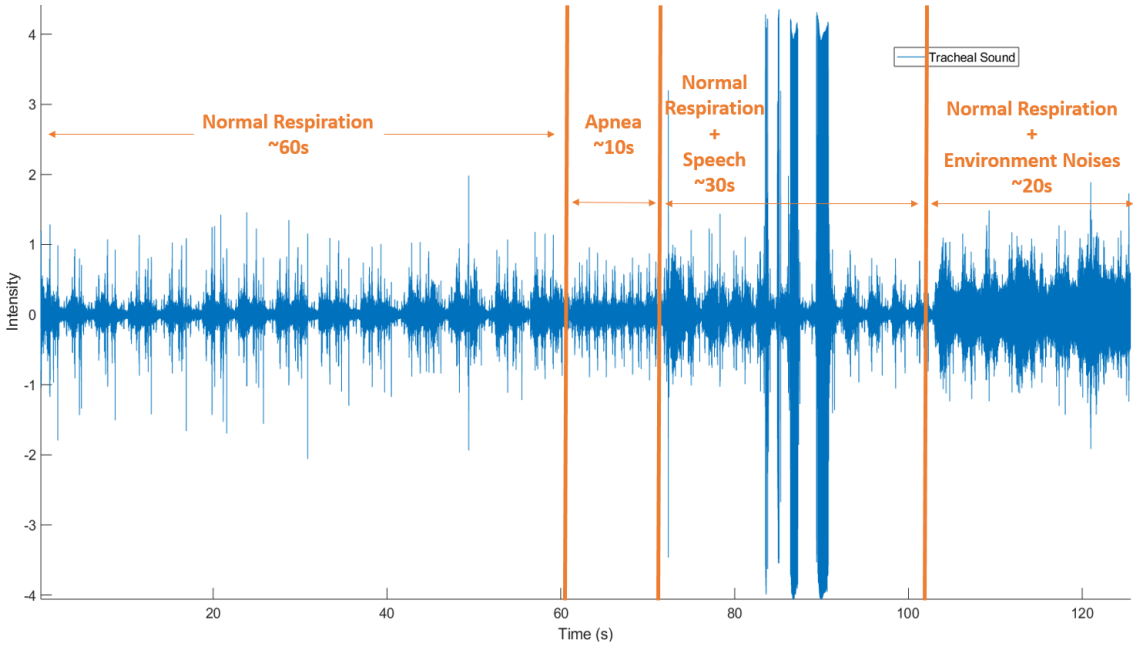


Figure 20. One example of a 2-min recording of tracheal sounds.

6.10. Attenuation and Delay of Remote Potentials Evoked by Direct Electrical Stimulation During Brain Surgery

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

Direct electrical stimulation (DES) is used to perform functional brain mapping during awake surgery but its electrophysiological effects 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. EPs were similar in shape but delayed in time and attenuated in amplitude when elicited from a different gyrus or remotely from the recording site. We were able to trigger remote EPs using low stimulation intensities. We propose different activation and electrophysiological propagation mechanisms following DES based on activated neural elements [15].

Electrodes of both ECoG strips are numbered from 1 to 4 and from 5 to 8. The Sylvian fissure and central sulcus are highlighted by a white dashed lines and annoted "SyF" and "Cs" respectively. For P1, 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. Strip 1 spans over both temporal and parietal lobe with: electrode 1 over the most posterior part of the superior temporal gyrus; electrode 2 over the Sylvian fissure; electrodes 3 and 4 over the adjacent supramarginal gyrus. Strip 2 spans over the precentral gyrus with: electrodes 5 to 7 over the ventral premotor cortex; electrode 8 is bordering with the most posterior part of the partially resected dorsolateral prefrontal cortex. For P2, experimental DES was applied on: (1) the middle part of the superior temporal gyrus (S1) which led to complete anomia; (2) the precentral gyrus (S2), which induced articulatory disorders. Strip 1 spans over the superior temporal gyrus with: electrodes 1 and 2 over its middle third; electrodes 3 and 4 over its most posterior part. Strip 2 spans over the precentral

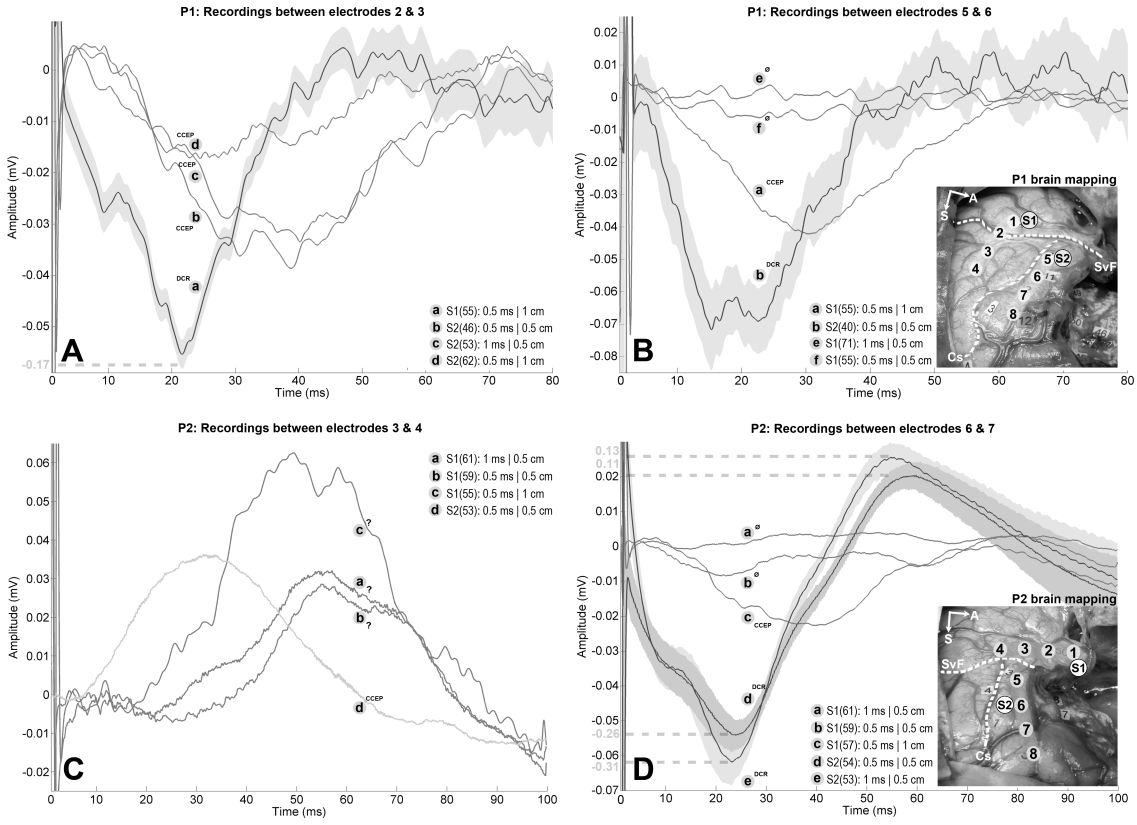


Figure 21. P1 and P2 brain mappings: Pictures illustrating the stimulation sites (S1, S2) and ECoG positioning with respect to the initial 60 Hz cortical brain mapping (numbered paper tags).

and dorsolateral prefrontal gyri with: electrodes 5 and 6 over the ventral premotor cortex; electrodes 7 and 8 are respectively bordering and within the adjacent dorsolateral prefrontal cortex. Tumor was about 164 cm3 for P1 and 150 cm3 for P2. The number of averaged stimuli is reported within parentheses for each trace. 99% confidence interval estimated for DCRs are represented by grey surfaces to demonstrate that CCEPs do not belong to them. Additional traces corresponding to variations of stimulation parameters were added if available, regardless of the presence of EPs. A: Differential recordings between electrodes 2 and 3 for P1 while stimulating S1 (-170 μV, 21 ms delay) and S2 (amplitudes ranging from -40 μV to -17 μV, delays ranging from 25 ms to 38 ms). EPs following S2 stimulation are CCEPs because of the presence of the central fissure between the stimulation and recording sites. The EP measured after stimulating S1 is ambiguous because electrode 2 lies on the Sylvian fissure, but the short latency and enhanced amplitude with regard to the CCEPs suggest a DCR. Note the dashed line indicating a different amplitude scale for the DCR, which was reduced by a factor 3 for visualization purposes. B: Differential recordings between electrodes 5 and 6 for P1 while stimulating S2 (-75 μV, 20 ms delay) and S1 (-44 μV, 30 ms delay). EP following S1 stimulation is a CCEP because of the presence of the Sylvian fissure between the stimulation and recording sites. EP following S2 stimulation should be viewed as DCR as it was recorded on the same gyrus and it showed shorter latency and enhanced amplitude in comparison with the CCEP. C: Differential recordings between electrodes 3 and 4 for P2 while stimulating S1 (amplitudes ranging from +29 μV to +62 μV, delays ranging from 52 ms to 62 ms) and S2 (+36 μV, 32 ms delay). EP following S2 stimulation is a CCEP because of the presence of the Sylvian fissure between the stimulation and recording sites. EPs following S1 stimulations should be viewed as DCRs as they are recorded on the same gyrus but the latencies and amplitudes appeared unusual. EPs are positive because of differential measure. D: Differential recordings between electrodes 6 and 7 for P2 while stimulating S2 (amplitudes ranging from -260 μV to -310 μV, 20 ms delay) and S1 (-24 μV, 38 ms delay). EP following S1 stimulation is a CCEP because of the presence of the Sylvian fissure and the operative cavity between the stimulation and recording sites. EPs following S2 are likely DCRs as they are recorded on the same gyrus, which is corroborated by their short latencies and maximized amplitudes with regard to the CCEP. Note the dashed lines indicating different amplitude scales for the DCRs, which were reduced by a factor 5 for visualization purposes.

EMPENN 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 following major methodological topics:

7.1.1. Diffusion imaging

7.1.1.1. Free water estimation using single-shell diffusion-weighted images

Participant: Emmanuel Caruyer.

Free-water estimation requires the fitting of a bi-compartment model, which is an ill-posed problem when using only single-shell data. Its solution requires optimization, which relies on an initialization step. We propose a novel initialization approach, called "Freewater EstimatoR using iNtErpolated iniTialization" (FERNET), which improves the estimation of free water in edematous and infiltrated peritumoral regions, using single-shell diffusion MRI data. The method has been extensively investigated on simulated data and healthy and brain tumor datasets, demonstrating its applicability on clinically acquired data. Additionally, it has been applied to data from brain tumor patients to demonstrate the improvement in tractography in the peritumoral region [57].

7.1.1.2. Multi-dimensional diffusion MRI sampling scheme: B-tensor design and accurate signal reconstruction

Participant: Emmanuel Caruyer.

B-tensor encoding enables the separation of isotropic and anisotropic tensors. However, little consideration has been given as to how to design a B-tensor encoding sampling scheme. In this work, we propose the first 4D basis for representing the diffusion signal acquired with B-tensor encoding. We study the properties of the diffusion signal in this basis to give recommendations for optimally sampling the space of axisymmetric b-tensors. We show, using simulations, that the proposed sampling scheme enables accurate reconstruction of the diffusion signal by expansion in this basis using a clinically feasible number of samples [24].

This work was done in collaboration with A. Bates, Australian National University and Al. Daducci, University of Verona.

7.1.1.3. Optimal selection of diffusion-weighting gradient waveforms using compressed sensing and dictionary learning

Participants: Raphaël Truffet, Emmanuel Caruyer, Christian Barillot.

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 propose here a method to select only a subset of the measurements, while being able to predict the unseen data using compressed sensing. We learn 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 [49].

7.1.1.4. Geometric evaluation of distortion correction methods in diffusion MRI of the spinal cord

Participants: Haykel Snoussi, Emmanuel Caruyer, Olivier Commowick, Benoit Combès, Élise Bannier, Christian Barillot.

Acquiring and processing Diffusion MRI in spinal cord present inherent challenges. Differences in magnetic susceptibility between soft tissues, air and bones make the magnetic field non uniform in spinal cord. In this context, various procedures were proposed for correcting inhomogeneity-induced distortions; in this work, we propose novel geometric statistics to measure the alignment of the reconstructed diffusion model with the apparent centerline of the spine. In parallel of the correlation with an anatomical T2-weighted image, we show the utility of these statistics to study and evaluate the impact of distortion correction by comparing three correction methods using a pair of images acquired with reversed gradient polarity [48].

This work was done in collaboration with Anne Kerbrat, Neuropoly Montréal and Julien Cohen-Adad from NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, QC, Canada.

7.1.2. Arterial Spin Labeling

7.1.2.1. Acquisition duration in resting-state arterial spin labeling. How long is enough?
Participants: Corentin Vallée, Pierre Maurel, Isabelle Corouge, Christian Barillot.

Resting-state Arterial Spin Labeling (rs-ASL) is a rather confidential method compared to resting-state BOLD but it drives great prospects with respect to potential clinical applications. By enabling the study of cerebral blood flow maps, rs-ASL can lead to significant clinical subject-scaled applications as CBF is a biomarker in neuropathology. An important parameter to consider in functional imaging is the acquisition duration. Despite directly impacting practicability and functional networks representation, there is no standard for rs-ASL. Our work here focuses on strengthening the confidence in ASL as a rs-fMRI method, and on studying the influence of the acquisition duration. To this end, we acquired a long rs-ASL sequence and assessed the quality of typical functional brain networks quality over time compared to gold-standard networks. Our results show that after 14min of duration acquisition, functional networks representation can be considered as stable [58], [50].

7.1.2.2. 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 resolution 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 [16].

The development of this super-resolution algorithm in the context of the PhD of Cédric Meurée founded 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 him to integrate the proposed solution with tools in use within this team. Part of the work also consisted in reducing the computation, a factor of 5 being achieved at the end of these four weeks.

7.1.3. Atlases

7.1.3.1. Unbiased longitudinal brain atlas creation using robust linear registration and log-Euclidean framework for diffeomorphisms
Participants: Antoine Legouhy, Olivier Commowick, Christian Barillot.

We have defined a new method to create a diffeomorphic longitudinal (4D) atlas composed of a set of 3D atlases each representing an average model at a given age. This is achieved by generalizing atlasing methods to produce atlases unbiased with respect to the initial reference up to a rigid transformation and ensuring diffeomorphic deformations thanks to the Baker-Campbell-Hausdorff formula and the log-Euclidean framework for diffeomorphisms. Subjects are additionally weighted using an asymmetric function to closely match specified target ages. Creating a longitudinal atlas also implies dealing with subjects with large brain differences that can lead to registration errors. This is overcome by a robust rigid registration based on polar decomposition. We illustrated these techniques for the creation of a 4D pediatric atlas, showing their ability to create a temporally consistent atlas [22].

This work was done in collaboration with François Rousseau, IMT Atlantique, LaTIM U1101 INSERM, Brest, France, under the ANR MAIA project.

7.1.3.2. Online atlasing using an iterative centroid

Participants: Antoine Legouhy, Olivier Commowick, Christian Barillot.

Online atlasing, i.e. incrementing an atlas with new images as they are acquired, is key when performing studies on databases very large or still being gathered. We proposed to this end a new diffeomorphic online atlasing method without having to perform again the atlasing process from scratch. New subjects are integrated following an iterative procedure gradually shifting the centroid of the images to its final position, making it computationally cheap to update regularly an atlas as new images are acquired (only needing a number of registrations equal to the number of new subjects). We evaluated this iterative centroid approach through the analysis of the sharpness and variance of the resulting atlases, and the transformations of images, comparing their deviations from a conventional method using Guimond's method. We demonstrated that the transformations divergence between the two approaches is small and stable, and that both atlases reach equivalent levels of image quality [42].

This work was done in collaboration with François Rousseau, IMT Atlantique, LaTIM U1101 INSERM, Brest, France, under the ANR MAIA project.

7.1.4. Neurofeedback

7.1.4.1. Learning bi-modal EEG-fMRI neurofeedback to improve neurofeedback in EEG only

Participants: Claire Cury, Pierre Maurel, Giulia Lioi, Christian Barillot.

In neurofeedback (NF), a new kind of data are available: electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) acquired simultaneously during bi-modal EEG-fMRI neurofeedback. These two complementary techniques have only recently been integrated in the context of NF for brain rehabilitation protocols. Bi-modal NF (NF-EEG-fMRI) combines information coming from two modalities sensitive to different aspect of brain activity, therefore providing a higher NF quality. However, the use of the MRI scanner is cumbersome and exhausting for patients. We presented, a novel methodological development, able to reduce the use of fMRI while providing to subjects NF-EEG sessions of quality comparable to the bi-modal NF sessions. We proposed an original alternative to the ill-posed problem of source reconstruction. We designed a non-linear model considering different frequency bands, electrodes and temporal delays, with a structured sparse regularisation. Results show that our model is able to significantly improve the quality of NF sessions over what EEG could provide alone. We tested our method on 17 subjects that performed three NF-EEG-fMRI sessions each [30].

7.1.4.2. Can we learn from coupling EEG-fMRI to enhance neuro-feedback in EEG only?

Participants: Claire Cury, Pierre Maurel, Christian Barillot.

Neurofeedback (NF) measures brain activation during a task, and gives back to the subject a score reflecting his/her performance that he/she tries to improve. Among noninvasive functional brain imaging modalities, the most used in NF, are electro-encephalography (EEG) and the functional magnetic resonance imaging (fMRI). EEG measures the electrical activity of the brain through channels located on the scalp, with an excellent temporal resolution (milliseconds), but has a limited spatial resolution due to the well-known ill-posed inverse problem of source reconstruction. Also NF-EEG (NF session with NF scores extracted from EEG) is not easy to control since it comes from mixtures of propagating electric potential fluctuations. Blood oxygenation level dependent (BOLD) fMRI measures neuro-vascular activity, easier to control, with an excellent spatial resolution, making NF-fMRI (NF session with NF scores extracted from BOLD-fMRI) an adequate modality for NF. However its temporal resolution is only of a few seconds, and it is a costly, exhausting for subjects and time consuming modality. Since those modalities are complementary, their combined acquisition is actively investigated, as well as the methodology to extract information from fMRI with EEG which is the easiest modality to use [Abreu et al. 2018]. Our challenge is to learn EEG activation patterns from NF-fMRI scores extracted during a NF session using coupled EEG-fMRI data (NF-EEG-fMRI) to improve NF scores when using EEG only [29].

7.1.5. Deep learning

7.1.5.1. Unsupervised domain adaptation with optimal transport in multi-site segmentation of multiple sclerosis lesions from MRI data

Participants: Antoine Ackaouy, Olivier Commowick, Christian Barillot, Francesca Galassi.

Automatic segmentation of Multiple Sclerosis (MS) lesions from Magnetic Resonance Imaging (MRI) images is essential for clinical assessment and treatment planning of MS. Recent years have seen an increasing use of Convolutional Neural Networks (CNNs) for this task. Although these methods provide accurate segmentation, their applicability in clinical settings remains limited due to a reproducibility issue across different image domains. MS images can have highly variable characteristics across patients, MRI scanners and imaging protocols. Retraining a supervised model with data from each new domain is not a feasible solution because it requires manual annotation from expert radiologists. In this work, we explored an unsupervised solution to the problem of domain shift. We presented a framework, Seg-JDOT, which adapts a deep model so that samples from a source domain and samples from a target domain sharing similar representations will be similarly segmented. We evaluated the framework on a multi-site dataset, MICCAI 2016, and showed that the adaptation towards a target site can bring remarkable improvements in a model performance over standard training [54].

This work was done in collaboration with Nicolas Courty, Obelix team, IRISA laboratory from University of Bretagne Sud.

7.1.5.2. Deep learning for multi-site MS lesions segmentation: two-step intensity standardization and generalized loss function.

Participants: Francesca Galassi, Olivier Commowick, Christian Barillot.

We presented an improved CNN framework for the segmentation of Multiple Sclerosis (MS) lesions from multi-modal MRI. It uses a two-step intensity normalization and a cascaded network with cost sensitive learning. Performance was assessed on a public multi-site data-set [35].

7.2. Research axis 2: Applications in Neuroradiology and Neurological Disorders

Our objectives is also to provide new computational solutions for our target clinical applications (radiology, neurology, psychiatry and rehabilitation...), allowing a more appropriate representation of data for image analysis and the detection of biomarkers specific to a form or grade of pathology, or specific to a population of subjects. In this section, we present our contributions in different clinical applications.

7.2.1. Rehabilitation

7.2.1.1. Efficacy of EEG-fMRI Neurofeedback for stroke rehabilitation in relation to the DTI structural damage: a pilot study.
Participants: Giulia Lioi, Mathis Fleury, Christian Barillot, Isabelle Bonan.

Recent studies have shown the potential of neurofeedback (NF) for motor rehabilitation after stroke. The majority of these NF approaches have relied solely on one imaging technique: mostly on EEG recordings. Recent study have gone further, revealing the potential of integrating complementary techniques such as EEG and fMRI to achieve a more specific regulation. In this exploratory work, multisession bimodal EEG-fMRI NF for upper limb motor recovery was tested in four stroke patients.The feasibility of the NF training was investigated with respect to the integrity of the corticospinal tract (CST), a well-established predictor of the potential for clinical improvement. Results indicated that patients exhibiting a high degree of integrity of the ipsilesional CST showed significant increased activation of the ipsilesional M1 at the end of the training ($p<<0.001$, Wilcoxon test). These preliminary findings confirm the critical role of the CST integrity for stroke motor recovery and indicate that this is importantly related also to functional brain regulation of the ipsilesional motor cortex [43].

7.2.2. Multiple sclerosis

7.2.2.1. Tissue microstructure information from T2 relaxometry and diffusion MRI can identify multiple sclerosis (MS) lesions undergoing blood-brain barrier breakdown (BBB)
Participants: Olivier Commowick, Christian Barillot.

Gadolinium-based contrast agents (GBCA) play a critical role in identifying MS lesions undergoing BBB which is of high clinical importance. However, repeated use of GBCAs over a long period of time and the risks associated with administering it to patients with renal complications has mandated for greater caution in its usage. In this work we explored the plausibility of identifying MS lesions undergoing BBB from tissue microstructure information obtained from T2 relaxometry and dMRI data. We also proposed a framework to predict MS lesions undergoing BBB using the tissue microstructure information and demonstrated its potential on a test case [26].

7.2.2.2. Neural basis of irony in patients with Multiple Sclerosis: an exploratory fMRI study
Participants: Quentin Duché, Élise Bannier.

Irony is a form of non-literal language that is characterized by the opposition between the literal meaning of a statement and the message that the speaker wishes to convey. Knowledge about the neural bases of non-literal language has largely developed in recent years from injury studies or more recently through data from functional imaging studies. Multiple sclerosis (MS) is a neurodegenerative disease that, in addition to cognitive dysfunction, results in variable impairment of theory of mind and non-literal language skills. This work aims at exploring neural basis underpinning the comprehension of irony in MS patients compared to a group of healthy subjects. The results suggest that multiple sclerosis patients require higher left hemisphere resources than healthy controls to understand irony [32].

This work is done in collaboration with by Florian Chapelain (Pôle Saint Hélier), Philippe Gallien (Pôle Saint Hélier) and Virginie Dardier (Université Rennes 2).

7.2.2.3. Joint assessment of brain and spinal cord motor tract damage in patients with early relapsing remitting multiple sclerosis (RRMS): predominant impact of spinal cord lesions on motor function
Participants: Benoit Combès, Élise Bannier, Haykel Snoussi, Jean-Christophe Ferré, Christian Barillot.

The effect of structural multiple sclerosis damage to the corticospinal tract (CST) has been separately evaluated in the brain and spinal cord (SC), even though a cumulative impact is suspected. In this work, we evaluated CST damages on both the cortex and cervical SC, and examine their relative associations with motor function, measured both clinically and by electrophysiology. This study highlights the major contribution of SC lesions to CST damage and motor function abnormalities [8].

This work was done in collaboration with Anne Kerbrat (Neuropoly Montréal) and Raphael Chouteau (CHU Rennes).

7.2.2.4. Spatial distribution of multiple sclerosis lesions in the cervical spinal cord

Participants: Élise Bannier, Gilles Edan.

Spinal cord lesions detected on MRI hold important diagnostic and prognostic value for multiple sclerosis. Our aim was to explore the spatial distribution of multiple sclerosis lesions in the cervical spinal cord, with respect to clinical status. We included 642 suspected or confirmed multiple sclerosis patients (31 clinically isolated syndrome, and 416 relapsing-remitting, 84 secondary progressive, and 73 primary progressive multiple sclerosis) from 13 clinical sites. With an automatic publicly-available analysis pipeline we produced voxelwise lesion frequency maps to identify predilection sites in various patient groups characterized by clinical subtype, Expanded Disability Status Scale score and disease duration. We also measured absolute and normalized lesion volumes in several regions of interest using an atlas-based approach, and evaluated differences within and between groups. The lateral funiculi were more frequently affected by lesions in progressive subtypes than in relapsing in voxelwise analysis ($P < 0.001$), which was further confirmed by absolute and normalized lesion volumes ($P < 0.01$). The central cord area was more often affected by lesions in primary progressive than relapse-remitting patients ($P < 0.001$). Between white and grey matter, the absolute lesion volume in the white matter was greater than in the grey matter in all phenotypes ($P < 0.001$); however when normalizing by each region, normalized lesion volumes were comparable between white and grey matter in primary progressive patients. Lesions appearing in the lateral funiculi and central cord area were significantly correlated with Expanded Disability Status Scale score ($P < 0.001$). High lesion frequencies were observed in patients with a more aggressive disease course, rather than long disease duration. Lesions located in the lateral funiculi and central cord area of the cervical spine may influence clinical status in multiple sclerosis. This work shows the added value of cervical spine lesions, and provides an avenue for evaluating the distribution of spinal cord lesions in various patient groups [14].

This work was done in collaboration with Julien Cohen-Adad (Neuropoly, Montreal) and Anne Kerbrat (Neuropoly Montréal).

7.2.2.5. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks

Participants: Élise Bannier, Gilles Edan.

The goal of this study was to develop a fully-automatic framework - robust to variability in both image parameters and clinical condition - for segmentation of the spinal cord and intramedullary MS lesions from conventional MRI data of MS and non-MS cases. Scans of 1042 subjects (459 healthy controls, 471 MS patients, and 112 with other spinal pathologies) were included in this multi-site study (n=30). Data spanned three contrasts (T1-, T2-, and $T2_*$-weighted) for a total of 1943vol and featured large heterogeneity in terms of resolution, orientation, coverage, and clinical conditions. The proposed cord and lesion automatic segmentation approach is based on a sequence of two Convolutional Neural Networks (CNNs). CNNs were trained independently with the Dice loss. When compared against manual segmentation, our CNN-based approach showed a median Dice of 95% vs. 88% for PropSeg ($p \leq 0.05$), a state-of-the-art spinal cord segmentation method. Regarding lesion segmentation on MS data, our framework provided a Dice of 60%, a relative volume difference of -15%, and a lesion-wise detection sensitivity and precision of 83% and 77%, respectively. In this study, we introduce a robust method to segment the spinal cord and intramedullary MS lesions on a variety of MRI contrasts. The proposed framework is open-source and readily available in the Spinal Cord Toolbox.

This work was done in collaboration with Julien Cohen-Adad (Neuropoly, Montreal) and Anne Kerbrat (Neuropoly Montréal).

7.2.3. Arterial Spin Labeling in pediatric populations

7.2.3.1. Changes in brain perfusion in successive arterial spin labeling MRI scans in neonates with hypoxic-ischemic encephalopathy

Participants: Maia Proisy, Isabelle Corouge, Antoine Legouhy, Christian Barillot, Jean-Christophe Ferré.

The primary objective of this study was to evaluate changes in cerebral blood flow (CBF) using arterial spin labeling MRI between day 4 of life (DOL4) and day 11 of life (DOL11) in neonates with hypoxic-ischemic encephalopathy (HIE) treated with hypothermia. The secondary objectives were to compare CBF values between the different regions of interest (ROIs) and between infants with ischemic lesions on MRI and infants with normal MRI findings. We prospectively included all consecutive neonates with HIE admitted to the neonatal intensive care unit of our institution who were eligible for therapeutic hypothermia. Each neonate systematically underwent two MRI examinations as close as possible to day 4 (early MRI) and day 11 (late MRI) of life. We proposed an innovative processing pipeline for morphological and ASL data suited to neonates that enable automated segmentation to obtain CBF values over ROIs. We evaluated CBF on two successive scanswithin thefirst 15 days of life in the same subjects. ASL imaging in asphyxiated neonates seems more relevant when used relatively early, in the first days of life. The correlation of intra-subject changes in cerebral perfusion between early and late MRI with neurodevelopmental outcome warrants investigation in a larger cohort, to determine whether the CBF pattern change can provide prognostic information beyond that provided by visible structural abnormalities on conventional MRI [18], [47].

7.2.4. Cerebral blood flow in sickle cell populations

7.2.4.1. White matter has impaired resting oxygen delivery in sickle cell patients

Participant: Julie Coloigner.

Although modern medical management has lowered overt stroke occurrence in patients with sickle cell disease (SCD), progressive white matter (WM) damage remains common. It is known that cerebral blood flow (CBF) increases to compensate for anemia, but sufficiency of cerebral oxygen delivery, especially in the WM, has not been systematically investigated. Cerebral perfusion was measured by arterial spin labeling in 32 SCD patients (age range: 10-42 years old, 14 males, 7 with hemoglobin SC, 25 hemoglobin SS) and 25 age and race-matched healthy controls (age range: 15-45 years old, 10 males, 12 with hemoglobin AS, 13 hemoglobin AA); 8/24 SCD patients were receiving regular blood transfusions and 14/24 non-transfused SCD patients were taking hydroxyurea. Imaging data from control subjects were used to calculate maps for CBF and oxygen delivery in SCD patients and their T-score maps. Whole brain CBF was increased in SCD patients with a mean T-score of 0.5 and correlated with lactate dehydrogenase ($r2 = 0.58$, $P < 0.0001$). When corrected for oxygen content and arterial saturation, whole brain and gray matter (GM) oxygen delivery were normal in SCD, but WM oxygen delivery was 35% lower than in controls. Age and hematocrit were the strongest predictors for WM CBF and oxygen delivery in patients with SCD. There was spatial co-localization between regions of low oxygen delivery and WM hyperintensities on T2 FLAIR imaging. To conclude, oxygen delivery is preserved in the GM of SCD patients, but is decreased throughout the WM, particularly in areas prone to WM silent strokes [7].

This work was done in collaboration with Natasha Leporé and her team, Children's hospital Los Angeles, University of Southern California, USA.

7.2.5. Alzheimer disease

7.2.5.1. Abnormal fMRI response in sub-hippocampal structures: how prior knowledge impairs memory in AD

Participants: Quentin Duché, Pierre-Yves Jonin.

Early Alzheimer's disease typically impairs associative learning abilities, up to 18 years before dementia. Importantly, patients' concerns refer to their daily routine, meaning that they lack associative memory for highly familiar stimuli. However, most of the tests involve much less familiar stimuli (e.g. isolated words). It follows that we ignore whether prior knowledge about memoranda alters memory formation and its neural correlates in Alzheimer's Disease. Here, we aimed at manipulating prior knowledge available at encoding and repetition to investigate whether prior knowledge could alter the neural underpinnings of associative encoding, in a way sensitive to early AD. The results suggest that distinct forms of prior knowledge may drive partly non-overlapping brain networks at encoding, and in turn these regions differentially contribute to successful memory formation. Thus, our finding that sub-hippocampal, not hippocampal, activation underlie the inability of the patients to benefit remote prior knowledge in new learning opens perspectives for further diagnostic and prognostic markers development [37].

7.2.5.2. Learning what you know: how prior knowledge impairs new associative learning in early AD.

Participants: Pierre-Yves Jonin, Quentin Duché, Élise Bannier, Isabelle Corouge, Jean-Christophe Ferré, Christian Barillot.

While associative memory impairment is a core feature of prodromal Alzheimer's Disease (AD), whether prior knowledge affects associative learning is largely overlooked. Stimuli repetition yields suppression or enhancement of the BOLD signal, allowing the functional mapping of brain networks. We addressed the role of prior knowledge in associative encoding by manipulating repetition and familiarity of the memoranda in a subsequent memory fMRI study design. 17 patients with prodromal AD (AD-MCI) and 19 Controls learned face-scene associations presented twice in the scanner. Pre-experimental knowledge trials (PEK) involved famous faces while in Experimental Knowledge trials (EK), unknown faces familiarized before scanning were used. Study events were sorted as associative hits, associative misses or misses after a recognition test outside the scanner. We computed the Repetition X Prior knowledge interaction contrast to test whether the encoding networks differed along with prior knowledge, then looked for subsequent associative memory effects in the resulting clusters. PEK and EK yielded similar associative memory performance in AD-MCI, while PEK increased associative memory by 28% in Controls. Repetition effects were modulated by Prior knowledge in Controls, but AD-MCI showed aberrant repetition effects. Subsequent memory effects were observed only in Controls for PEK in the right subhippocampal structures. By contrast, in both groups, EK triggered a subsequent memory effect in the right hippocampus. Provided that tau pathology starts within anterior subhippocampal regions in early AD, our findings that subhippocampal, not hippocampal, involvement underlies the inability of the patients to benefit from PEK open innovative clinical and research perspectives [38].

7.2.6. Depression

7.2.6.1. White matter abnormalities in depression: a categorical and phenotypic diffusion MRI study.

Participants: Julie Coloigner, Olivier Commowick, Isabelle Corouge, Christian Barillot.

Mood depressive disorder is one of the most disabling chronic diseases with a high rate of everyday life disability that affects 350 million people around the world. Recent advances in neuroimaging have reported widespread structural abnormalities, suggesting a dysfunctional frontal-limbic circuit involved in the pathophysiological mechanisms of depression. However, a variety of different white matter regions has been highlighted and these results lack reproducibility of such categorical-based biomarkers. These inconsistent results might be attributed to various factors: actual categorical definition of depression as well as clinical phenotype variability. In this study, we 1/ examined WM changes in a large cohort (114 patients) compared to a healthy control group and 2/ sought to identify specific WM alterations in relation to specific depressive phenotypes such as anhedonia (i.e. lack of pleasure), anxiety and psychomotor retardation –three core symptoms involved in depression. Consistent with previous studies, reduced white matter was observed in the genu of the corpus callosum extending to the inferior fasciculus and posterior thalamic radiation, confirming a frontal-limbic circuit abnormality. Our analysis also reported other patterns of increased fractional anisotropy and axial diffusivity as well as decreased apparent diffusion coefficient and radial diffusivity in the splenium of the corpus callosum and posterior limb of the internal capsule. Moreover, a positive correlation between FA and anhedonia was found in the superior longitudinal fasciculus as well as a negative correlation in the cingulum. Then, the analysis of the anxiety and diffusion metric revealed that increased anxiety was associated with greater FA values in genu and splenium of corpus callosum, anterior corona radiata and posterior thalamic radiation. Finally, the motor retardation analysis showed a correlation between increased Widlöcher depressive retardation scale scores and reduced FA in the body and genu of the corpus callosum, fornix, and superior striatum. Through this twofold approach (categorical and phenotypic), this study has underlined the need to move forward to a symptom-based research area of biomarkers, which help to understand the pathophysiology of mood depressive disorders and to stratify precise phenotypes of depression with targeted therapeutic strategies [9]. This work was done with Centre Hospitalier Guillaume Régnier, Academic Psychiatry Department, 35703 Rennes, France.

7.2.6.2. Structural connectivity analysis in treatment-resistant depression

Participants: Julie Coloigner, Isabelle Corouge, Christian Barillot.

Depressive disorder is characterized by a profound dysregulation of affect and mood as well as additional abnormalities including cognitive dysfunction, insomnia, fatigue and emotional disturbance. Converging evidence shows that a dysfunction in prefrontal-subcortical circuits is associated with depressive state. However, the process of treatment resistance was poorly studied. One study of functional magnetic resonance imaging has reported more disrupted connectivity in prefrontal areas and in thalamus for resistant (R) group (Lui et al., 2011). These observations suggest a modification of functional connectivity in the prefrontal-subcortical circuits in the R patients. Using graph theory-based analysis, we examined white matter changes in the organization of networks in R patients compared with non-resistant (NR) group. We revealed 15 areas with significant density differences in R patients compared to NR subjects. The NR depression seems associated with decreased connectivity among distributed limbic areas, particularly in the ACC and in basal ganglia. However, the R patients exhibit a reduced connectivity in anterior limb of internal capsule and genu of corpus callosum compared with NR patients. Combined with previous studies, which described a widespread disruption in prefrontal-subcortical networks, this result suggests a more important connectivity decrease in the frontal cortex, as well as a smaller reduction in the limbic circuit for the patients with pejorative outcome. These results were consistent with connectivity studies, which suggested that the degree of disruption could influence the resistance severity and that two distinct networks could be implicated in NR end R depression. [27].

7.2.7. Prenatal exposure

7.2.7.1. Prenatal exposure to glycol ethers and motor inhibition function evaluated by functional MRI at the age of 10 to 12years in the PELAGIE mother-child cohort

Participants: Élise Bannier, Christian Barillot.

Pregnant women are ubiquitously exposed to organic solvents, such as glycol ethers. Several studies suggest potential developmental neurotoxicity following exposure to glycol ethers with a lack of clarity of possible brain mechanisms. We investigated the association between urinary levels of glycol ethers of women during early pregnancy and motor inhibition function of their 10- to 12-year-old children by behavioral assessment and brain MR imaging. Prenatal urinary levels of two glycol ether metabolites were associated with poorer Go/No-Go task performance. Differential activations were observed in the brain motor inhibition network in relation with successful inhibition, but not with cognitive demand. Nevertheless, there is no consistence between performance indicators and cerebral activity results. Other studies are highly necessary given the ubiquity of glycol ether exposure [5].

This work is done in collaboration with Fabienne Pelé and Cécile Chevrier (IRSET). Anne Claire Binter defended her PhD in December 2019 supervised by Fabienne Pelé, Cécile Chevrier and Élise Bannier. t

7.2.7.2. Effect of prenatal organic solvent exposure on structural connectivity at childhood

Participants: Julie Coloigner, Élise Bannier, Jean-Christophe Ferré, Christian Barillot.

Glycol ethers are part of organic solvents. They are used in industry and at home during manufacturing or usage of products such as paints, cleaning agents and cosmetics. The specific detection of subtle, low-dose effects of early-life exposure to these solvents on neuropsychological performance in children is a trendy subject of investigation. Neuroimaging allows looking into brain function and identifying different cerebral connections that may be affected by these neurotoxicants. In this paper, we investigated the specific effects of prenatal low-level exposure to different glycol ethers, on brain development of children between 10 and 12 years old. Based on previous studies suggesting cognitive disabilities in the attention, inhibition and working memory, we proposed a structural connectivity analysis using graph theory restricted to the regions involved in these functions. Our results suggest a possible relationship between the attention, working memory and inhibition and prenatal exposure to specific glycol ethers, such as ethoxyacetic acid, ethoxyethoxyacetic acid and 2-butoxyacetic acid [28].

7.2.8. Cognitive food-choice task

7.2.8.1. Implemenation of a new food picture database in the context of fMRI and visual cognitive food-choice task in healthy volunteers

Participant: Élise Bannier.

This pilot study aimed at implementing a new food picture database in the context of functional magnetic resonance imaging (fMRI) cognitive food-choice task, with an internal conflict or not, in healthy normal-weight adults. The fMRI analyses showed that the different liking foods (i.e. foods with different hedonic appraisals) condition elicited the activation of dorsal anterior cingulate cortex, involved in internal conflict monitoring, whereas similar liking (ie, foods with similar hedonic appraisals) condition did not, and that low-energy (LE) food choice involved high-level cognitive processes with higher activation of the hippocampus (HPC) and fusiform gyrus compared to high-energy (HE) food choice. Overall, this pilot study validated the use of the food picture database and fMRI-based procedure assessing decision-making processing during a food choice cognitive task with and without internal conflict[15].

This work was done in collaboration with Yentl Gautier, Paul Meurice, Yann Serrand, Nicolas Coquery Romain Moirand and David Val-Laillet from the NuMeCan Institute (Nutrition Metabolisms Cancer, UMR 1241, Inserm - Université de Rennes 1) and INRA.

7.3. Research axis 3: Management of Information in Neuroimaging

In the context of population imaging, we have made progress in three main areas this year. First we were involved in the development of infrastructures for open science with OpenAIRE, we also participated in the collaborative definition of standards that will ensure that infrastructures remain interoperable. Finally, we started a research new axis looking at how variations in analytical pipelines impact neuroimaging results (i.e. analytic variability).

7.3.1. Infrastructures

7.3.1.1. Open research: linking the bits and pieces with OpenAIRE-connect

Participants: Camille Maumet, Christian Barillot, Xavier Rolland.

Open research is growing in neuroimaging. The community — supported by funders who want best use of public funding but also by the general public who wants more transparent and participatory research practices — is constantly expanding online resources including: data, code, materials , tutorials, etc. This trend will likely amplify in the future and is also observed in other areas of experimental sciences. Open resources are typically deposited in dedicated repositories that are tailored to a particular type of artefact. While this is best practice, it makes it difficult to get the big picture: artefacts are scattered across the web in a multitude of databases. Although one could claim that the publication is here to link all related artefacts together, it is not machine-readable and does not me toallow searching for artefacts using filters (e.g. all datasets created in relation with a given funder). We presented OpenAIRE-connect, an overlay platform that links together research resources stored on the web: https://beta.ni.openaire.eu/ [45].

This work was done in collaboration with Dr. Sorina Caramasu-Pop and Axel Bonnet from Creatis in Lyon and with collaborators of the OpenAIRE-Connect project.

7.3.2. Standardisation and interoperability

7.3.2.1. The best of both worlds: using semantic web with JSON-LD. An example with NIDM-Results & Datalad

Participant: Camille Maumet.

The Neuroimaging data model (NIDM-Results) provides a harmonised representation for fMRI results reporting using Semantic Web technologies. While those technologies are particularly well suited for aggregation across complex datasets, using them can be costly in terms of initial development time to generate and read the corresponding serialisations. While the technology is machine accessible, it can be difficult to comprehend by humans. This hinders adoption by scientific communities and by software developers used to more-lightweight data-exchange formats, such as JSON. JSON-LD: a JSON representation for semantic graphs ("JSON-LD 1.1" n.d.) was created to address this limitation and recent extensions to the specification allow creating JSON-LD documents that are structured more similar to simple JSON. This representation is simultaneously readable by a large number of JSON-based applications and by Semantic Web tools. Here we review our work on building a JSON-LD representation for NIDM-Results data and exposing it to Datalad, a data-management tool suitable for neuroimaging datasets with built-in support for metadata extraction and search [44].

This work was done in collaboration with Prof. Michael Hanke from Institute of Neuroscience and Medicine in Julich and with members of the INCF.

7.3.2.2. Tools for FAIR Neuroimaging Experiment Metadata Annotation with NIDM Experiment

Participant: Camille Maumet.

Acceleration of scientific discovery relies on our ability to effectively use data acquired by consortiums and/or across multiple domains to generate robust and replicable findings. Efficient use of existing data relies on metadata being FAIR1 - Findable, Accessible, Interoperable and Reusable. Typically, data are shared using formats appropriate for the specific data types with little contextual information. Therefore, scientists looking to reuse data must contend with data originating from multiple sources, lacking complete acquisition information and often basic participant information (e.g. sex, age). What is required is a rich metadata standard that allows annotation of participant and data information throughout the experiment workflow, thereby allowing consumers easy discovery of suitable data. The Neuroimaging Data Model (NIDM)2 is an ongoing effort to represent, in a single core technology, the different components of a research activity, their relations, and derived data provenance3. NIDM-Experiment (NIDM-E) is focused on experiment design, source data descriptions, and information on the participants and acquisition information. In this work we report on annotation tools developed as part of the PyNIDM4 application programming interface (API) and their application to annotating and extending the BIDS5 versions of ADHD2006 and ABIDE7 datasets hosted in DataLad[40].

This work was led by Dr David Keator from UCI Irvine and done in collaboration with members of the INCF.

7.3.3. Quantifying analytic variability

7.3.3.1. Exploring the impact of analysis software on task 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 is to understand how choice of software package impacts on analysis results. We use 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 obtain all information on how to process, analyze, and model each dataset from the publications. We make 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 find broad similarities between packages, we also discover marked differences, such as Dice similarity coefficients ranging from 0.000-0.743 in comparisons of thresholded statistic maps between software. We discuss the challenges involved in trying to reanalyse the published studies, and highlight our own efforts to make this research reproducible [6].

This work was done in collaboration with Alexander Bowring and Prof. Thomas Nichols from the Oxford Big Data Institute in the UK.

EPIONE Project-Team

6. New Results

6.1. Medical Image Analysis

6.1.1. Learning a Probabilistic Model for Diffeomorphic Registration and Motion Modeling

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 motion modeling, artificial intelligence, latent variable model, deformation transport

We developed a probabilistic approach for multi-scale deformable image registration in 3-D using conditional variational autoencoder [16], [58] and extended it to a motion model by using cardiac MRI image sequences [40]. This includes:

- A probabilistic formulation of the registration problem through unsupervised learning of an encoded deformation model.
- A generative motion model using explicit time-dependent temporal convolutional networks (Fig. 4).
- Demonstration on cardiac cine-MRI for cardiac motion tracking, simulation, transport and temporal super-resolution.

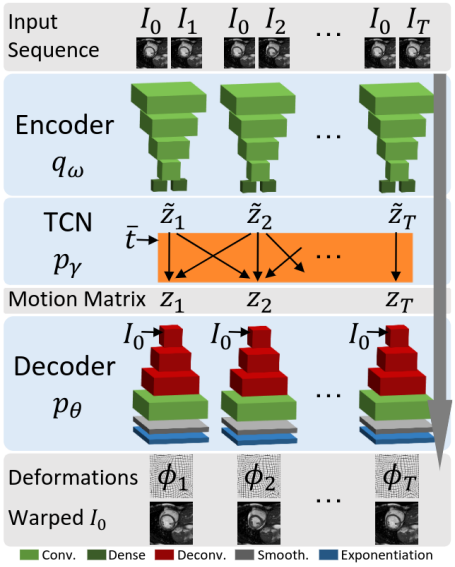


Figure 4. Probabilistic motion model: the encoder q_ω projects the image pair (I_0, I_t) to a probabilistic low-dimensional deformation encoding $\tilde{z}_t$ from which the temporal convolutional network p_γ constructs the motion matrix $z \in \mathbb{R}^{d \times T}$. The decoder p_θ maps the motion matrix to the deformations ϕ_t.

6.1.2. Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis

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

By using multiparametric MRI, we proposed to use a 3D FCNN to predict FLAIR MRI which is used clinically for the detection of WM lesions [26]. In addition, we proposed Sketcher-Refiner GANs to predict PET-derived demyelination from multiparametric MRI [25] with the following contributions:

- Learning the complex relationship between myelin content and multimodal MRI data;
- Comparing quantitatively our approach to other state-of-the-art techniques;
- Proposing visual attention saliency maps to better interpret the neural networks;
- Comparing different combinations of MRI modalities and features to assess which is the optimal input;

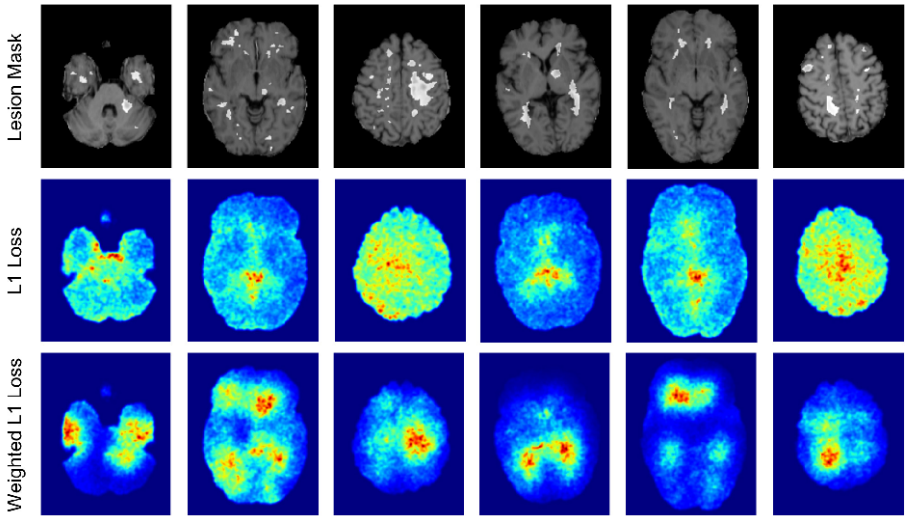


Figure 5. The proposed visual attention saliency map. The white regions shown in first row are MS lesion masks. The second row shows some examples of the attention of neural networks when L1 loss is used as the traditional constraint in the loss function, without the specific weighting scheme that we proposed. The third row shows the corresponding attention of neural networks when our proposed weighted L1 loss is applied. It is clear that our designed loss function is able to effectively shift the attention of neural networks towards MS lesions.

6.1.3. Patch Based Bayesian Mesh Registration

Participants: Paul Blanc-Durand [Correspondant], Hervé Delingette.

A 1 year grant from APHP

Bayesian Modeling, Mesh deformation, Mechanical model

The objective of this work is to co-register two lung CT scans of the same patient acquired at different breathing cycle based on an elastic and Bayesian model of lung deformation. Its originality stems from the joint estimation of a displacement fields and its derivatives (gradient matrix) defined from a tetrahedral mesh. Inference is performed in two alternating steps including the optimization of local affine transforms and the global optimiztion of the displacement.

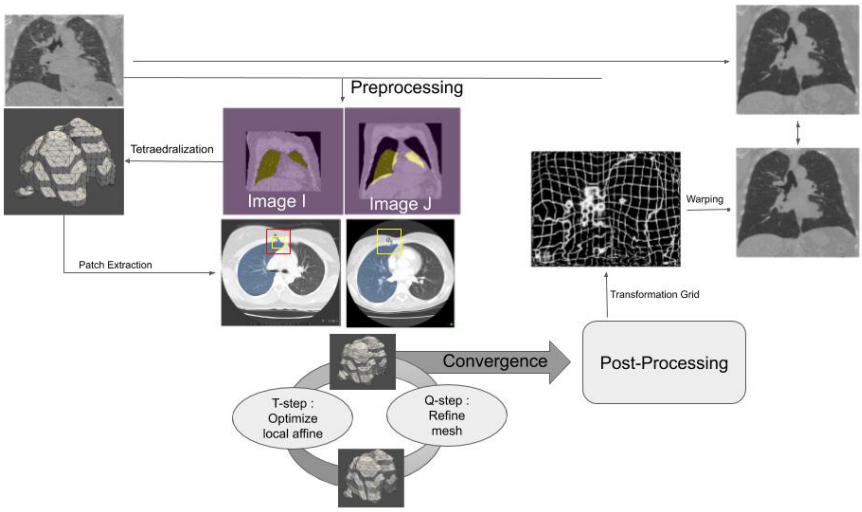


Figure 6. Patches are extracted around vertices of mesh. During T-step, we aim to optimize an affine transform centered on a vertice of image I (the moving image) to image J (the fixed image). The affine transform is regularized under a probabilistic model taking into account the deformation of the mesh. During Q-step, we developed an elastical model of lung which homogenize predictions. After few epochs, convergence is achieved.

6.2. Imaging & Phenomics, Biostatistics

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

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

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.

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

We developed a probabilistic latent variable model able to account for heterogeneous data modalities jointly [6]. In the latent space, this is achieved by constraining the variational distribution of each modality to a common target prior. Moreover, we added *ad hoc* prior distribution and parameterization for the latent space to induce sparsity (Fig. 7 a). This approach is capable to highlight meaningful relationships among biomarkers in the context of Alzheimer's disease (Fig. 7 b) that can be used to develop optimal strategies for disease quantification and prediction.

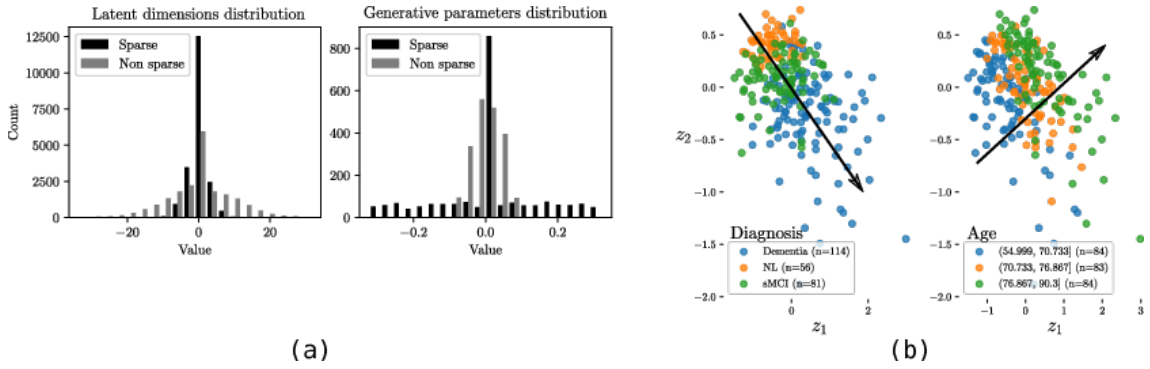


Figure 7. (a) Effect of variational dropout on a synthetic experiment modeled with the Multi-Channel VAE. As expected, the minimum amount of non-zero components of the latent variables (left) and generative parameters (right) is obtained with the sparse model. (b) Stratification of the ADNI subjects (test data) in the sparse latent space. In the same space it is possible to stratify subjects in the test-set by disease status (left) and by age (right) in almost orthogonal directions.

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

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

Lung cancer, Early detection, Biomarkers, Segmentation quality control

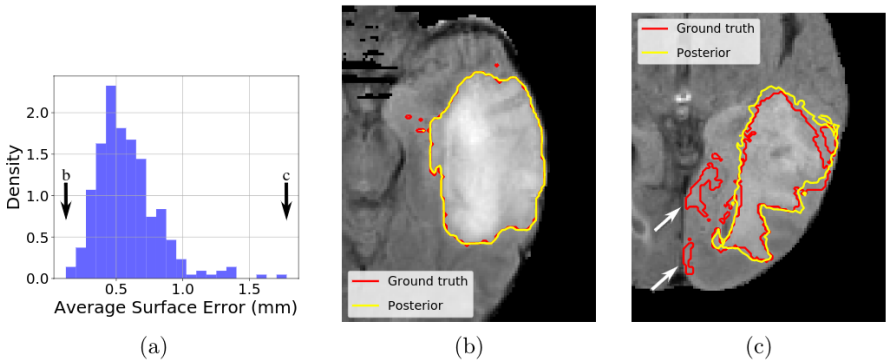
Lung cancer is among the most common cancer and is considered to be one of the most important public health problem. In recent years, immunotherapy has revolutionized cancer treatments but its efficiency is varying among patients. To prevent possible negative side effects there is a critical need in reliable biomarkers capable of predicting the response to immunotherapy treatments. We analyzed the performance of different biomarkers and studied their combination through logistic regression and decision tree models, as part of a joint project with the IRCAN laboratory (Pr P. Hofman, Dr S. Heeke) at Nice hospital [13].

Furthermore, we investigated the issue of automated quality control assessment of image segmentations, which are a key point of medical image processing pipelines. We propose a novel unsupervised quality control approach based on simple intensity and smoothness assumptions [30]. We introduce a novel spatial prior which allows an automatic estimation of all parameters through Bayesian learning. The approach was tested on various medical imaging datasets (Fig. 8).

6.2.3. *Modelling and inference of protein dynamics in neurodegenerative diseases across brain networks*

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".



(a) (b) (c)

Figure 8. Unsupervised quality control of the BRATS 2017 challenge training set. Distribution of the Average Surface Error (a). Example of a segmentation explained by the model (b). Example of a segmentation not explained by the model (c).

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

In this project we propose the first unified framework for the joint estimation of long term neurodegenerative disease progression and kinetic parameters describing pathological protein dynamics across brain networks [48]. The model is expressed within a constrained Gaussian Process regression setting. We use stochastic variational inference for scalable inference and uncertainty quantification. Experiments on simulated data and on AV45-PET brain imaging data measuring topographic amyloid deposition in Alzheimer's disease show that our model accurately recovers prescribed rates along graph dynamics and precisely reconstructs the underlying progression.

6.3. Computational Anatomy & Geometric Statistics

6.3.1. Riemannian Geometric Statistics in Medical Image Analysis

Participants: Xavier Pennec [Correspondant], Stefan Sommer [CPH Univ, DK], Tom Fletcher [University of Virginia at Charlottesville, USA].

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

Geometric statistics, Riemannian geometry, medical image analysis, computational anatomy

There has been a growing need in the medical image computing community for principled methods to process nonlinear geometric data. Riemannian geometry has emerged as one of the most powerful mathematical and computational frameworks for analyzing such data. In the book *Riemannian Geometric Statistics in Medical Image Analysis* [53], we provided an introduction to the core methodology for performing statistics on Riemannian manifolds and more general nonlinear spaces followed by a presentation of state-of-the-art methods in medical image analysis.

We provided more specifically an introduction chapter on differential and Riemannian geometry [56] (with S. Sommer and T. Fletcher), a comprehensive chapter on symmetric positive definite matrices (SPD) and manifold value image processing [55], and reference chapter on the affine connection setting for transformation groups including the stationary velocity fields parametrisation of diffeomorphisms and its use in medical image registration for longitudinal modeling of Alzheimer's disease [54] (with M. Lorenzi) and a chapter on the statistical bias on the estimation in quotient space [52] (with N. Miolane and L Devillier).

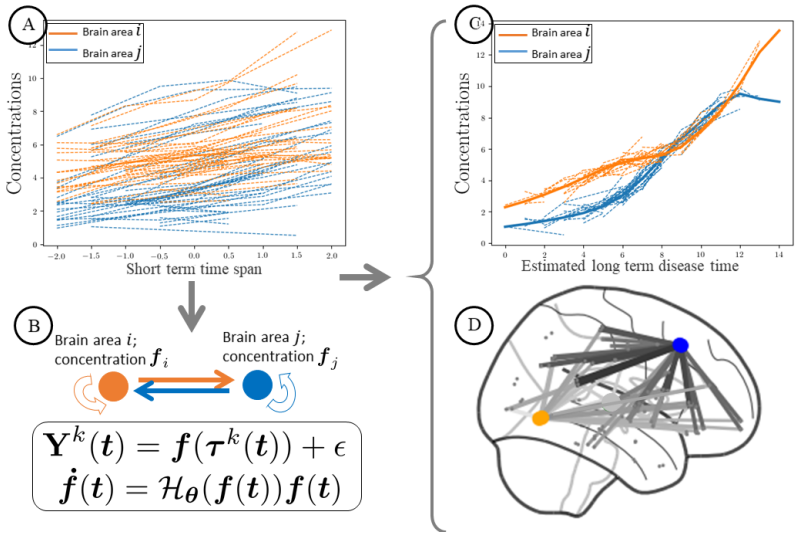


Figure 9. Schematic representation of the proposed framework. Regional protein concentrations are collected for a number of subjects over a short term time span (A). The dynamics of such concentrations are described in terms of a dynamical system for the vector of concentrations (B). The proposed framework estimates such parameters encoding the strength of propagation (D) and the long term protein concentrations with respect to the estimated long term time axis (C).

6.3.2. Effect of curvature on the Empirical Fréchet mean estimation in manifolds

Participant: Xavier Pennec [Correspondant].

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

Geometric statistics, empirical Fréchet mean

Statistical inference in manifolds most often rely on the Fréchet mean in the Riemannian case, or on exponential barycenters in affine connection spaces. The uncertainty of the empirical mean estimation with a fixed number of samples is a key question. In sufficient concentration conditions, a central limit theorem was established in Riemannian manifolds by Bhattacharya and Patrangenaru in 2005. We propose in [62] an asymptotic development valid in Riemannian and affine cases which better explain the role of the curvature in the modulation of the speed of convergence of the empirical mean. We also establish a non-asymptotic development in high concentration which shows a statistical bias on the empirical mean in the direction of the average gradient of the curvature. These curvature effects become important with large curvature and can drastically modify the estimation of the mean. They could partly explain the phenomenon of sticky means recently put into evidence in stratified spaces, notably in the case of negative curvature.

6.3.3. Shape Analysis with diffeomorphisms

Participants: Nicolas Guigui [Correspondant], Shuman Jia, Maxime Sermesant, Xavier Pennec.

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

Shape Analysis, parallel transport, LDDMM, symmetry

The statistical analysis of temporal deformations and inter-subject variability relies on shape registration and parallel transport of deformations (Figure 10). However, the numerical integration and optimization required lead to important numerical errors. This work aims at improving the numerical consistency and reproducibility

of the Pole Ladder scheme to perform parallel transport. We propose a modification of this scheme using registration errors [39] and define different types of errors to evaluate the accuracy: the involutivity and transvectivity. We test our method on 138 cardiac shapes and demonstrate improved numerical consistency for both types of errors.

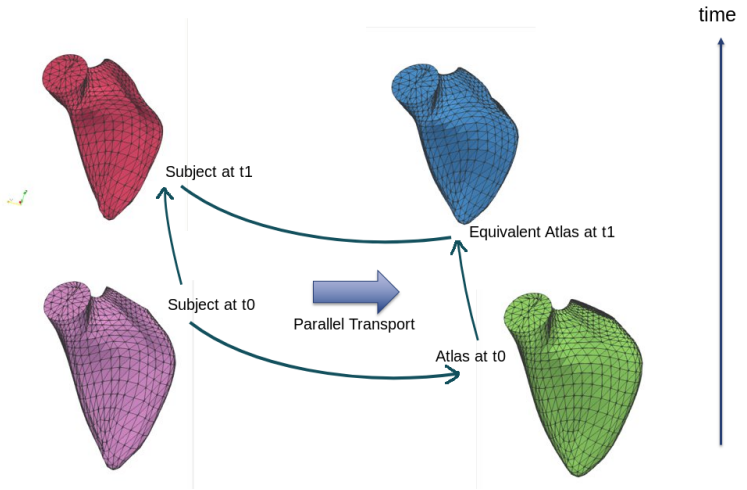


Figure 10. Illustration of our framework using parallel transport to normalize individual temporal deformations to an atlas.

6.3.4. Classification of Riemannian metrics on the manifold of symmetric positive definite matrices

Participants: Yann Thanwerdas [Correspondant], Xavier Pennec.

This work is partially funded by the ERC-Adv G-Statistics and by the IDEX UCA-JEDI ANR-15-IDEX-01 through an excellence OPhD fellowship.

Symmetric Positive Definite matrices, Riemannian metrics, dually flat manifolds

Symmetric Positive Definite matrices have been used in many fields of medical data analysis. Many Riemannian metrics have been defined on this manifold but the choice of the Riemannian structure lacks a set of principles that could lead one to choose properly the metric. We introduced several families of Riemannian metrics supported by a deformation principle and a principle of balanced metrics:

1. Power-Affine and Deformed-Affine metrics [43], that highlight relations between the affine-invariant, the polar-affine and the log-Euclidean metrics ;
2. Mixed-Power-Euclidean and Mixed-Power-Affine metrics [42], that highlight relations between many Riemannian metrics, as shown on Figure 11 .

6.3.5. Statistical shape analysis of faces for computer aided dermatology and plastic surgery

Participants: Florent Jousse [Correspondant], Xavier Pennec, Hervé Delingette, Matilde Gonzalez.

Supported by the company Quantificare through a CIFRE funding.

Gaussian Processes, non rigid registration

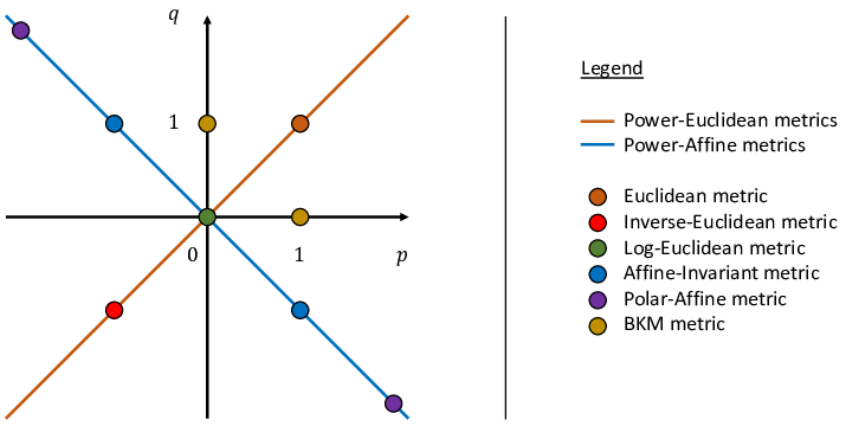


Figure 11. The family of Mixed-Power-Euclidean metrics

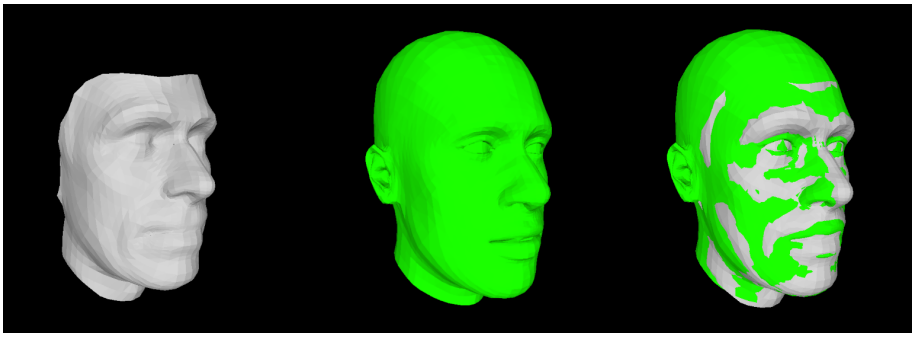


Figure 12. Example of facial template fitting. The white mesh (target) has been acquired by stereoscopy while the green one is a template mesh that has been deformed to fit the target.

The objective of this work is to model complex face deformations related to natural aging, facial expressions, surgical interventions or posture motions to improve the 3D reconstruction of faces and to normalize their analysis. It includes the development of non-rigid registration methods of textured meshes and their statistical modeling through Gaussian processes.

6.3.6. Brain Morphometry in the MAPT clinical trial

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

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

Longitudinal deformation modeling, multivariate statistics, brain morphology, Alzheimer's disease, clinical trial.

- We proposed a complete framework for statistical hypothesis testing on mass-multivariate data. This framework builds on the recent works on multivariate statistics in neuroimaging to propose a generic approach adapted to the study of longitudinal deformations [3].
- This framework is used in the context of the MAPT study to highlight a significant effect of the multidomain intervention on the brain morphological changes (see Figure 13) [22].

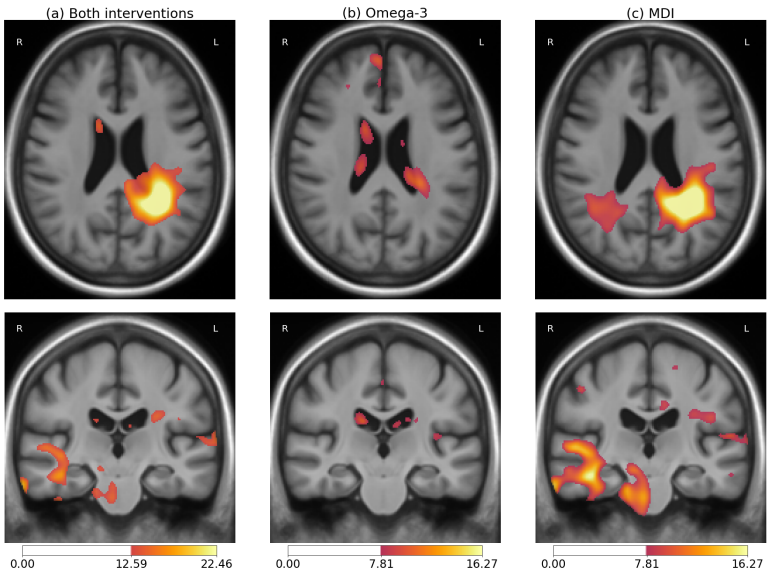


Figure 13. Localization of the MAPT treatments effect on the longitudinal morphological changes for: (a) both categorical variables associated with the omega-3 supplementation and the multidomain intervention, (b) omega-3 only, (c) multidomain intervention only. Color bars indicate the magnitude of the z-values for the likelihood-ratio test. High values indicate a difference in the morphological changes that is associated with the treatment status.

6.3.7. 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 thesis is to develop a spatio-temporal model of Alzheimer's Disease (AD) progession based on multi-modal brain data. 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 Gaussian processes (Figure 14). We also compute an individual time-shift parameter to assess the disease stage of each subject. This work has been accepted for publication in the journal NeuroImage [5]

6.4. Computational Physiology

6.4.1. Deep Learning based Metal Artifacts Reduction in post-operative Cochlear Implant CT Imaging

Participants: Zihao Wang [Correspondant], Clair Vandersteen, Thomas Demarcy, Dan Gnansia, Charles Raffaelli, Nicolas Guevara, Hervé Delingette.

This work is funded by the Provence-Alpes-Côte-d'Azur region, the Université Côte d'Azur and Oticon Medical through CIMPLE https://team.inria.fr/epione/en/research/cimple/ research project.

Generative Adversarial Network, Metal Artifacts Reduction, Cochlea Implantation

We propose a 3D metal artifact reduction method using convolutional neural networks for post-operative cochlear implant imaging.[44]

- Learn metal artifacts reduction by using pre-operative images and metal artifacts simulation to create image pairs for training GANs.
- Metal artifacts simulation starts from a cochlea implantation fusion image and ends with the simulated post-operative image.(Fig. 15)
- A 3D generative adversarial network (MARGANs) to create an image with a reduction of metal artifacts.
- Evaluations on ten patients show the effectiveness of artefact reduction compared to two classical methods.

6.4.2. Kinematic Spiral Shape Recognition in the Human Cochlea

Participants: Wilhelm Wimmer [Correspondant], Clair Vandersteen, Nicolas Guevara, Marco Caversaccio, Hervé Delingette.

Supported by the Swiss National Science Foundation (no. P400P2_180822) and the French government (UCA JEDI - ANR-15-IDEX-01).

Approximate maximum likelihood, kinematic surface recognition, natural growth

To improve therapies for hearing loss and deafness, e.g., with auditory neuroprostheses, we developed a reliable detection algorithm for the cochlear modiolar axis in CT images (Fig. 16). The algorithm was tested in an experimental study with 4 experts in 23 human cochlea CT data sets [45] [27]. Our experiments showed that the algorithm reduces the alignment error providing more reliable modiolar axis detection for clinical and research applications.

6.5. Computational Cardiology & Image-Based Cardiac Interventions

6.5.1. Cardial Electrophysiological Model Learning and Personalisation

Participants: Nicolas Cedilnik [Correspondant], Ibrahim Ayed [Sorbonne, LIP6, Paris], Hubert Cochet [IHU Liryc, Bordeaux], Patrick Gallinari [Sorbonne, LIP6, Paris], Maxime Sermesant.

This work is funded by the IHU Liryc, Bordeaux.

modelling, electrophysiology, ventricular tachycardia, ischemic cardiomyopathy

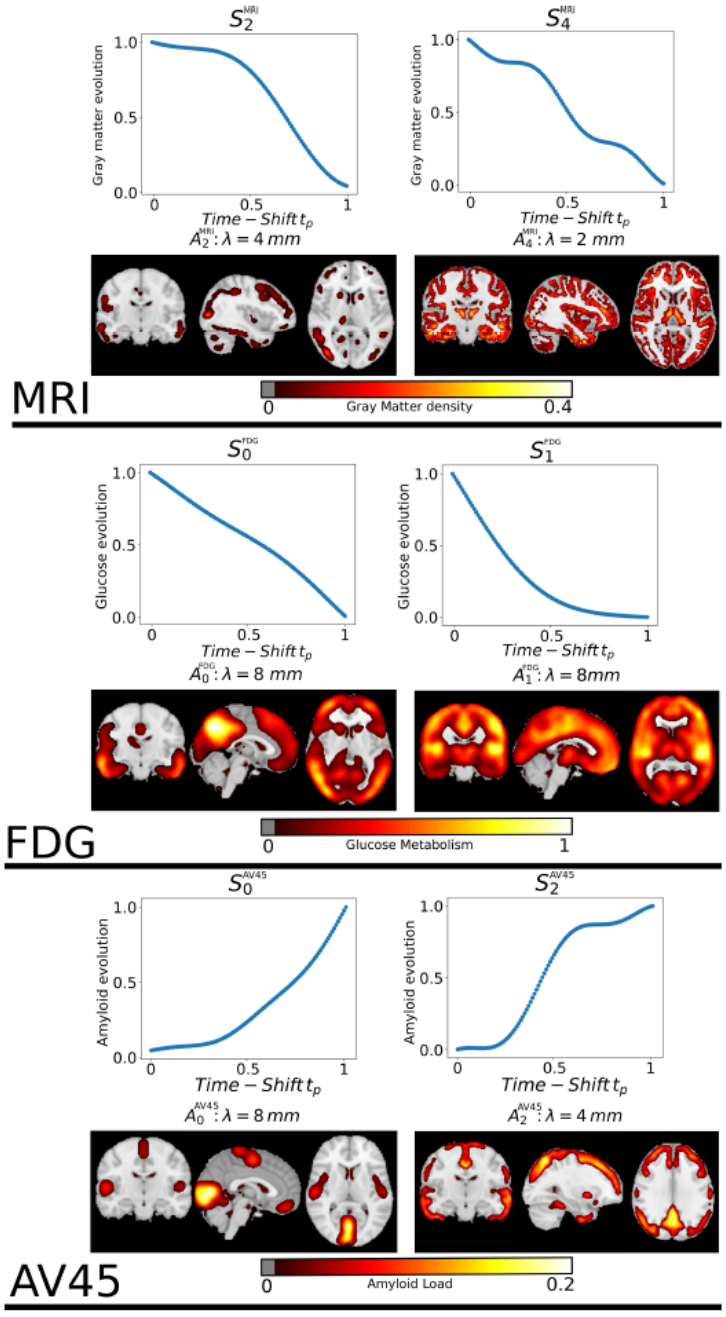


Figure 14. Estimated spatio-temporal processes affecting the brain during Alzheimer's Disease for three different imaging markers.

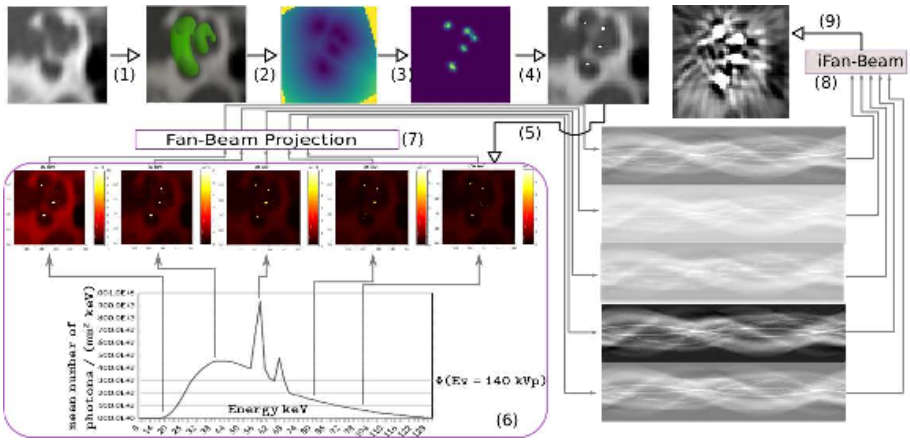


Figure 15. CI metal artifacts simulation workflow starting from a pre-operative image and ending with the simulated post-operative image after 9 processing steps.

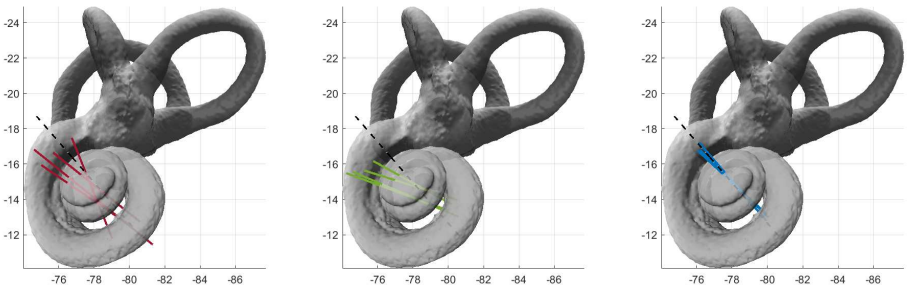


Figure 16. Visualization of the bony labyrinth with reference modiolar axis (dashed line). Modiolar axes after manual landmark-based (left), PCA-based (middle), and robust kinematic detection (right) in CT data are shown for comparison.

This project aims at making electrophysiological model personalisation enter clinical practice in interventional cardiology. During this year:

- we evaluated a fully automated computed tomography-based model personalisation framework in the context of post-ischemic ventricular tachycardia [35],
- we developped a model personalisation methodology based on invasive data in our participation in the STACOM2019 modelling challenge [37],
- we proposed a deep learning based approach to replace numerical integration of partial differential equations used in cardiac modelling [32], see Figure 17 .

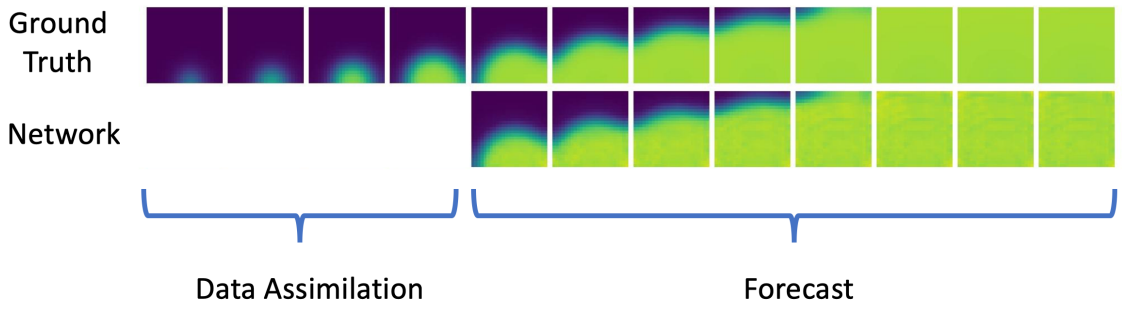


Figure 17. Transmembrane potential obtained with a reaction diffusion model (top) and forecasted by EP-Net (bottom) for one slice of a tissue slab

6.5.2. Deep Learning Formulation of ECGI for Data-driven Integration of Spatiotemporal Correlations and Imaging Information

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.

Deep Learning, Electrocardiographic Imaging, Inverse problem of ECG, Electrical simulation, Generative Model.

Electrocardiographic imaging (ECGI) aims at reconstructing the electrical activity of the heart using body surface potentials.To achieve this one has to solve the ill-posed inverse problem of the torso propagation. We propose in [33] a novel Deep Learning method based on Conditional Variational Autoencoder able to solve ECGI inverse problem in 2D. This generative probabilistic model learns geometrical and spatio-temporal information and enables to generate the corresponding activation map of the specific heart.

120 activation maps and the corresponding Body Surface Potentials (BSP) were generated using the dipole formulation. 80% of the simulated data was used for training and 20% for testing.We generate 10 propbable solutions for each given input using our model. The Mean Squarre Error (MSE) metric over all the tests was 0.095. As results we were able to observe that the reconstruction performs well. Next, we will extend the model in 3D and test it on real data provided by the IHU Liryc.

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

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

Université Côte d'Azur (UCA)

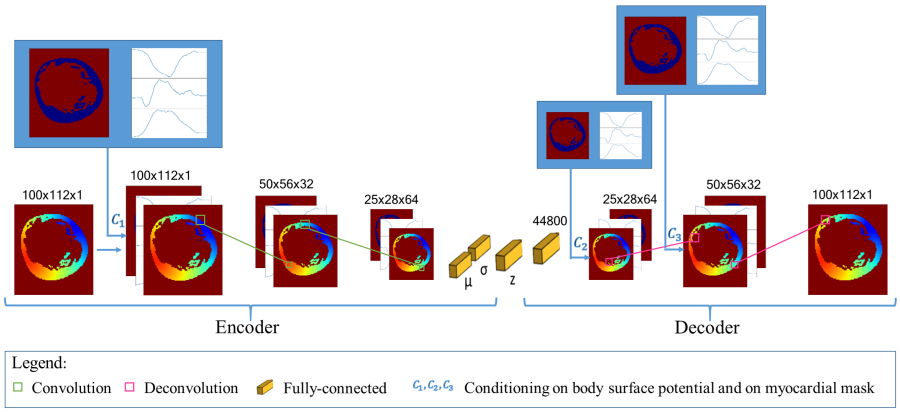


Figure 18. Architecture of our conditioned generative model (encoder) and our conditioned variational approximation (decoder)

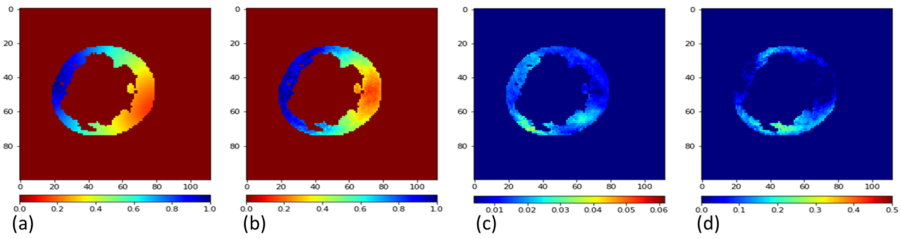


Figure 19. (a) Simulated and (b) predicted mean activation maps for proposed deep learning based ECGI, (c) Standard deviation map calculated over 10 predictions, (d) error map, difference between predicted and simulated activation maps.

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. [34]

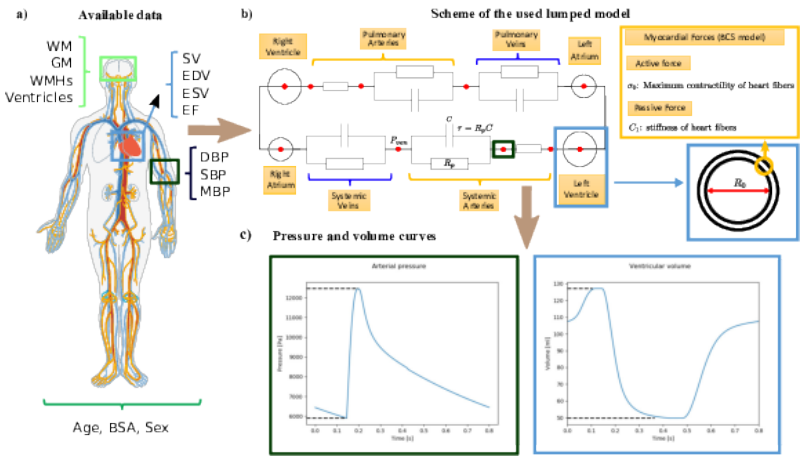


Figure 20. a) Summary of the available data for each subject, including cardiac data, socio-demographic information, blood pressure measurements and brain volumetric indicators. b) Simplified representation of the lumped model showing the parameters used in the personalisation. τ characterizes the contractility of the main systemic arteries, R_p the peripheral resistance, P_{ven} the venous pressure right after the capillaries, R_0 the radius of the left ventricle, σ_0 the contractility of the cardiac fibers and C_1 their stiffness. A more detailed representation of the myocardial forces is omitted for the sake of clarification. c) Example of the pressure and volume curves that can be obtained from the model, from these curves we extract scalar indicators to match the available clinical data.

6.5.4. Parallel transport of surface deformations from pole ladder to symmetrical extension

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

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

We proposed a general scheme to perform statistical modeling of the temporal deformation of the heart, directly based on meshes. We encoded the motion and the intersubject shape variations, with diffeomorphisms parameterized either by stationary SVFs or by time-varying velocity fields in the LDDMM framework.

Experiments on a 4D right-ventricular endocardial meshes database demonstrated the stability of our transport algorithm, of importance for the assessment of pathological changes. The method is adaptable to other anatomies with temporal or longitudinal data.

6.5.5. Machine Learning and Pulmonary hypertension

Participants: Yingyu Yang [Correspondant], Stephane Gillon, Jaume Banus Cobo, Pamela Moceri, Maxime Sermesant.

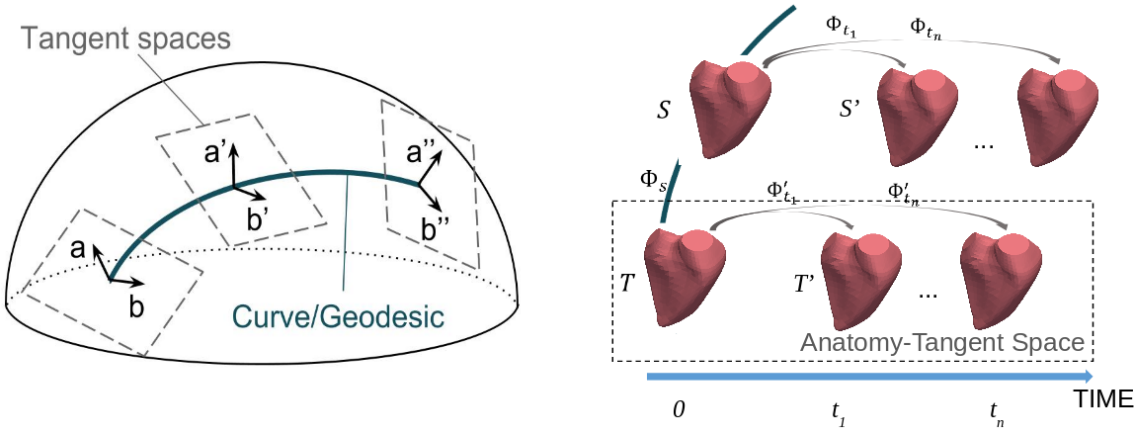


Figure 21. Illustration of parallel transport of vectors a and b along a curve (left) and its application to cardiac imaging (right) with a focus on surfaces.

cardiac modelling, machine learning

Right heart catheterisation is considered as the gold standard for the assessment of patients with suspected pulmonary hyper-tension. It provides clinicians with meaningful data, such as pulmonary capillary wedge pressure and pulmonary vascular resistance, however its usage is limited due to its invasive nature. Non-invasive alternatives, like Doppler echocardiography could present insightful measurements of right heart but lack detailed information related to pulmonary vasculature. In order to explore non-invasive means, we studied a dataset of 95 pulmonary hypertension patients, which includes measurements from echocardiography and from right-heart catheterisation. We used data extracted from echocardiography to conduct cardiac circulation model personalisation and tested its prediction power of catheter data. Standard machine learning methods were also investigated for pulmonary artery pressure prediction. Our preliminary results demonstrated the potential prediction power of both data-driven and model-based approaches. It was published as "Non-Invasive Pressure Estimation in Patients with Pulmonary Arterial Hypertension: Data-driven or Model-based?" accepted at 10th Workshop on Statistical Atlases and Computational Modelling of the Heart, Oct 2019, Shenzhen, China [46]

6.5.6. *Style Data Augmentation for Robust Segmentation of Multi-Modality Cardiac MRI*

Participants: Buntheng Ly [Correspondent], Hubert Cochet [IHU Liryc, Bordeaux], Maxime Sermesant.

Image Segmentation. Multi-modality, Cardiac Magnetic Resonance Imaging, Late Gadolinium Enhanced, Deep Learning

We propose a data augmentation method to improve the segmentation accuracy of the convolutional neural network on multi-modality cardiac magnetic resonance dataset [41].

The strategy aims to reduce over-fitting of the network toward any specific intensity or contrast of the training images by introducing diversity in these two aspects, as shown in figure 23 .

6.5.7. *Towards Hyper-Reduction of Cardiac Models using Poly-Affine Deformation*

Participants: Gaëtan Desrues [Correspondant], Hervé Delingette, Maxime Sermesant.

Model Order Reduction, Finite Elements Method, Affine Transformation, Meshless

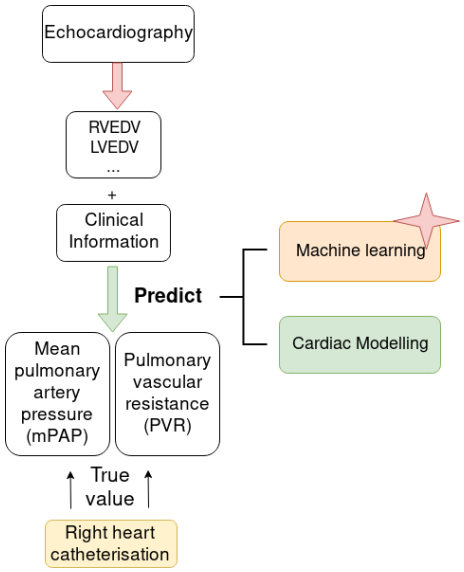


Figure 22. The main idea and logic of this work

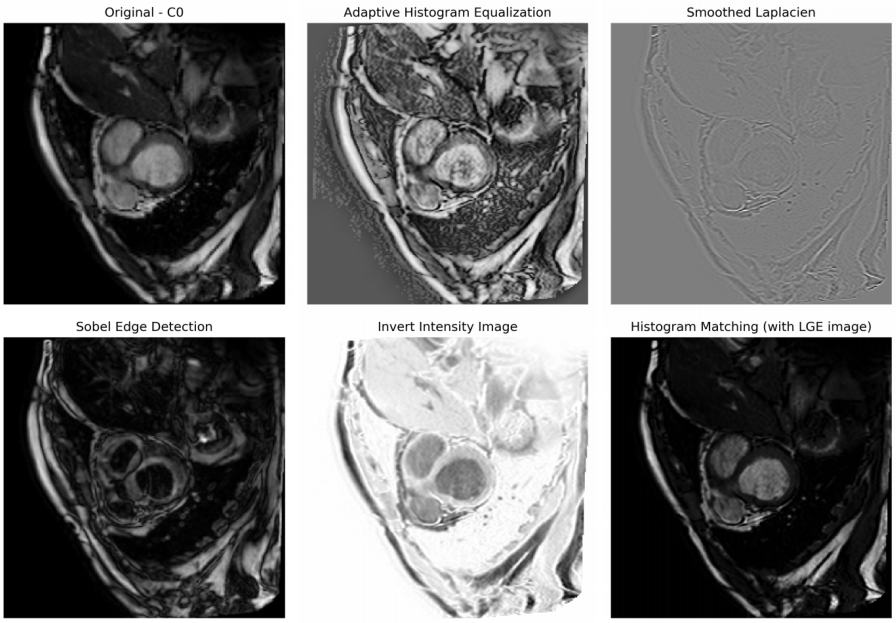


Figure 23. Different variation of input images and the image processing methods used. C0 denotes the steady-state free precessing CMR modality image.

Patient-specific 3D models can help in improving therapy selection, treatment optimization and interventional training. However, these simulations generally have an important computational cost. The aim of this project is to optimize a 3D electromechanical model of the heart for faster simulations [38]. The cardiac deformation is approached by a reduced number of degrees of freedom represented by affine transformations (frames in Figure 24 b) located at the center of the AHA regions (Figure 24 a). The displacement of the material points are computed using region-based shape functions (Figure 24 c).

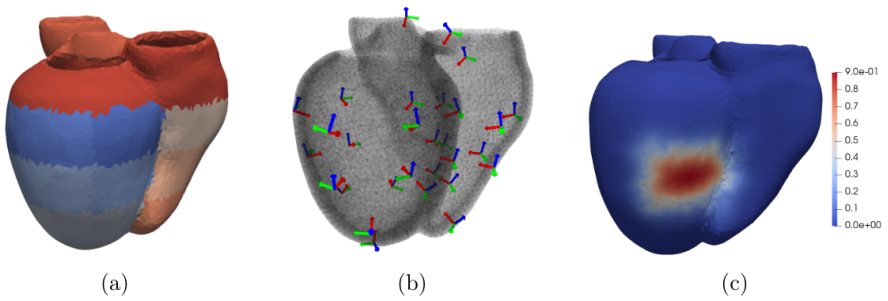


Figure 24. Framework on a cardiac topology. AHA regions (a). Affine degrees of freedom (b). Shape function in one region (c).

MATHNEURO Project-Team

5. New Results

5.1. Neural Networks as dynamical systems

5.1.1. Metastable Resting State Brain Dynamics

Participants: Peter Beim Graben [Brandenburg University of Technology Cottbus, Germany], Antonio Jimenez-Marin [Computational Neuroimaging Lab, BioCruces-Bizkaia Health Research Institute, Spain], Ibai Diez [Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA], Jesus M Cortes [Computational Neuroimaging Lab, BioCruces-Bizkaia Health Research Institute, Spain], Mathieu Desroches, Serafim Rodrigues [Ikerbasque & MCEN team, Basque Center for Applied Mathematics, Spain].

Metastability refers to the fact that the state of a dynamical system spends a large amount of time in a restricted region of its available phase space before a transition takes place, bringing the system into another state from where it might recur into the previous one. Beim Graben and Hutt (2013) [74] suggested to use the recurrence plot (RP) technique introduced by Eckmann et al. (1987) [57] for the segmentation of system's trajectories into metastable states using recurrence grammars. Here, we apply this recurrence structure analysis (RSA) for the first time to resting-state brain dynamics obtained from functional magnetic resonance imaging (fMRI). Brain regions are defined according to the brain hierarchical atlas (BHA) developed by Diez et al. (2015) [56], and as a consequence, regions present high-connectivity in both structure (obtained from diffusion tensor imaging) and function (from the blood-level dependent-oxygenation–BOLD–signal). Remarkably, regions observed by Diez et al. were completely time-invariant. Here, in order to compare this static picture with the metastable systems dynamics obtained from the RSA segmentation, we determine the number of metastable states as a measure of complexity for all subjects and for region numbers varying from 3 to 100. We find RSA convergence toward an optimal segmentation of 40 metastable states for normalized BOLD signals, averaged over BHA modules. Next, we build a bistable dynamics at population level by pooling 30 subjects after Hausdorff clustering. In link with this finding, we reflect on the different modeling frameworks that can allow for such scenarios: heteroclinic dynamics, dynamics with riddled basins of attraction, multiple timescale dynamics. Finally, we characterize the metastable states both functionally and structurally, using templates for resting state networks (RSNs) and the automated anatomical labeling (AAL) atlas, respectively.

This work has been published in Frontiers in Computational Neuroscience and is available as [20].

5.1.2. Controlling seizure propagation in large-scale brain networks

Participants: Simona Olmi, Spase Petkoski [Institut de Neurosciences des Systèmes, Marseille], Maxime Guye [CEMEREM, Hôpital de la Timone, Marseille], Fabrice Bartolomei [Hôpital de la Timone, 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 published in PLoS Computational Biology and is available as [29].

5.1.3. *Chimera states in pulse coupled neural networks: the influence of dilution and noise*

Participants: Simona Olmi, Alessandro Torcini [Institute of Complex Systems, Florence, Italy].

We analyse the possible dynamical states emerging for two symmetrically pulse coupled populations of leaky integrate-and-fire neurons. In particular, we observe broken symmetry states in this set-up: namely, breathing chimeras, where one population is fully synchronized and the other is in a state of partial synchronization (PS) as well as generalized chimera states, where both populations are in PS, but with different levels of synchronization. Symmetric macroscopic states are also present, ranging from quasi-periodic motions, to collective chaos, from splay states to population anti-phase partial synchronization. We then investigate the influence disorder, random link removal or noise, on the dynamics of collective solutions in this model. As a result, we observe that broken symmetry chimera-like states, with both populations partially synchronized, persist up to 80 % of broken links and up to noise amplitudes 8 % of threshold-reset distance. Furthermore, the introduction of disorder on symmetric chaotic state has a constructive effect, namely to induce the emergence of chimera-like states at intermediate dilution or noise level.

This work has been published as a chapter in the book Nonlinear Dynamics in Computational Neuroscience (Springer, 2019) and is available as [35].

5.1.4. *Enhancing power grid synchronization and stability through time delayed feedback control*

Participants: Halgurd Taher, Simona Olmi, Eckehard Schöll [Technical University Berlin, Germany].

We study the synchronization and stability of power grids within the Kuramoto phase oscillator model with inertia with a bimodal frequency distribution representing the generators and the loads. We identify critical nodes through solitary frequency deviations and Lyapunov vectors corresponding to unstable Lyapunov exponents. To cure dangerous deviations from synchronization we propose time-delayed feedback control, which is an efficient control concept in nonlinear dynamic systems. Different control strategies are tested and compared with respect to the minimum number of controlled nodes required to achieve synchronization and Lyapunov stability. As a proof of principle, this fast-acting control method is demonstrated using a model of the German power transmission grid.

This work has been published in Physical Review E and is available as [32].

5.1.5. *Stability and control of power grids with diluted network topology*

Participants: Liudmila Tumash [gipsa-lab, CNRS, Grenoble], Simona Olmi, Eckehard Schöll [Technical University Berlin, Germany].

In the present study we consider a random network of Kuramoto oscillators with inertia in order to mimic and investigate the dynamics emerging in high-voltage power grids. The corresponding natural frequencies are assumed to be bimodally Gaussian distributed, thus modeling the distribution of both power generators and consumers: for the stable operation of power systems these two quantities must be in balance. Since synchronization has to be ensured for a perfectly working power grid, we investigate the stability of the desired synchronized state. We solve this problem numerically for a population of N rotators regardless of the level of quenched disorder present in the topology. We obtain stable and unstable solutions for different initial phase conditions, and we propose how to control unstable solutions, for sufficiently large coupling strength, such that they are stabilized for any initial phase. Finally, we examine a random Erdös-Renyi network under the impact of white Gaussian noise, which is an essential ingredient for power grids in view of increasing renewable energy sources.

This work has been published in Chaos: An Interdisciplinary Journal of Nonlinear Science and is available as [33].

5.1.6. Modeling dopaminergic modulation of clustered gamma rhythms

Participants: Denis Zakharov [Center for Cognition and Decision Making, HSE, Moscow, Russia], Martin Krupa [UCA, LJAD, Inria MathNeuro], Boris Gutkin [Laboratoire de Neurosciences Cognitives, ENS, Paris].

Gamma rhythm (20-100Hz) plays a key role in numerous cognitive tasks: working memory, sensory processing and in routing of information across neural circuits. In comparison with lower frequency oscillations in the brain, gamma-rhythm associated firing of the individual neurons is sparse and the activity is locally distributed in the cortex. Such "weak" gamma rhythm results from synchronous firing of pyramidal neurons in an interplay with the local inhibitory interneurons in a "pyramidal-interneuron gamma" or PING. Experimental evidence shows that individual pyramidal neurons during such oscillations tend to fire at rates below gamma, with the population showing clear gamma oscillations and synchrony. One possible way to describe such features is that this gamma oscillation is generated within local synchronous neuronal clusters. The number of such synchronous clusters defines the overall coherence of the rhythm and its spatial structure. The number of clusters in turn depends on the properties of the synaptic coupling and the intrinsic properties of the constituent neurons. We previously showed that a slow spike frequency adaptation current in the pyramidal neurons can effectively control cluster numbers. These slow adaptation currents are modulated by endogenous brain neuromodulators such as dopamine, whose level is in turn related to cognitive task requirements. Hence we postulate that dopaminergic modulation can effectively control the clustering of weak gamma and its coherence. In this paper we study how dopaminergic modulation of the network and cell properties impacts the cluster formation process in a PING network model.

This work has been accepted for publication in Communications in Nonlinear Science and Numerical Simulation and is available as [34].

5.2. Mean field theory and stochastic processes

5.2.1. Mean-field limit of interacting 2D nonlinear stochastic spiking neurons

Participants: Benjamin Aymard, Fabien Campillo, Romain Veltz.

In this work, we propose a nonlinear stochastic model of a network of stochastic spiking neurons. We heuristically derive the mean-field limit of this system. We then design a Monte Carlo method for the simulation of the microscopic system, and a finite volume method (based on an upwind implicit scheme) for the mean-field model. The finite volume method respects numerical versions of the two main properties of the mean-field model, conservation and positivity, leading to existence and uniqueness of a numerical solution. As the size of the network tends to infinity, we numerically observe propagation of chaos and convergence from an individual description to a mean-field description. Numerical evidences for the existence of a Hopf bifurcation (synonym of synchronised activity) for a sufficiently high value of connectivity, are provided.

This work has been submitted for publication and is available as [38].

5.2.2. Stochastic modeling for biotechnologies Anaerobic model AM2b

Participants: Fabien Campillo, Mohsen Chebbi [ENIT, University of Tunis, Tunisia], Salwa Toumi [INSAT, University of Carthage, Tunisia].

The model AM2b is conventionally represented by a system of differential equations. However, this model is valid only in a large population context and our objective is to establish several stochastic models at different scales. At a microscopic scale, we propose a pure jump stochastic model that can be simulated exactly. But in most situations this exact simulation is not feasible, and we propose approximate simulation methods of Poisson type and of diffusive type. The diffusive type simulation method can be seen as a discretization of a stochastic differential equation. Finally, we formally present a result of law of large numbers and of functional central limit theorem which demonstrates the convergence of these stochastic models towards the initial deterministic models.

This work has been published in ARIMA and is available as [22].

5.2.3. Cross frequency coupling in next generation inhibitory neural mass models

Participants: Andrea Ceni [University of Exeter, UK], Simona Olmi, Alessandro Torcini [Institute of Complex Systems, Florence, Italy], David Angulo-Garcia [Polytechnic University of Cartagena, Colombia].

Coupling among neural rhythms is one of the most important mechanisms at the basis of cognitive processes in the brain. In this study we consider a neural mass model, rigorously obtained from the microscopic dynamics of an inhibitory spiking network with exponential synapses, able to autonomously generate collective oscillations (COs). These oscillations emerge via a super-critical Hopf bifurcation, and their frequencies are controlled by the synaptic time scale, the synaptic coupling and the excitability of the neural population. Furthermore, we show that two inhibitory populations in a master-slave configuration with different synaptic time scales can display various collective dynamical regimes: namely, damped oscillations towards a stable focus, periodic and quasi-periodic oscillations, and chaos. Finally, when bidirectionally coupled the two inhibitory populations can exhibit different types of theta-gamma cross-frequency couplings (CFCs): namely, phase-phase and phase-amplitude CFC. The coupling between theta and gamma COs is enhanced in presence of a external theta forcing, reminiscent of the type of modulation induced in Hippocampal and Cortex circuits via optogenetic drive.

This work has been submitted for publication and is available as [40].

5.2.4. Conductance-Based Refractory Density Approach for a Population of Bursting Neurons

Participants: Anton Chizhov [IOFFE Institute, St Petersburg, Russia], Fabien Campillo, Mathieu Desroches, Antoni Guillamon [Polytechnic University of Catalonia, Barcelona, Spain], Serafim Rodrigues [Ikerbasque & MCEN team, Basque Center for Applied Mathematics, Spain].

The conductance-based refractory density (CBRD) approach is a parsimonious mathematical-computational framework for modelling interacting populations of regular spiking neurons, which, however, has not been yet extended for a population of bursting neurons. The canonical CBRD method allows to describe the firing activity of a statistical ensemble of uncoupled Hodgkin-Huxley-like neurons (differentiated by noise) and has demonstrated its validity against experimental data. The present manuscript generalises the CBRD for a population of bursting neurons; however, in this pilot computational study, we consider the simplest setting in which each individual neuron is governed by a piecewise linear bursting dynamics. The resulting population model makes use of slow-fast analysis, which leads to a novel methodology that combines CBRD with the theory of multiple timescale dynamics. The main prospect is that it opens novel avenues for mathematical explorations, as well as, the derivation of more sophisticated population activity from Hodgkin-Huxley-like bursting neurons, which will allow to capture the activity of synchronised bursting activity in hyper-excitable brain states (e.g. onset of epilepsy).

This work has been published in Bulletin of Mathematical Biology and is available as [23].

5.2.5. 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 dynamics 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 published in Stochastic Processes and their Applications and is available as [24].

5.2.6. Effective low-dimensional dynamics of a mean-field coupled network of slow-fast spiking lasers

Participants: Axel Dolcemascolo [INPHYNI, Nice], Alexandre Miazek [INPHYNI, Nice], Romain Veltz, Francesco Marino [National Institute of Optics, Italy], Stéphane Barland [INPHYNI, Nice].

Low dimensional dynamics of large networks is the focus of many theoretical works, but controlled laboratory experiments are comparatively very few. Here, we discuss experimental observations on a mean-field coupled network of hundreds of semiconductor lasers, which collectively display effectively low-dimensional mixed mode oscillations and chaotic spiking typical of slow-fast systems. We demonstrate that such a reduced dimensionality originates from the slow-fast nature of the system and of the existence of a critical manifold of the network where most of the dynamics takes place. Experimental measurement of the bifurcation parameter for different network sizes corroborate the theory.

This work has been submitted for publication and is available as [42].

5.2.7. The mean-field limit of a network of Hopfield neurons with correlated synaptic weights

Participants: Olivier Faugeras, James Maclaurin [NJIT, USA], Étienne Tanré [Inria Tosca].

We study the asymptotic behaviour for asymmetric neuronal dynamics in a network of Hopfield neurons. The randomness in the network is modelled by random couplings which are centered Gaussian correlated random variables. We prove that the annealed law of the empirical measure satisfies a large deviation principle without any condition on time. We prove that the good rate function of this large deviation principle achieves its minimum value at a unique Gaussian measure which is not Markovian. This implies almost sure convergence of the empirical measure under the quenched law. We prove that the limit equations are expressed as an infinite countable set of linear non Markovian SDEs.

This work has been submitted for publication and is available as [43].

5.2.8. Asymptotic behaviour of a network of neurons with random linear interactions

Participants: Olivier Faugeras, Émilie Soret, Étienne Tanré [Inria Tosca].

We study the asymptotic behaviour for asymmetric neuronal dynamics in a network of linear Hopfield neurons. The randomness in the network is modelled by random couplings which are centered i.i.d. random variables with finite moments of all orders. We prove that if the initial condition of the network is a set of i.i.d random variables with finite moments of all orders and independent of the synaptic weights, each component of the limit system is described as the sum of the corresponding coordinate of the initial condition with a centered Gaussian process whose covariance function can be described in terms of a modified Bessel function. This process is not Markovian. The convergence is in law almost surely w.r.t. the random weights. Our method is essentially based on the CLT and the method of moments.

This work has been submitted for publication and is available as [44].

5.2.9. On a toy network of neurons interacting through their dendrites

Participants: Nicolas Fournier [LPSM, Sorbonne Université], 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 $w_n \simeq N^{-1/2}$. We make use of some recent versions of the results of Deuschel and Zeitouni [55] 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 accepted for publication in Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques and is available as [27].

5.2.10. Bumps and oscillons in networks of spiking neurons

Participants: Helmut Schmidt [Max Planck Institute for Human Cognitive and Brain Science, Germany], Daniele Avitabile [VU Amsterdam, Inria MathNeuro].

We study localized patterns in an exact mean-field description of a spatially-extended network of quadratic integrate-and-fire (QIF) neurons. We investigate conditions for the existence and stability of localized solutions, so-called bumps, and give an analytic estimate for the parameter range where these solutions exist in parameter space, when one or more microscopic network parameters are varied. We develop Galerkin methods for the model equations, which enable numerical bifurcation analysis of stationary and time-periodic spatially-extended solutions. We study the emergence of patterns composed of multiple bumps, which are arranged in a snake-and-ladder bifurcation structure if a homogeneous or heterogeneous synaptic kernel is suitably chosen. Furthermore, we examine time-periodic, spatially-localized solutions (oscillons) in the presence of external forcing, and in autonomous, recurrently coupled excitatory and inhibitory networks. In both cases we observe period doubling cascades leading to chaotic oscillations.

This work has been submitted for publication and is available as [46].

5.2.11. Slow-fast dynamics in the mean-field limit of neural networks

Participants: Daniele Avitabile [VU Amsterdam, Inria MathNeuro], Emre Baspinar, Mathieu Desroches, Olivier Faugeras.

In the context of the Human Brain Project (HBP, see section 5.1.1.1. below), we have recruited Emre Baspinar in December 2018 for a two-year postdoc. Within MathNeuro, Emre is working on analysing slow-fast dynamical behaviours in the mean-field limit of neural networks.
In a first project, he has been analysing the slow-fast structure in the mean-field limit of a network of FitzHugh-Nagumo neuron models; the mean-field was previously established in [3] but its slow-fast aspect had not been analysed. In particular, he has proved a persistence result of Fenichel type for slow manifolds in this mean-field limit, thus extending previous work by Berglund *et al.* [47], [48]. A manuscript is in preparation.
In a second project, he has been looking at a network of Wilson-Cowan systems whose mean-field limit is an ODE, and he has studied elliptic bursting dynamics in both the network and the limit: its slow-fast dissection, its singular limits and the role of canards. In passing, he has obtained a new characterisation of ellipting bursting via the construction of periodic limit sets using both the slow and the fast singular limits and unravelled a new singular-limit scenario giving rise to elliptic bursting via a new type of torus canard orbits. A manuscript is in preparation.

5.3. Neural fields theory

5.3.1. Next-generation neural field model: The evolution of synchrony within patterns and waves

Participants: Áine Byrne [Center for Neural Science, New York University, USA], Daniele Avitabile [VU Amsterdam, Inria MathNeuro], Stephen Coombes [University of Nottingham, UK].

Neural field models are commonly used to describe wave propagation and bump attractors at a tissue level in the brain. Although motivated by biology, these models are phenomenological in nature. They are built on the assumption that the neural tissue operates in a near synchronous regime, and hence, cannot account for changes in the underlying synchrony of patterns. It is customary to use spiking neural network models when examining within population synchronization. Unfortunately, these high-dimensional models are notoriously hard to obtain insight from. In this paper, we consider a network of θ-neurons, which has recently been shown to admit an exact mean-field description in the absence of a spatial component. We show that the inclusion of space and a realistic synapse model leads to a reduced model that has many of the features of a standard neural field model coupled to a further dynamical equation that describes the evolution of network synchrony. Both Turing instability analysis and numerical continuation software are used to explore the existence and stability of spatiotemporal patterns in the system. In particular, we show that this new model can support states above and beyond those seen in a standard neural field model. These states are typified by structures within bumps and waves showing the dynamic evolution of population synchrony.

This work has been published in Physical Review E and is available as [21].

5.3.2. *The hyperbolic model for edge and texture detection in the primary visual cortex*

Participant: Pascal Chossat [CNRS, Inria MathNeuro].

The modelling of neural fields in the visual cortex involves geometrical structures which describe in mathematical formalism the functional architecture of this cortical area. The case of contour detection and orientation tuning has been extensively studied and has become a paradigm for the mathematical analysis of image processing by the brain. Ten years ago an attempt was made to extend these models by replacing orientation (an angle) with a second-order tensor built from the gradient of the image intensity and named the structure tensor. This assumption does not follow from biological observations (experimental evidence is still lacking) but from the idea that the effectiveness of texture processing with the stucture tensor in computer vision may well be exploited by the brain itself. The drawback is that in this case the geometry is not Euclidean but hyperbolic instead, which complicates substantially the analysis. The purpose of this review is to present the methodology that was developed in a series of papers to investigate this quite unusual problem, specifically from the point of view of tuning and pattern formation. These methods, which rely on bifurcation theory with symmetry in the hyperbolic context, might be of interest for the modelling of other features such as color vision, or other brain functions.

This work has been accepted for publication in Journal of Mathematical Neuroscience and is available as [41].

5.3.3. *A neural field model for color perception unifying assimilation and contrast*

Participants: Anna Song [ENS Paris, France], Olivier Faugeras, Romain Veltz.

We address the question of color-space interactions in the brain, by proposing a neural field model of color perception with spatial context for the visual area V1 of the cortex. Our framework reconciles two opposing perceptual phenomena, known as simultaneous contrast and chromatic assimilation. They have been previously shown to act synergistically, so that at some point in an image, the color seems perceptually more similar to that of adjacent neighbors, while being more dissimilar from that of remote ones. Thus, their combined effects are enhanced in the presence of a spatial pattern, and can be measured as larger shifts in color matching experiments. Our model supposes a hypercolumnar structure coding for colors in V1, and relies on the notion of color opponency introduced by Hering. The connectivity kernel of the neural field exploits the balance between attraction and repulsion in color and physical spaces, so as to reproduce the sign reversal in the influence of neighboring points. The color sensation at a point, defined from a steady state of the neural activities, is then extracted as a nonlinear percept conveyed by an assembly of neurons. It connects the cortical and perceptual levels, because we describe the search for a color match in asymmetric matching experiments as a mathematical projection on color sensations. We validate our color neural field alongside this color matching framework, by performing a multi-parameter regression to data produced by psychophysicists and ourselves. All the results show that we are able to explain the nonlinear behavior of shifts observed along one or two dimensions in color space, which cannot be done using a simple linear model.

This work has been published in PLoS Computational Biology and is available as [31].

5.4. Slow-fast dynamics in Neuroscience

5.4.1. Local theory for spatio-temporal canards and delayed bifurcations

Participants: Daniele Avitabile [VU Amsterdam, Inria MathNeuro], Mathieu Desroches, Romain Veltz, Martin Wechselberger [University of Sydney, Australia].

We present a rigorous framework for the local analysis of canards and slow passages through bifurcations in a wide class of infinite-dimensional dynamical systems with time-scale separation. The framework is applicable to models where an infinite-dimensional dynamical system for the fast variables is coupled to a finite-dimensional dynamical system for slow variables. We prove the existence of centre-manifolds for generic models of this type, and study the reduced, finite-dimensional dynamics near bifurcations of (possibly) patterned steady states in the layer problem. Theoretical results are complemented with detailed examples and numerical simulations covering systems of local- and nonlocal-reaction diffusion equations, neural field models, and delay-differential equations. We provide analytical foundations for numerical observations recently reported in literature, such as spatio-temporal canards and slow-passages through Hopf bifurcations in spatially-extended systems subject to slow parameter variations. We also provide a theoretical analysis of slow passage through a Turing bifurcation in local and nonlocal models.

This work has been submitted for publication and is available as [37].

5.4.2. Pseudo-plateau bursting and mixed-mode oscillations in a model of developing inner hair cells

Participants: Harun Baldemir, Daniele Avitabile [VU Amsterdam, Inria MathNeuro], Krasimira Tsaneva-Atanasova [University of Exeter, UK].

Inner hair cells (IHCs) are excitable sensory cells in the inner ear that encode acoustic information. Before the onset of hearing IHCs fire calcium-based action potentials that trigger transmitter release onto developing spiral ganglion neurones. There is accumulating experimental evidence that these spontaneous firing patterns are associated with maturation of the IHC synapses and hence involved in the development of hearing. The dynamics organising the IHCs' electrical activity are therefore of interest.
Building on our previous modelling work we propose a three-dimensional, reduced IHC model and carry out non-dimensionalisation. We show that there is a significant range of parameter values for which the dynamics of the reduced (three-dimensional) model map well onto the dynamics observed in the original biophysical (four-dimensional) IHC model. By estimating the typical time scales of the variables in the reduced IHC model we demonstrate that this model could be characterised by two fast and one slow or one fast and two slow variables depending on biophysically relevant parameters that control the dynamics. Specifically, we investigate how changes in the conductance of the voltage-gated calcium channels as well as the parameter corresponding to the fraction of free cytosolic calcium concentration in the model affect the oscillatory model bahaviour leading to transition from pseudo-plateau bursting to mixed-mode oscillations. Hence, using fast-slow analysis we are able to further our understanding of this model and reveal a path in the parameter space connecting pseudo-plateau bursting and mixed-mode oscillations by varying a single parameter in the model.

This work has been accepted for publication in Communications in Nonlinear Science and Numerical Simulation and is available as [18].

5.4.3. Parabolic bursting, spike-adding, dips and slices in a minimal model

Participants: Mathieu Desroches, Jean-Pierre Françoise [LJLL, Sorbonne Université, Paris], Martin Krupa [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 one-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 published in Mathematical Modelling of Natural Phenomena and is available as [26].

5.4.4. Anticipation via canards in excitable systems

Participants: Elif Köksal Ersöz [INSERM, Rennes, Inria MathNeuro], 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 that is 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 and receiver follow the evolution of a multi-scale excitable system. We present a novel theoretical viewpoint 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-dimensional 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 by a wider class of excitable systems with appropriate theoretical extensions.

This work has been published in Chaos: An Interdisciplinary Journal of Nonlinear Science and is available as [28].

5.4.5. Canard-induced complex oscillations in an excitatory network

Participants: Elif Köksal Ersöz, Mathieu Desroches, Antoni Guillamon [Polytechnic University of Catalunya, Spain], John Rinzel [Center for Neural Science and Courant Institute of Mathematical Sciences, New York University, USA], 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 [69]. The dynamics is that of a classical square-wave burster, with alternation of silent and active phases. Tabak et al. [69] 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 [66], [67], [68] 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 [69] 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 [10]. Finally, we related 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 [68].

This work has been submitted for publication and is available as [12].

5.5. Mathematical modeling of neuronal excitability

5.5.1. Modeling cortical spreading depression induced by the hyperactivity of interneurons

Participants: Mathieu Desroches, Olivier Faugeras, Martin Krupa [LJAD, UCA, Inria MathNeuro], Massimo Mantegazza [Institut de Pharmacologie Moléculaire et Cellulaire (IPMC), Sophia Antipolis].

Cortical spreading depression (CSD) is a wave of transient intense neuronal firing leading to a long lasting depolarizing block of neuronal activity. It is a proposed pathological mechanism of migraine with aura. Some forms of migraine are associated with a genetic mutation of the $Na_{v1.1}$ channel, resulting in its gain of function and implying hyperexcitability of interneurons. This leads to the counterintuitive hypothesis that intense firing of interneurons can cause CSD ignition. To test this hypothesis in silico, we developed a computational model of an E-I pair (a pyramidal cell and an interneuron), in which the coupling between the cells in not just synaptic, but takes into account also the effects of the accumulation of extracellular potassium caused by the activity of the neurons and of the synapses. In the context of this model, we show that the intense firing of the interneuron can lead to CSD. We have investigated the effect of various biophysical parameters on the transition to CSD, including the levels of glutamate or GABA, frequency of the interneuron firing and the efficacy of the KCC_2 co-transporter. The key element for CSD ignition in our model was the frequency of interneuron firing and the related accumulation of extracellular potassium, which induced a depolarizing block of the pyramidal cell. This constitutes a new mechanism of CSD ignition.

This work has been published in Journal of Computational Neuroscience and is available as [25].

The extension of this work is the topic of the PhD of Louisiane Lemaire, who started in October 2018. A first part of Louisiane's PhD has been to improve and extend the model published in [25] in a number of ways: replace the GABAergic neuron model used in [25], namely the Wang-Buszáki model, by a more recent fast-spiking cortical interneuron model due to Golomb and collaborators; implement the effect of the HM1a toxin used by M. Mantegazza to mimic the genetic mutation of sodium channels responsible for the hyperactivity of the GABAergic neurons; take into account ionic concentration dynamics (relaxing the hypothesis of constant reversal potentials) for the GABAergic as well whereas in [25] this was done only for the Pyramidal neuron. This required a great deal of modelling and calibration and the simulation results are closer to the actual experiments by Mantegazza than in our previous study. A manuscript is in preparation.

5.6. Modelling the visual system

5.6.1. Uniqueness of viscosity mean curvature flow solution in two sub-Riemannian structures

Participants: Emre Baspinar, Giovanna Citti [University of Bologna, Italy].

We provide a uniqueness result for a class of viscosity solutions to sub-Riemannian mean curvature flows. In a sub-Riemannian setting, uniqueness cannot be deduced by the comparison principle, which is known only for graphs and for radially symmetry surfaces. Here we use a definition of continuous viscosity solutions of sub-Riemannian mean curvature flows motivated from a regularized Riemannian approximation of the flow. With this definition, we prove that any continuous viscosity solution of the equation is a limit of a sequence of solutions of Riemannian flow and obtain as a consequence uniqueness and the comparison principle. The results are provided in the settings of both 3-dimensional rototranslation group $SE(2)$ and Carnot groups of step 2, which are particularly important due to their relation to the surface completion problem of a model of the visual cortex.

This work has been published in SIAM Journal on Mathematical Analysis and is available as [19].

5.6.2. A sub-Riemannian model of the visual cortex with frequency and phase

Participants: Emre Baspinar, Alessandro Sarti [CAMS, EHESS, Paris, France], Giovanna Citti [University of Bologna, Italy].

In this paper we present a novel model of the primary visual cortex (V1) based on orientation, frequency and phase selective behavior of the V1 simple cells. We start from the first level mechanisms of visual perception: receptive profiles. The model interprets V1 as a fiber bundle over the 2-dimensional retinal plane by introducing orientation, frequency and phase as intrinsic variables. Each receptive profile on the fiber is mathematically interpreted as a rotated, frequency modulated and phase shifted Gabor function. We start from the Gabor function and show that it induces in a natural way the model geometry and the associated horizontal connectivity modeling the neural connectivity patterns in V1. We provide an image enhancement algorithm employing the model framework. The algorithm is capable of exploiting not only orientation but

also frequency and phase information existing intrinsically in a 2-dimensional input image. We provide the experimental results corresponding to the enhancement algorithm.

This work has been submitted for publication and is available as [39].

MIMESIS Team

7. New Results

7.1. Real-time simulation of hyperelastic materials using Deep Learning

Participants: Andrea Mendizabal, Pablo Márquez-Neila, Stéphane Cotin.

The finite element method (FEM) is among the most commonly used numerical methods for solving engineering problems. Due to its computational cost, various ideas have been introduced to reduce computation times, such as domain decomposition, parallel computing, adaptive meshing, and model order reduction. In this work we propose the U-Mesh: a data-driven method based on a U-Net architecture that approximates the non-linear relation between a contact force and the displacement field computed by FE algorithm. We show that deep learning, one of the latest machine learning methods based on artificial neural networks, can enhance computational mechanics through its ability to encode highly non-linear models in a compact form. Our method is applied to three benchmark examples: a cantilever beam, an L-shape and a liver model subject to moving punctual loads. A comparison between our method and proper orthogonal decomposition (POD) is done. The results show that U-Mesh can perform very fast simulations on various geometries and topologies, mesh resolutions and number of input forces with very small errors. results were published in the Journal of Medical Image Analysis [23] (impact factor 8.5).

7.2. FEM-based confidence assessment of non-rigid registration

Participants: Paul Baksic, Hadrien Courtecuisse, Matthieu Chabanas, Bernard Bayle.

Non-rigid registration is often used for 3D representations during surgical procedures. It needs to provide good precision in order to guide the surgeon properly. We proposed in [25] a method that allows the computation of a local upper bound of the registration confidence over the whole organ volume. Using a bio-mechanical model, we apply tearing forces over the whole organ to compute the upper bound of the degrees of freedom left by the registrations constraints. Confrontation of our method with experimental data shows promising results to estimate the registration confidence. Indeed, the computed maximum error appears to be a real upper bound(see figure 4). A more advanced method was submitted at IPCAI 2020.

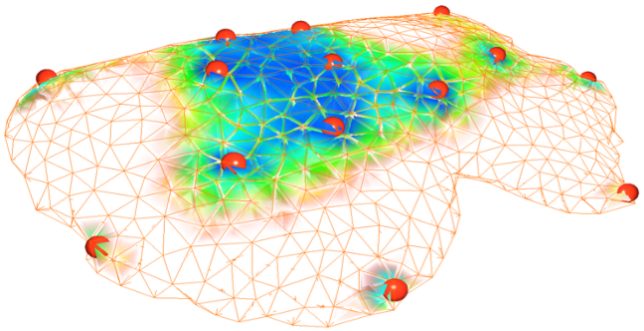


Figure 4. This is an example of confidence map given by our method on a registration of a lamb liver. The red dots are the registration constraint given by sensors. High confidence area are presented in blue. The area where the confidence is below the one needed for the surgery appears transparent.

7.3. Physics-based Deep Neural Network for Augmented Reality

Participants: Jean-Nicolas Brunet, Andrea Mendizabal, Antoine Petit, Nicolas Golse, Eric Vibert, Stéphane Cotin.

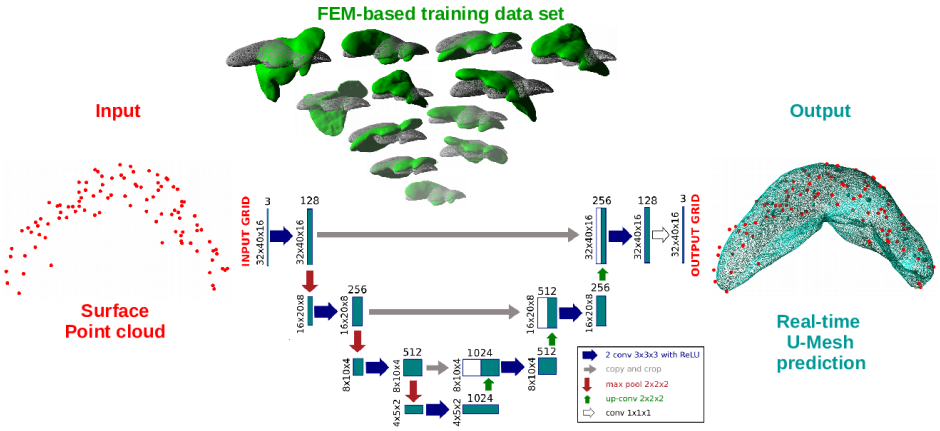


Figure 5. The U-Mesh framework allows for extremely fast simulations of soft tissues accounting for large non linear deformations.

We propose an approach combining a finite element method and a deep neural network to learn complex elastic deformations with the objective of providing augmented reality during hepatic surgery. Derived from the U-Net architecture, our network is built entirely from physics-based simulations of a preoperative segmentation of the organ (see figure 5). These simulations are performed using an immersed-boundary method, which offers several numerical and practical benefits, such as not requiring boundary-conforming volume elements. We perform a quantitative assessment of the method using synthetic and *ex vivo* patient data. Results show that the network is capable of solving the deformed state of the organ using only a sparse partial surface displacement data and achieve similar accuracy as a FEM solution, while being about 100x faster. When applied to an *ex vivo* liver example, we achieve the registration in only $3\,ms$ with a mean target registration error (TRE) of $2.9\,mm$. This results were presented at MICCAI 2019 [22].

7.4. Estimation of boundary conditions for patient-specific liver simulation during augmented surgery

Participants: Sergei Nikolaev, Stéphane Cotin.

Augmented liver surgery is an active research area that aims at improving the surgical outcome by enhancing the view of internal structures. However, to precisely estimate the position of these, an accurate model of the liver is required. While researchers have focused on proposing new models, algorithms for real-time computation or estimation of the tissue properties, very few have addressed the question of boundary conditions. Yet, they play a key role in the computation of the deformation. Boundary conditions mainly result from ligaments connecting the liver to its surrounding anatomy and limiting its motion. Unfortunately, ligaments' shapes and properties cannot be identified with preoperative imaging. We propose to estimate both the location and stiffness of ligaments by using a combination of a statistical atlas, numerical simulation, and Bayesian inference (fig. 6). Ligaments are modeled as polynomial springs connected to a liver finite element model. Their original location on the liver is based on an anatomical atlas, and their initial stiffness is taken

from the literature. These characteristics are then corrected using a reduced order unscented Kalman filter based on observations taken from the laparoscopic image stream. Our approach is evaluated using synthetic data and phantom data. Results show that our estimation of the boundary conditions improves the accuracy of the simulation by 75% when compared to typical methods involving Dirichlet boundary conditions. The results were submitted for a presentation at IPCAI 2020

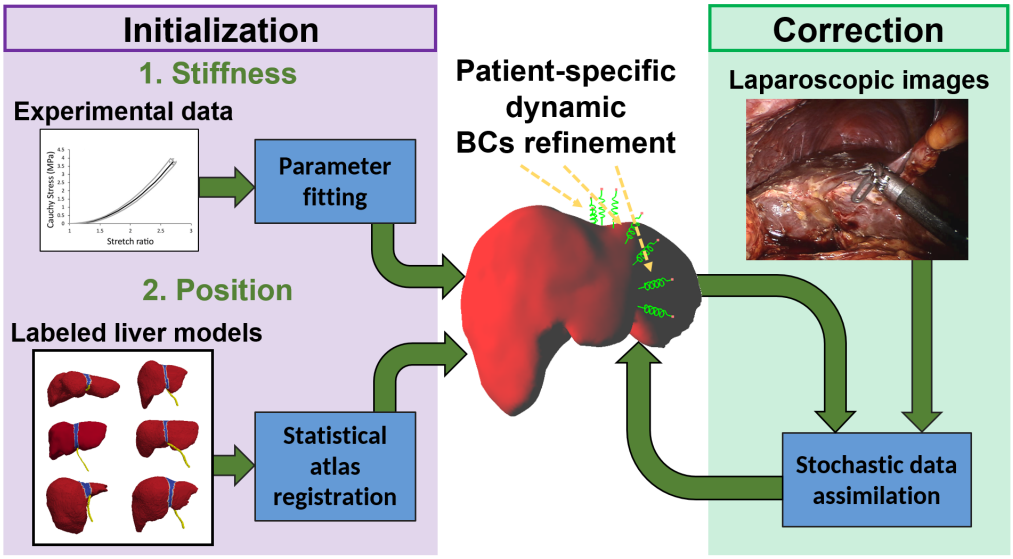


Figure 6. Overview of the boundary condition identification process. It contains two main steps. 1 - Initial approximation based on statistics from the processed model database and experimental data. 2 - Identification based on intraoperative patient-specific images.

7.5. Corotated meshless implicit dynamics for deformable bodies

Participants: Jean-Nicolas Brunet, Vincent Magnoux, Benoît Ozell, Stéphane Cotin.

We proposed a fast, stable and accurate meshless method to simulate geometrically non-linear elastic behaviors. To address the inherent limitations of finite element (FE) models, the discretization of the domain is simplified by removing the need to create polyhedral elements. The volumetric locking effect exhibited by incompressible materials in some linear FE models is also completely avoided. Our approach merely requires that the volume of the object be filled with a cloud of points (see figure 7). To minimize numerical errors, we constructed a corotational formulation around the quadrature positions that is well suited for large displacements containing small deformations. The equations of motion was integrated in time following an implicit scheme. The convergence rate and accuracy were validated through both stretching and bending case studies. Finally, results were presented using a set of examples that show how we can easily build a realistic physical model of various deformable bodies with little effort spent on the discretization of the domain. We presented our work at WSCG 2019 [21]. (Fig. 7).

7.6. The effect of discretization on parameter identification. Application to patient-specific simulations

Participants: Nava Schulmann, Igor Peterlik, Stéphane Cotin.

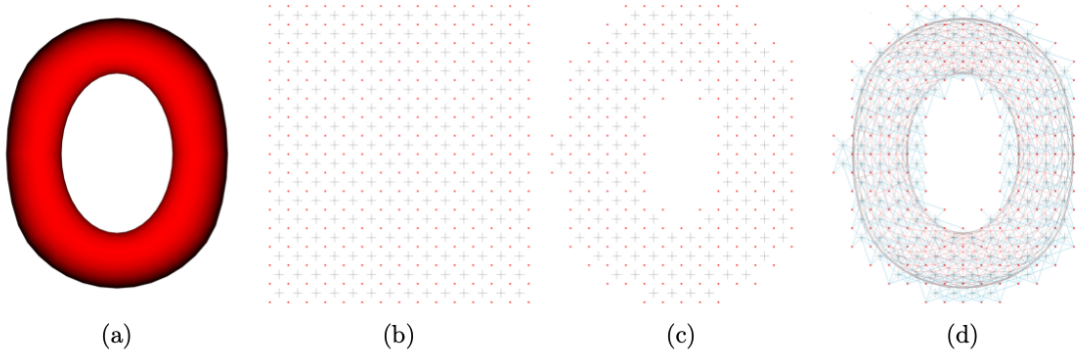


Figure 7. Volumetric discretizations of a 3D surface. (a) Surface mesh provided by the user. (b) Background grid where the grid's cubes are used to place the DOFs(degrees of freedom) and the integration points. (c) DOFs and integration points are cropped to fit the surface mesh. (d) A neighborhood of the closest particles is built around each integration point.

Identifying the elastic parameters of a finite element model from a dynamically acquired set of observations is a fundamental challenge in many data-driven medical applications, from soft surgical robotics to image-guided per-operative simulations. While various strategies exist to tackle the parameter-identification inverse problem [29], the effect of sub-optimal discretization, as often required in real-time applications, is largely overlooked. Indeed, the need to tune the parameter values in order to account for discretization-induced stiffening in specific models is reported in different works (e.g. [Chen et al., 2015, Anna et al., 2018]). However, to the best of our knowledge, no systematic study of this phenomenon exists to date, nor has any strategy to select optimal effective values been developed. Our work addresses the issue of parameter identification in coarsened meshes with special attention to the dynamical nature of the identification. We focus on the estimation of Young's moduli in simplified systems and show that the estimated stiffnesses are underestimated in a systematic manner when reducing the number of degrees of freedom. We also show that the effective stiffness of a given coarse mesh, when associated with an undersampled mesh discretization, is not constant but strongly depends on the prescribed deformations. These results show that the estimated parameters should not be considered as the true parameter value of the organ or tissue but instead are model-dependent values. We argue that Bayesian methods present a clear advantage w.r.t. classical minimization methods by their ability to efficiently adapt the local parameter values. The results were presented at CMBBE 2019 [26].

7.7. Elastic registration based on biomechanical graph matching

Participants: Jaime Garcia Guevara, Igor Peterlik, Marie-Odile Berger, Stéphane Cotin.

An automatic elastic registration method suited for vascularized organs is proposed. The vasculature in both the preoperative and intra-operative images is represented as a graph. A typical application of this method is the fusion of pre-operative information onto the organ during surgery, to compensate for the limited details provided by the intra-operative imaging modality (e.g. CBCT) and to cope with changes in the shape of the organ. Due differences in image modalities and organ deformation, each graph has a different topology and shape. The Adaptive Compliance Graph Matching (ACGM) method presented does not require any manual initialization, handles intra-operative nonrigid deformations of up to 65 mm and computes a complete displacement field over the organ from only the matched vasculature. ACGM is better than the previous Biomechanical Graph Matching method [3] (BGM) because it uses an efficient biomechanical vascularized liver model to compute the organ's transformation and compliance of vessel bifurcations. It allows to efficiently find the best graph matches with a novel compliance-based adaptive search. These contributions are evaluated

on ten realistic synthetic and two real porcine automatically segmented datasets. ACGM obtains better target registration error (TRE) than BGM, with an average TRE in the real datasets of 4.2 mm compared to 6.5 mm, respectively. It also is up to one order of magnitude faster, less dependent on the parameters used and more robust to noise. Figure 8 depicts the large deformation and the registered porcine CBCT and CTA data. Results were published in Annals of Biomedical Engineering (2019) [4].

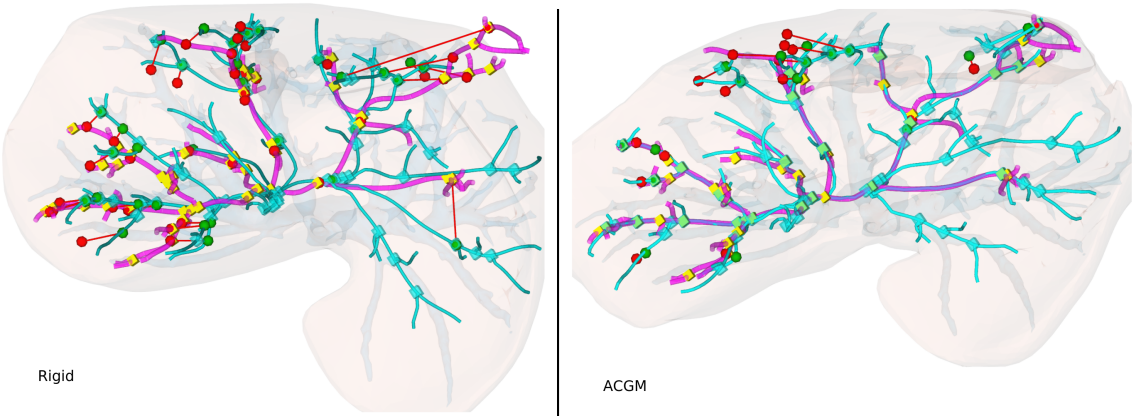


Figure 8. Registration between CTA and CBCT images. The target CBCT (in pink) and source CTA (in cyan) portal vein graphs are rendered with tubular structures. The graph nodes (bifurcations) are shown as cubic markers (in yellow for the target, cyan for the source and green for the matched). The augmented hepatic vein, which was only visible in the CTA image, is in transparent blue behind the portal veins graphs. The 37 target evaluation landmarks (red spheres) and their corresponding connected source landmarks (green spheres) and the liver structures are rigidly aligned and show the large intra-operative deformation (left image). The result of the registration process (right) shows the 16 registered landmarks.

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 and rule manipulation. 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.

Finally, we study the links between our bio-inspired modeling work and other domains like Machine Learning, computer science and educational science (*cf.* § 7.6).

7.2. The limbic loop

Our main contribution this year to advancing our view of the limbic loop is the defense of the PhD of B. T. Nallapu [1] related to the modeling of the orbital and medial loops the two main constituents of the limbic loop, as described in [33]. In short, this work proposes, instead of a global view of the orbital loop as generally proposed, a view with two loops, one corresponding to the lateral part of the orbitofrontal cortex related to respondent conditioning and one to the medial part related to operant conditioning and closely linked to the medial loop. The work shows that such a model replicates a variety of observations in neuroscience and can be used for the autonomous behavior of an agent in the Minecraft video game.

7.3. The associative loop

The prefrontal cortex is known to be involved in many high-level cognitive functions, in particular working memory. Gated working memory is defined as the capacity of holding arbitrary information at any time in order to be used at a later time. Based on electrophysiological recordings, several computational models have tackled the problem using dedicated and explicit mechanisms. We propose instead to consider an implicit mechanism based on a random recurrent neural network. We introduce a robust yet simple reservoir model of gated working memory with instantaneous updates [11]. The model is able to store an arbitrary real value at random time over an extended period of time. The dynamics of the model is a line attractor that learns to exploit reentry and a non-linearity during the training phase using only a few representative values. A deeper study of the model shows that there is actually a large range of hyper parameters for which the results hold (number of neurons, sparsity, global weight scaling, etc.) such that any large enough population, mixing excitatory and inhibitory neurons can quickly learn to realize such gated working memory. This suggests this property could be an implicit property of any random population, that can be acquired through learning. Furthermore, considering working memory to be a physically open but functionally closed system, we give account on some counter-intuitive electrophysiological recordings.

We also developed a model of working memory combining short-term and long-term components [24]. For the long-term component, we used Conceptors in order to store constant temporal patterns. For the short-term component , we used the Gated-Reservoir model [11]. We combined both components in order to obtain a model in which information can go from long-term memory to short-term memory and vice-versa.

The prefrontal cortex is also known to be the place where complex and abstract behavioral rules are implemented. In order to study the mechanisms related to the manipulation of such rules, we have begun the study of networks able to build rules to manipulate such framework as the Wisconsin Card Sorting Test, widely used in the clinical domain [3].

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 examplified by song learning in birds or human complex vocalizations[35]. Sensorimotor learning represents a challenging problem for natural and artificial systems. Several computational models have been proposed to explain the neural and cognitive mechanisms at play in the brain [34]. In general, these models can be decomposed in three common components: a sensory system, a motor control device and a learning framework. The latter includes the architecture, the learning rule or optimisation method, and the exploration strategy used to guide learning. In this review, we focus on imitative vocal learning, that is exemplified in song learning in birds and speech acquisition in humans. We aim to synthesise, analyse and compare the various models of vocal learning that have been proposed, highlighting their common points and differences. We first introduce the biological context, including the behavioural and physiological hallmarks of vocal learning and sketch the neural circuits involved. Then, we detail the different components of a vocal learning model and how they are implemented in the reviewed models.

On this topic, X. Hinaut is also collaborating with Catherine del Negro's team (CNRS, NeuroPSI, 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 transcription and analyses has been done by M1 intern (in Neuroscience) Juliette Giraudon, and preliminary work on song segmentation and classification has been done by L3 intern (ENS) Pierre Marcus. In December, Aurore Cazala (student in the NeuroPSI collaborator team since 2014) defended her PhD at the University of Paris-Saclay on "Codage neuronal de l'ordre des signaux acoustiques dans le chant des oiseaux chanteurs"; X. Hinaut participated in studies done in this PhD.

7.5. Systemic integration

Several global approaches corresponding to the systemic integration of several loops have been studied this year. The PhD work [1] evoked in section § 7.2 for its contributions to modeling the limbic loop, has also been partly devoted to the definition of a global cognitive model and its embodiment in an agent in the Minecraft game, in order to illustrate the performances of autonomous behavior of a system endowed by such a limbic loop [33].

7.6. Association to other scientific domains

Concerning Machine Learning and 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 parser (ResPars) with the representational system of Spranger: IRL (Incremental Recruitment Language) [17]. 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 [5]. 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. Some people can mix two languages within the same sentence: this is known as intra-sentential code-switching. With M1 intern Pauline Detraz, we collected data from human subjects that were required to mix pairs of given sentences in French and English. The corpus obtained have some very complex mixed sentences: there can be until eleven language switches within the same sentence. Then, we trained our Reservoir-based sentence Parsing model, with the collected corpus.

Surprisingly the model is able to learn and generalize on the mixed corpus with performances nearly as good as the unmixed French-English corpus [16]. A post-doc joined the team in Nov 2019 to work on the project HuRRiCane ("Hierarchical Reservoir Computing for Language Comprehension") project founded by Inria. This project aims at extending the ResPars model to work from speech inputs to sentence comprehension including coherancy checking. In other words, the objective is to experiment how a sentence comprehension model, based on reservoir computing, can learn to understand sentences by exploring which meanings can have the sentences, implying several steps from stream of phonemes to words and from stream of words to sentence comprehension. The model will be implemented on a virtual agent first and then on the Nao humanoid robot.

This project is linked to other projects in the team on the hierarchical organization of the prefrontal cortex (including Broca's area, involved in language). This hierarchy corresponds to an increasingly higher abstraction, which is made by different sub-areas. We will therefore be able to link this post-doc project to existing projects of the team, where different levels of abstractions are necessary for sentence comprehension.

As explained in § 7.6 , song segmentation and classification has been done by L3 intern (ENS) Pierre Marcus.

The on-going work on an original prototype based approach of deep-learning considering not so big data sets, and targeting also interpretability of the result, has been finalized [4], including a fine study on metaparameter adjustment in this context, while both standard learning and meta-learning paradigms have been considered. The capability to easily the "how it works" mechanism to no specialist of the field is an important outcome of the paper.

Co-leaded by Margarida Romero scientific director of the LINE laboratory of the UCA and researchers of our team, a preliminary work regarding artificial intelligence devoted to education (AIDE) was developed to study applications to educational science. This first year has been devoted to study to which extents the existing collaboration between Inria science outreach regarding computational thinking initiation and educational science research in order to be understand the underlying cognitive processes of the former actions and evaluate them, could be enlarged to multi-disciplinary research in both fields. The first outcome of this collaboration has been an analysis of a computational thinking unplugged activity under the perspective of embodied cognition [9] and deep and large review in the field, analyzing how computational thinking in K-12 education could be developed [22], within the scope of studies regarding co-creativity, robotics and maker education. [20], while a qualitative analysis one very large audience (more then 18000 inscriptions) on-line course outcomes has been published [18], with some operational outcomes regarding enlarging comptational thinking training from teachers to all citizens [12].

Software is a fundamental pillar of modern scientific research, across all fields and disciplines. However, there is a lack of adequate means to cite and reference software due to the complexity of the problem in terms of authorship, roles and credits. This complexity is further increased when it is considered over the lifetime of a software that can span up to several decades. Building upon the internal experience of Inria, the French research institute for digital sciences, we provide in this paper a contribution to the ongoing efforts in order to develop proper guidelines and recommendations for software citation and reference. Namely, we recommend: (1) a richer taxonomy for software contributions with a qualitative scale; (2) to put humans at the heart of the evaluation; and (3) to distinguish citation from reference.

NEUROSYS Project-Team

7. New Results

7.1. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Amélie Aussel, Nathalie Azevedo Carvalho, Dominique Martinez (CNRS), Radu Ranta (Univ. Lorraine, CRAN).
In collaboration with Harry Tran (Univ. Lorraine, CRAN), Louise Tyvaert (Univ. Lorraine, CRAN, CHRU Nancy), Olivier Aron (Univ. Lorraine, CRAN, CHRU Nancy), Sylvain Contassot-Vivier (Univ. Lorraine),

7.1.1. Hippocampal oscillatory activity

7.1.1.1. Healthy hippocampus

We proposed a detailed anatomical and mathematical model of the hippocampal formation for the generation of healthy hippocampal activity, especially sharp-wave ripples and theta-nested gamma oscillations [24], [25]. 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 [24].

We further extended this work with an extensive study of the parameter range of the healthy hippocampus activity and showed that the "healthy model" was unable to reproduce pathological hippocampal oscillations observed in temporal lobe epilepsy.

7.1.1.2. Modeling LFP measures

The development of this model was also the opportunity to extend our model of the measure of the local field potential (LFP) and to study the contribution of spikes (not only synaptic currents) to the generation of the LFP. Indeed, simulating extracellular recordings of neuronal populations is a challenging task for understanding the nature of extracellular field potentials (LFPs), investigating specific brain structures and mapping cognitive functions. In general, it is assumed that extracellular recording devices (micro and/or macro-electrodes) record a mixture of low frequency patterns, mainly attributed to the synaptic currents and high-frequency components reflecting action potential (APs) activity. Simulating such signals often requires a high computational burden due to the multicompartmental neuron models used. Therefore, different LFP proxies coexist in the literature, most of them only reproducing some of the features of experimental signals. This may be an issue in producing and validating computational models of phenomena where the fast and slow components of neural activity are equally important, such as hippocampal oscillations. In this part of the work, we proposed an original approach for simulating large-scale neural networks efficiently while computing a realistic approximation of the LFP signal including extracellular signatures of both synaptic and action potentials [26]. We applied this method on the hippocampal network we developed earlier and compared the simulated signal with intracranial measurements from human patients.

7.1.1.3. Epilepsy of the mesial temporal lobe

The model described above has then been extended to include pathological changes observed in temporal lobe epilepsy, the future goal being to better understand the generation and propagation of epileptic activity throughout the brain, and therefore to investigate new potential therapeutic targets.

The mechanisms underlying the generation of hippocampal epileptic seizures and interictal events during the sleep-wake cycle are not yet fully understood. In this article, based on our previous computational modeling work of the hippocampal formation based on realistic topology and synaptic connectivity, we study the role of network specificity and channel pathological conditions of the epileptic hippocampus in the generation and maintenance of seizures and interictal oscillations. Indeed, the epilepsies of the mesial temporal lobe are associated with hippocampal neuronal and axonal loss, mossy fiber sprouting and channelopathies, namely impaired potassium and chlore dynamics. We show, through the simulations of hippocampal activity during slow-wave sleep and wakefulness that: (i) both mossy fiber sprouting and sclerosis account for epileptic seizures, (ii) high hippocampal sclerosis with low sprouting suppresses seizures, (iii) impaired potassium and chloride dynamics have little influence on the generation of seizures, (iv) but do have an influence on interictal spikes that decreases with high mossy fiber sprouting. A manuscript is in preparation for the Journal of Neuroscience.

7.1.2. Synchronization phenomena in neuronal network models

From a more computational point of view, we got interested in interneuronal gamma oscillations and synchronization in hippocampus-like networks via different models, especially in adaptive exponential integrate-and-fire neurons. Fast neuronal oscillations in gamma frequencies are observed in neocortex and hippocampus during essential arousal behaviors. Through a four-variable Hodgkin–Huxley type model, Wang and Buzsáki have numerically demonstrated that such rhythmic activity can emerge from a random network of GABAergic interneurons via minimum synaptic inputs. In this case, the intrinsic neuronal characteristics and network structure act as the main drive of the rhythm. We investigate inhibitory network synchrony with a low complexity, two-variable adaptive exponential integrate-and-fire (AdEx) model, whose parameters possess strong physiological relevances, and provide a comparison with the two-variable Izhikevich model and Morris–Lecar model. Despite the simplicity of these three models, the AdEx model shares two important results with the previous biophysically detailed Hodgkin–Huxley type model: the minimum number of synaptic inputs necessary to initiate network gamma-band rhythms remains the same, and this number is weakly dependent on the network size. Meanwhile, Izhikevich and Morris–Lecar neurons demonstrate different results in this study. We further investigated the necessary neuronal, synaptic and connectivity properties, including gap junctions and shunting inhibitions, for AdEx model leading to sparse and random network synchrony in gamma rhythms and nested theta gamma rhythms. These findings suggest a computationally more tractable framework for studying synchronized networks in inducing cerebral gamma band activities.

7.1.3. Event-driven simulation of large scale neural models with on-demand connectivity generation

Accurate simulations of brain structures is a major problem in neuroscience. Many works are dedicated to design better models or to develop more efficient simulation schemes. In this work, we propose to combine time-stepping numerical integration of Hodgkin-Huxley type neurons with event-driven updating of the synaptic currents. A spike detection method was also developed to determine the spike time more precisely in order to preserve the second-order Runge-Kutta methods. This hybrid approach allows us to regenerate the outgoing connections at each event, thereby avoiding the storage of the connectivity. Consequently, memory consumption and execution time are significantly reduced while preserving accurate simulations, especially spike times of detailed point neuron models. The efficiency of the method has been demonstrated on the simulation of 10^6 interconnected MSN neurons with Parkinson disease (an article has been submitted to Frontiers in Neuroinformatics) [23].

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Sébastien Rimbert, Oleksii Avilov, Rahaf Al-Chwa, Anais Coster, Elina Ortega Herrera, Nicolas Rault, Radu Ranta (univ. Lorraine).
In collaboration with Stéphanie Fleck (Univ. Lorraine)

7.2.1. On source space resolution in EEG brain imaging for motor imagery

Brain source localization accuracy is known to be dependent on the EEG sensor placement over the head surface. In Brain-Computer Interfaces (BCI), according to the paradigm used, Motor Imagery (MI) and Steady-State Visual Evoked Potential (SSVEP) in particular, electrodes are not well distributed over the head, and their number is not standardized as in classical clinical applications. We proposed a method for quantifying the expected quality of source localization with respect of the sensor placement, known as EEG montage. Our method, based on a subspace correlation metric, can be used to assess which brain sources can be distinguished (as they generate sufficiently different potentials on the electrodes), and also to identify regions/volumes in which precise source localization is impossible (i.e. all sources inside those regions could generate similar electrode potentials). In particular, for a MI dedicated montage, we show that source localization is less precise than for standard montages, although the local density of electrodes over the areas of interest is higher [13].

7.2.2. 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, 0.2 to 1.3% 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 reduce 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 [15], [7].

Moreover the way in which propofol (i.e., an anesthetic commonly used for the general anesthesia induction) affects motor brain activity within the electroencephalographic (EEG) signal has been poorly investigated and is not clearly understood. For this reason, a detailed study of the motor activity behavior with a step-wise increasing dose of propofol is required and would provide a proof of concept for such an innovative BCI. We started a study to highlight the occurrence of movement attempt patterns, mainly changes in oscillations called event-related desynchronization (ERD) and event-related synchronization (ERS), in the EEG signal over the motor cortex, in healthy subjects, without and under propofol sedation, during four different motor tasks [8], [12].

7.2.3. 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. We aimed 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 revealed that BCI performance is correlated to habits and frequency of practicing manual activities [6].

7.2.3.1. Hypnotic State Modulates Sensorimotor Beta Rhythms During Real Movement and Motor Imagery

Hypnosis techniques are currently used in the medical field and directly influence the patient's state of relaxation, perception of the body, and its visual imagination. There is evidence to suggest that a hypnotic state may help patients to better achieve tasks of motor imagination, which is central in the rehabilitation protocols after a stroke. However, the hypnosis techniques could also alter activity in the motor cortex. To the best of our knowledge, the impact of hypnosis on the EEG signal during a movement or an imagined movement is poorly investigated. In particular, how event-related desynchronization (ERD) and event-related synchronization (ERS) patterns would be modulated for different motor tasks may provide a better understanding of the potential benefits of hypnosis for stroke rehabilitation. To investigate this purpose, we recorded EEG signals from 23 healthy volunteers who performed real movements and motor imageries in a closed eye condition. Our results suggest that the state of hypnosis changes the sensorimotor beta rhythm during the ERD phase but maintains the ERS phase in the mu and beta frequency band, suggesting a different activation of the motor cortex in a hypnotized state [14], [9].

OPIS Project-Team

7. New Results

7.1. General risk measures for robust machine learning

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Henri Gérard, ENPC, Paris)
A wide array of machine learning problems are formulated as the minimization of the expectation of a convex loss function on some parameter space. Since the probability distribution of the data of interest is usually unknown, it is is often estimated from training sets, which may lead to poor out-of-sample performance. In the work [11], we bring new insights in this problem by using the framework which has been developed in quantitative finance for risk measures. We show that the original min-max problem can be recast as a convex minimization problem under suitable assumptions. We discuss several important examples of robust formulations, in particular by defining ambiguity sets based on φ-divergences and the Wasserstein metric. We also propose an efficient algorithm for solving the corresponding convex optimization problems involving complex convex constraints. Through simulation examples, we demonstrate that this algorithm scales well on real data sets.

7.2. Deep Latent Factor Model for Collaborative Filtering

Participants: Emilie Chouzenoux (Collaboration: Aanchal Mongia, Neha Jhamb, Angshul Majumdar, IIIT Delhi, India)
Latent factor models have been used widely in collaborative filtering based recommender systems. In recent years, deep learning has been successful in solving a wide variety of machine learning problems. Motivated by the success of deep learning, we propose in [44], [23] a deeper version of latent factor model. Experiments on benchmark datasets shows that our proposed technique significantly outperforms all stateof-the-art collaborative filtering techniques.

7.3. A Proximal Interior Point Algorithm with Applications to Image Processing

Participants: Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet
In the work [10], we introduce a new proximal interior point algorithm (PIPA). This algorithm is able to handle convex optimization problems involving various constraints where the objective function is the sum of a Lipschitz differentiable term and a possibly nonsmooth one. Each iteration of PIPA involves the minimization of a merit function evaluated for decaying values of a logarithmic barrier parameter. This inner minimization is performed thanks to a finite number of subiterations of a variable metric forward-backward method employing a line search strategy. The convergence of this latter step as well as the convergence the global method itself are analyzed. The numerical efficiency of the proposed approach is demonstrated in two image processing applications.

7.4. Deep Unfolding of a Proximal Interior Point Method for Image Restoration

Participants: Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet (Collaboration: Carla Bertocchi, Marco Prato, Universita di Modena, Italy)

Variational methods are widely applied to ill-posed inverse problems for they have the ability to embed prior knowledge about the solution. However, the level of performance of these methods significantly depends on a set of parameters, which can be estimated through computationally expensive and time consuming methods. In contrast, deep learning offers very generic and efficient architectures, at the expense of explainability, since it is often used as a black-box, without any fine control over its output. Deep unfolding provides a convenient approach to combine variational-based and deep learning approaches. Starting from a variational formulation for image restoration, we developped in [36], [5], iRestNet, a neural network architecture obtained by unfolding a proximal interior point algorithm. Hard constraints, encoding desirable properties for the restored image, are incorporated into the network thanks to a logarithmic barrier, while the barrier parameter, the stepsize, and the penalization weight are learned by the network. We derive explicit expressions for the gradient of the proximity operator for various choices of constraints, which allows training iRestNet with gradient descent and backpropagation. In addition, we provide theoretical results regarding the stability of the network for a common inverse problem example. Numerical experiments on image deblurring problems show that the proposed approach compares favorably with both state-of-the-art variational and machine learning methods in terms of image quality.

7.5. Preconditioned P-ULA for Joint Deconvolution-Segmentation of Ultrasound Images

Participants: Emilie Chouzenoux, Marie-Caroline Corbineau, Jean-Christophe Pesquet (Collaboration: Denis Kouamé, Jean-Yves Tourneret, IRIT, Toulouse)
Joint deconvolution and segmentation of ultrasound images is a challenging problem in medical imaging. By adopting a hierarchical Bayesian model, we propose in [15] an accelerated Markov chain Monte Carlo scheme where the tissue reflectivity function is sampled thanks to a recently introduced proximal unadjusted Langevin algorithm. This new approach is combined with a forward-backward step and a preconditioning strategy to accelerate the convergence, and with a method based on the majorization-minimization principle to solve the inner nonconvex minimization problems. As demonstrated in numerical experiments conducted on both simulated and in vivo ultrasound images, the proposed method provides high-quality restoration and segmentation results and is up to six times faster than an existing Hamiltonian Monte Carlo method.

7.6. A Random Block-Coordinate Douglas-Rachford Splitting Method with Low Computational Complexity for Binary Logistic Regression

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Giovanni Chierchia, ESIEE Paris, Luis Bricenos Arias, Universidad Técnica Federico Santa Maria, Valparaiso, Chile)
In the paper [6], we proposed a new optimization algorithm for sparse logistic regression based on a stochastic version of the Douglas-Rachford splitting method. Our algorithm sweeps the training set by randomly selecting a mini-batch of data at each iteration, and it allows us to update the variables in a block coordinate manner. Our approach leverages the proximity operator of the logistic loss, which is expressed with the generalized Lambert W function. Experiments carried out on standard datasets demonstrate the efficiency of our approach w.r.t. stochastic gradient-like methods.

7.7. A probabilistic incremental proximal gradient method

Participant: Emilie Chouzenoux (Collaboration: Omer Deniz Akyildiz, Alan Turing Institute, London, UK, Victor Elvira, University of Edinburgh, Joaquin Miguez, Universidad Carlos III de Madrid, Spain)
In the paper [3], we proposed a probabilistic optimization method, named probabilistic incremental proximal gradient (PIPG) method, by developing a probabilistic interpretation of the incremental proximal gradient algorithm. We explicitly model the update rules of the incremental proximal gradient method and develop a systematic approach to propagate the uncertainty of the solution estimate over iterations. The PIPG algorithm takes the form of Bayesian filtering updates for a state-space model constructed by using the cost function. Our framework makes it possible to utilize well-known exact or approximate Bayesian filters, such as Kalman or extended Kalman filters, to solve large scale regularized optimization problems.

7.8. Optimal Multivariate Gaussian Fitting with Applications to PSF Modeling in Two-Photon Microscopy Imaging

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Claire Lefort, XLIM, Limoges, Tim Tsz-Kit Lau, Northwestern University, USA)

Fitting Gaussian functions to empirical data is a crucial task in a variety of scientific applications, especially in image processing. However, most of the existing approaches for performing such fitting are restricted to two dimensions and they cannot be easily extended to higher dimensions. Moreover, they are usually based on alternating minimization schemes which benefit from few theoretical guarantees in the underlying nonconvex setting. In the paper [12], we provided a novel variational formulation of the multivariate Gaussian fitting problem, which is applicable to any dimension and accounts for possible non-zero background and noise in the input data. The block multiconvexity of our objective function leads us to propose a proximal alternating method to minimize it in order to estimate the Gaussian shape parameters. The resulting FIGARO algorithm is shown to converge to a critical point under mild assumptions. The algorithm shows a good robustness when tested on synthetic datasets. To demonstrate the versatility of FIGARO, we also illustrate its excellent performance in the fitting of the Point Spread Functions of experimental raw data from a two-photon fluorescence microscope.

7.9. Calibration-less parallel imaging compressed sensing reconstruction based on OSCAR regularization

Participants: Emilie Chouzenoux, Loubna El Gueddari (Collaboration: Philippe Ciuciu, Alexandre Vignaut, Inria Saclay, Parietal)

Over the last decade, the combination of parallel imaging (PI) and compressed sensing (CS) in magnetic resonance imaging (MRI) has allowed to speed up acquisition while maintaining a good signal-to-noise ratio (SNR) for millimetric resolution. Self-calibrating techniques such as L1-ESPiRIT have emerged as a standard approach to estimate the coil sensitivity maps that are required at the reconstruction stage. Although straightforward in Cartesian acquisitions, these approaches become more computationally demanding in non-Cartesian scenarios especially for high resolution imaging (e.g. 500 μm in plane). Instead, calibration-less techniques no longer require this prior knowledge to perform multi-channel image reconstruction from undersampled k-space data. In this work, we introduce a new calibration-less PI-CS reconstruction method that is particularly suited to non-Cartesian data. It leverages structure sparsity of the multi-channel images in a wavelet transform domain while adapting to SNR inhomogeneities across receivers thanks to the OSCAR-norm regularization. Comparison and validation on 8 to 20-fold prospectively accelerated high-resolution ex-vivo human brain MRI data collected at 7 Tesla shows that the subbandwise OSCAR-norm regularization achieves the best trade-off between image quality and computational cost at the reconstructions stage compared to other tested versions (global, scalewise and pixelwise). This approach provides slight to moderate improvement over its state-of-the-art competitors (self-calibrating ‘1-ESPIRiT method and calibration-less AC-LORAKS and CaLM methods) in terms of closeness to the Cartesian reference magnitude image. Importantly, it also preserves much better phase information compared to other approaches [37], [57], [62].

7.10. Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet (Collaboration: Alessandro Benfenati, Universita di Milano)

In recent years, there has been a growing interest in mathematical models leading to the minimization, in a symmetric matrix space, of a Bregman divergence coupled with a regularization term. We address problems of this type within a general framework where the regularization term is split into two parts, one being a spectral function while the other is arbitrary. A Douglas–Rachford approach is proposed to address such problems, and a list of proximity operators is provided allowing us to consider various choices for the fit–to–data functional and for the regularization term. Based on our theoretical results, two novel approaches are proposed for the noisy graphical lasso problem, where a covariance or precision matrix has to be statistically estimated in presence of noise. The Douglas–Rachford approach directly applies to the estimation of the covariance matrix. When the precision matrix is sought, we solve a nonconvex optimization problem. More precisely, we propose a majorization–minimization approach building a sequence of convex surrogates and solving the inner optimization subproblems via the aforementioned Douglas–Rachford procedure. We establish conditions for the convergence of this iterative scheme. We illustrate the good numerical performance of the proposed approaches with respect to state–of–the–art approaches on synthetic and real-world datasets [4].

7.11. Representation Learning on Real-World Graphs

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat

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 *Kernel Node Embeddings* (KernelNE) [53], a model that aims to bring together two popular approaches for NRL, namely matrix factorization and random walk-based models. KernelNE is a weighted matrix factorization model which encodes random walk-based information about the nodes of the graph. The main benefit of this formulation is that it allows to utilize kernel functions on the computation of the embeddings.

Our second approach is motivated by the fact that the popular Skip-Gram algorithm models the conditional distribution of nodes within a random walk based on the softmax function, which might prohibit to capture richer types of interaction patterns among nodes that co-occur within a random walk. Here we argue that considering more expressive conditional probability models to relate nodes within a random walk sequence, might lead to more informative representations. That way, we have introduced the *Exponential Family Graph Embedding* (EFGE) model [54], that capitalizes on exponential family distribution models to capture interactions between nodes.

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 random walk-based models accompanied with kernels as well as exponential family distributions outperform widely-known baseline NRL methods.

7.12. Semi-supervised Learning for Misinformation Detection

Participants: Fragkiskos Malliaros (Collaboration: Adrien Benamira, Benjamin Devillers, Etienne Lesot, Ayush K. Ray, Manal Saadi, CentraleSupélec)

Social networks have become the main platforms for information dissemination. Nevertheless, due to the increasing number of users, social media platforms tend to be highly vulnerable to the propagation of disinformation – making the detection of fake news a challenging task. In our work, we have focused on content-based methods for detecting fake news – casting the problem to a binary text classification one (an article corresponds to either fake news or not). The main challenge here stems from the fact that the number of labeled data is limited; very few articles can be examined and annotated as fake. To this extend, we opted for semi-supervised learning approaches. In particular, we have proposed a graph-based semi-supervised fake news detection method, based on graph neural networks [34]. Our intuition is that, graphs are expressive

models that are able to capture contextual dependencies among articles, alleviating the label scarcity constraint. On a high level, our framework is composed of three components: i) embedding of articles in the Euclidean space; ii) construction of an article similarity graph; iii) inference of missing labels using graph learning techniques. The experimental results indicate that the proposed methodology achieves better performance compared to traditional classification techniques, especially when trained on limited number of labeled articles.

7.13. A Perturb and Combine Approach to Analyze Real-World Graphs

Participants: Fragkiskos Malliaros (Collaboration: Antoine J.-P. Tixier, Maria Evgenia G. Rossi, Jesse Read, Michalis Vazirgiannis, École Polytechnique)
Influential spreaders are nodes that can diffuse information to the largest part of the network in a minimum amount of time. Detecting influential spreaders is an important task with numerous real-world applications. Nevertheless, some of the most effective influential spreader detection algorithms (e.g., the k-core decomposition) are unstable to small perturbations of the network structure. Inspired by bagging in Machine Learning, we have proposed the first Perturb and Combine (P&C) procedure for networks [51]. It (1) creates many perturbed versions of a given graph, (2) applies a node scoring function separately to each graph, and (3) combines the results. Experiments conducted on real-world networks of various sizes with the k-core, generalized k-core, and PageRank algorithms reveal that P&C brings substantial improvements. Moreover, this performance boost can be obtained at almost no extra cost through parallelization. Finally, a bias-variance analysis suggests that P&C works mainly by reducing bias, and that therefore, it should be capable of improving the performance of all vertex scoring functions, including stable ones.

7.14. Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping: Mean-square and linear convergence

Participant: Jean-Christophe Pesquet (Collaboration: Patrick Louis Combettes, North Carolina University, USA)
In our previous work, we 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 paper [14] establishes results on the mean-square and linear convergence of the iterates. Applications to monotone operator splitting and proximal optimization algorithms are presented.

7.15. Rational optimization for non- linear 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 [7].

7.16. Deep neural network structures solving variational inequalities

Participant: Jean-Christophe Pesquet (Collaboration: Patrick Louis Combettes, North Carolina University, USA)
Motivated by structures that appear in deep neural networks, we investigate nonlinear composite models alternating proximity and affine operators defined on different spaces. We first show that a wide range of activation operators used in neural networks are actually proximity operators. We then establish conditions for the averagedness of the proposed composite constructs and investigate their asymptotic properties. It is shown that the limit of the resulting process solves a variational inequality which, in general, does not derive from a minimization problem [13].

7.17. Generation of patient-specific cardiac vascular networks: a hybrid image-based and synthetic geometric model

Participant: Hugues Talbot (Collaboration: Clara Jaquet, Laurent Najman, ESIEE Paris, Leo Grady, Michiel Schaap, Buzzy Spain, Hyun Kim, Charles Taylor, HeartFlow Inc, Irene Vignon-Clementel, Inria Paris)
In this work, we have proposed a blood-vessel generation procedure for extending known patient vasculature over and within the heart ventricle [19]. 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.

7.18. High throughput automated detection of axial malformations in Medaka fish embryo

Participant: Hugues Talbot (Collaboration: Diane Genest, Élodie Puybareau, Jean Cousty, ESIEE Paris, Marc Léonard, Noémie de Crozé, L'Oréal Recherche)
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 most important malformations in fish embryo. For examples those affecting the tail or the eyes, based on image analysis and machine learning [16]. We have also proposed an atlas-based automated procedure for detecting swim bladder malformations, which are very difficult to assess manually [39].

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.

7.19. Quantitative PET in the context of lymphoma

Participant: Hugues Talbot (Collaboration: Eloïse Grossiord, Laurent Najman, ESIEE Paris, Benoît Naegel, iCube, Strasbourg, Nicolas Passat, CRESTIC, Reims)
Lymphoma is type of cancer affecting the lymph system. Similar to blood disorders, these cancers can be difficult to cure because they affect a large portion of the body and metastasize easily. In contrast to leukemia, lymphoma also affects organs: the lymph nodes. Assessing the effectiveness of therapies implies to follow the impact of treatment on lymph nodes. This requires segmenting a large number of lesion, often several dozens. In [17], we have proposed an automated procedure based on hierarchical mathematical morphology, which has been extensively validated, and is now available as a plug-in for ImageJ/FIJI.

7.20. nD variational restoration of curvilinear structures with prior-based directional regularization

Participant: Hugues Talbot (Collaboration: Odyssée Merveille, Benoît Naegel, iCube, Strasbourg, Nicolas Passat, CRESTIC, Reims)

Curvilinear structure restoration in image processing procedures is a difficult task, which can be compounded when these structures are thin, i.e., when their smallest dimension is close to the resolution of the sensor. Many recent restoration methods involve considering a local gradient-based regularization term as prior, assuming gradient sparsity. An isotropic gradient operator is typically not suitable for thin curvilinear structures, since gradients are not sparse for these. In this paper [22], we propose a mixed gradient operator that combines a standard gradient in the isotropic image regions, and a directional gradient in the regions where specific orientations are likely. In particular, such information can be provided by curvilinear structure detectors (e.g., RORPO or Frangi filters). Our proposed mixed gradient operator, that can be viewed as a companion tool of such detectors, is proposed in a discrete framework and its formulation/computation holds in any dimension; in other words, it is valid in Z^n, $n \geq 1$. We show how this mixed gradient can be used to construct image priors that take edge orientation, as well as intensity, into account, and then involved in various image processing tasks while preserving curvilinear structures. The experiments carried out on 2D, 3D, real, and synthetic images illustrate the relevance of the proposed gradient, and its use in variational frameworks for both denoising and segmentation tasks.

7.21. Skin aging automated assessment

Participant: Hugues Talbot (Collaboration: Julie Robic, Alex Nkengne, Clarins laboratories, Benjamin Perret, Michel Couprie, ESIEE Paris)

With aging, human skin becomes drier, thinner and more irregular, but these characteristics are highly person-dependent, and can be brought about via exposure to heat, cold or Sun. It is important to the cosmetics industry to assess objectively the effect of their products on skin aging. With our partner Clarins laboratory, we have proposed a series of automated procedures based on graph-based image analysis. We have in particular proposed to detect the surface that correspond to the dermal-epidermal junction [25], and a series or procedures to link the appearance of this surfact to aging characteristics [48]. Both have been validated by dermatologists specialized in skin aging.

7.22. Particle tracking

Participant: Hugues Talbot (Collaboration: Alessandro Benfenati, Universita di Milano, Francesco Bonacci, Laboratoire Navier, Tarik Bourouina, ESIEE Paris)

Fluorescent bead tracking is important in biomedical application related to biomechanics, rheology and fluid dynamics. We have made several contributions for the detection and tracking of micrometer-scale fluorescent bead in 3D confocal microscopy [47], [61]. Many software packages exist for 2D tracking but almost none exist for 3D. It is a harder problem because in general beads are not fixed and move between plane acquisitions, due to the relatively slow scanning characteristics of confocal microscopy.

7.23. Artificial Intelligence Applications for Thoracic imaging

Participants: Guillaume Chassagnon, Maria Vakalopoulou (Collaboration: Marie-Pierre Revel and Nikos Paragios, AP-HP - Hopital Cochin Broca Hotel Dieu, Therapanacea)

Relevance and penetration of machine learning in clinical practice is a recent phenomenon with multiple applications being currently under development. Deep learning –and especially convolutional neural networks (CNNs)– is a subset of machine learning, which has recently entered the field of thoracic imaging. The structure of neural networks, organized in multiple layers, allows them to address complex tasks. For several clinical situations, CNNs have demonstrated superior performance as compared with classical machine learning algorithms and in some cases achieved comparable or better performance than clinical experts. Chest radiography, a high-volume procedure, is a natural application domain because of the large amount of stored images and reports facilitating the training of deep learning algorithms. Several algorithms for automated reporting have been developed. The training of deep learning algorithm CT images is more complex due to the dimension, variability, and complexity of the 3D signal. The role of these methods is likely to increase in clinical practice as a complement of the radiologist's expertise. The objective of these two reviews [9], [26] is to provide definitions for understanding the methods and their potential applications for thoracic imaging.

7.24. Use of Elastic Registration in Pulmonary MRI for the Assessment of Pulmonary Fibrosis in Patients with Systemic Sclerosis

Participants: Guillaume Chassagnon, Maria Vakalopoulou (Collaboration: Charlotte Martin, Rafael Marini Silva, Alexis Régent, Luc Mouthon, Nikos Paragios and Marie-Pierre Revel, AP-HP - Hopital Cochin Broca Hotel Dieu, Therapanacea)

Elastic registration of inspiratory and expiratory MRI revealed qualitative and quantitative differences in lung deformation in study participants with systemic sclerosis compared with healthy volunteers. Current imaging methods are not sensitive to changes in pulmonary function resulting from fibrosis. MRI with ultrashort echo time can be used to image the lung parenchyma and lung motion. To evaluate elastic registration of inspiratory to expiratory lung MRI for the assessment of pulmonary fibrosis in study participants with systemic sclerosis (SSc). This prospective study [8] was performed from September 2017 to March 2018 and recruited healthy volunteers and participants with SSc and high-resolution CT (within the previous 3 months) of the chest for lung MRI. Two breath-hold, coronal, three-dimensional, ultrashort–echo-time, gradient-echo sequences of the lungs were acquired after full inspiration and expiration with a 3.0-T unit. Images were registered from inspiration to expiration by using an elastic registration algorithm. Jacobian determinants were calculated from deformation fields and represented on color maps. Similarity between areas with marked shrinkage and logarithm of Jacobian determinants were compared between healthy volunteers and study participants with SSc. Receiver operating characteristic curve analysis was performed to determine the best Dice similarity coefficient threshold for diagnosis of fibrosis. Sixteen participants with SSc (seven with pulmonary fibrosis at high-resolution CT) and 11 healthy volunteers were evaluated. Areas of marked shrinkage during expiration with logarithm of Jacobian determinants less than -0.15 were found in the posterior lung bases of healthy volunteers and in participants with SSc without fibrosis, but not in participants with fibrosis. The sensitivity and specificity of MRI for presence of fibrosis at high-resolution CT were 86% and 75%, respectively (area under the curve, 0.81; P = .04) by using a threshold of 0.36 for Dice similarity coefficient. Elastic registration of inspiratory to expiratory MRI shows less lung base respiratory deformation in study participants with systemic sclerosis related pulmonary fibrosis compared with participants without fibrosis.

7.25. U-ReSNet: Ultimate Coupling of Registration and Segmentation with Deep Nets

Participants: Théo Estienne, Enzo Battistella, Marvin Lerousseau, Roger Sun, Maria Vakalopoulou (Collaboration: Stergios Christodoulidis, Alexandre Carre, Guillaume Klausner, Stavroula Mougiakakou, Charlotte Robert, Nikos Paragios and Eric Deutsch, Institute Gustave Roussy, University of Bern, Therapanacea)

We proposed in [58] a 3D deep neural network called U-ReSNet, a joint framework that can accurately register and segment medical volumes. The proposed network learns to automatically generate linear and elastic deformation models, trained by minimizing the mean square error and the local cross correlation similarity metrics. In parallel, a coupled architecture is integrated, seeking to provide segmentation maps for anatomies or tissue patterns using an additional decoder part trained with the dice coefficient metric. U-ReSNet is trained in an end to end fashion, while due to this joint optimization the generated network features are more informative leading to promising results compared to other deep learning-based methods existing in the literature. We evaluated the proposed architecture using the publicly available OASIS 3 dataset, measuring the dice coefficient metric for both registration and segmentation tasks. Our promising results indicate the potentials of our method which is composed from a convolutional architecture that is extremely simple and light in terms of parameters.

7.26. Gene Expression High-Dimensional Clustering Towards a Novel, Robust, Clinically Relevant and Highly Compact Cancer Signature

Participants: Enzo Battistella, Théo Estienne, Marvin Lerousseau, Roger Sun, Maria Vakalopoulou (Collaboration: Charlotte Robert, Nikos Paragios and Eric Deutsch, Institute Gustave Roussy, Therapanacea)
Precision medicine, a highly disruptive paradigm shift in healthcare targeting the personalizing treatment, heavily relies on genomic data. However, the complexity of the biological interactions, the important number of genes as well as the lack of substantial patient's clinical data consist a tremendous bottleneck on the clinical implementation of precision medicine. In this work [32], we introduce a generic, low dimensional gene signature that represents adequately the tumor type. Our gene signature is produced using LP-stability algorithm, a high dimensional center-based unsupervised clustering algorithm working in the dual domain, and is very versatile as it can consider any arbitrary distance metric between genes. The gene signature produced by LP-stability reports at least 10 times better statistical significance and 35% better biological significance than the ones produced by two referential unsupervised clustering methods. Moreover, our experiments demonstrate that our low dimensional biomarker (27 genes) surpass significantly existing state of the art methods both in terms of qualitative and quantitative assessment while providing better associations to tumor types than methods widely used in the literature that rely on several omics data.

7.27. A Novel Object-Based Deep Learning Framework for Semantic Segmentation of Very High-Resolution Remote Sensing Data: Comparison with Convolutional and Fully Convolutional Networks

Participants: Maria Papadomanolaki and Maria Vakalopoulou (Collaboration: Konstantinos Karantzalos, National Technical University of Athens)
Deep learning architectures have received much attention in recent years demonstrating state-of-the-art performance in several segmentation, classification and other computer vision tasks. Most of these deep networks are based on either convolutional or fully convolutional architectures. In this study [24], we propose a novel object-based deep-learning framework for semantic segmentation in very high-resolution satellite data. In particular, we exploit object-based priors integrated into a fully convolutional neural network by incorporating an anisotropic diffusion data preprocessing step and an additional loss term during the training process. Under this constrained framework, the goal is to enforce pixels that belong to the same object to be classified at the same semantic category. We compared thoroughly the novel object-based framework with the currently dominating convolutional and fully convolutional deep networks. In particular, numerous experiments were conducted on the publicly available ISPRS WGII/4 benchmark datasets, namely Vaihingen and Potsdam, for validation and inter-comparison based on a variety of metrics. Quantitatively, experimental results indicate that, overall, the proposed object-based framework slightly outperformed the current state-of-the-art fully convolutional networks by more than 1% in terms of overall accuracy, while intersection over union results are improved for all semantic categories. Qualitatively, man-made classes with more strict geometry such as buildings were the ones that benefit most from our method, especially along object boundaries, highlighting the great potential of the developed approach.

7.28. A multi-task deep learning framework coupling semantic segmentation and image reconstruction for very high resolution imagery

Participants: Maria Papadomanolaki and Maria Vakalopoulou (Collaboration: Konstantinos Karantzalos, National Technical University of Athens)
Semantic segmentation, especially for very high-resolution satellite data, is one of the pillar problems in the remote sensing community. Lately, deep learning techniques are the ones that set the state-of-the-art for a number of benchmark datasets, however, there are still a lot of challenges that need to be addressed, especially in the case of limited annotations. To this end, in this study [45], we propose a novel framework based on deep neural networks that is able to address concurrently semantic segmentation and image reconstruction in an end to end training. Under the proposed formulation, the image reconstruction acts as a regularization, constraining efficiently the solution in the entire image domain. This self-supervised component helps significantly the generalization of the network for the semantic segmentation, especially in cases of a low number of annotations. Experimental results and the performed quantitative evaluation on the publicly available ISPRS (WGIII/4) dataset indicate the great potential of the developed approach.

7.29. Detecting Urban Changes with Recurrent Neural Networks from Multitemporal Sentinel-2 Data

Participants: Maria Papadomanolaki, Sagar Verma and Maria Vakalopoulou (Collaboration: Siddharth Gupta and Konstantinos Karantzalos, Granular AI and National Technical University of Athens)
The advent of multitemporal high resolution data, like the Copernicus Sentinel-2, has enhanced significantly the potential of monitoring the earth's surface and environmental dynamics. In this study [45], we present a novel deep learning framework for urban change detection which combines state-of-the-art fully convolutional networks (similar to U-Net) for feature representation and powerful recurrent networks (such as LSTMs) for temporal modeling. We report our results on the recently publicly available bi-temporal Onera Satellite Change Detection (OSCD) Sentinel-2 dataset, enhancing the temporal information with additional images of the same region on different dates. Moreover, we evaluate the performance of the recurrent networks as well as the use of the additional dates on the unseen test-set using an ensemble cross-validation strategy. All the developed models during the validation phase have scored an overall accuracy of more than 95%, while the use of LSTMs and further temporal information, boost the F1 rate of the change class by an additional 1.5%.

7.30. Image Registration of Satellite Imagery with Deep Convolutional Neural Networks

Participants: Maria Vakalopoulou and Mihir Sahasrabudhe (Collaboration: Stergios Christodoulidis, Stavroula Mougiakakou and Nikos Paragios, University of Bern and Therapanacea)
Image registration in multimodal, multitemporal satellite imagery is one of the most important problems in remote sensing and essential for a number of other tasks such as change detection and image fusion. In this study [52], inspired by the recent success of deep learning approaches we propose a novel convolutional neural network architecture that couples linear and deformable approaches for accurate alignment of remote sensing imagery. The proposed method is completely unsupervised, ensures smooth displacement fields and provides real time registration on a pair of images. We evaluate the performance of our method using a challenging multitemporal dataset of very high resolution satellite images and compare its performance with a state of the art elastic registration method based on graphical models. Both quantitative and qualitative results prove the high potentials of our method.

7.31. Lifting AutoEncoders: Unsupervised Learning of a Fully-Disentangled 3D Morphable Model Using Deep Non-Rigid Structure From Motion

Participants: Mihir Sahasrabudhe (Collaboration: Eqward Bartrum, Riza Alp Guler, Dimitris Samaras and Iasonas Kokkinos, Stony Brook, UCL and Ariel AI)

We introduced, in [49], Lifting Autoencoders, a generative 3D surface-based model of object categories. We bring together ideas from non-rigid structure from motion, image formation, and morphable models to learn a controllable, geometric model of 3D categories in an entirely unsupervised manner from an unstructured set of images. We exploit the 3D geometric nature of our model and use normal information to disentangle appearance into illumination, shading, and albedo. We further use weak supervision to disentangle the non-rigid shape variability of human faces into identity and expression. We combine the 3D representation with a differentiable renderer to generate RGB images and append an adversarially trained refinement network to obtain sharp, photorealistic image reconstruction results. The learned generative model can be controlled in terms of interpretable geometry and appearance factors, allowing us to perform photorealistic image manipulation of identity, expression, 3D pose, and illumination properties.

PARIETAL Project-Team

7. New Results

7.1. The visual word form area (VWFA) is part of both language and attention circuitry

While predominant models of visual word form area (VWFA) function argue for its specific role in decoding written language, other accounts propose a more general role of VWFA in complex visual processing. However, a comprehensive examination of structural and functional VWFA circuits and their relationship to behavior has been missing. Here, using high-resolution multimodal imaging data from a large Human Connectome Project cohort (N = 313), we demonstrate robust patterns of VWFA connectivity with both canonical language and attentional networks. Brain-behavior relationships revealed a striking pattern of double dissociation: structural connectivity of VWFA with lateral temporal language network predicted language, but not visuo-spatial attention abilities, while VWFA connectivity with dorsal fronto-parietal attention network predicted visuo-spatial attention, but not language abilities. Our findings support a multiplex model of VWFA function characterized by distinct circuits for integrating language and attention, and point to connectivity-constrained cognition as a key principle of human brain organization.

More information can be found in [7].

7.2. SPARKLING: variable-density k-space filling curves for accelerated T2* -weighted MRI

Funding information Purpose: To present a new optimization-driven design of optimal k-space trajectories in the context of compressed sensing: Spreading Projection Algorithm for Rapid K-space sampLING (SPARKLING). Theory: The SPARKLING algorithm is a versatile method inspired from stippling techniques that automatically generates optimized sampling patterns compatible with MR hardware constraints on maximum gradient amplitude and slew rate. These non-Cartesian sampling curves are designed to comply with key criteria for optimal sampling: a controlled distribution of samples (e.g., variable density) and a locally uniform k-space coverage. Methods: Ex vivo and in vivo prospective $T2^{\star}$-weighted acquisitions were performed on a 7 Tesla scanner using the SPARKLING tra-jectories for various setups and target densities. Our method was compared to radial and variable-density spiral trajectories for high resolution imaging. Results: Combining sampling efficiency with compressed sensing, the proposed sampling patterns allowed up to 20-fold reductions in MR scan time (compared to fully-sampled Cartesian acquisitions) for two-dimensional $T2^{\star}$-weighted imaging without deterioration of image quality, as demonstrated by our experimental results at 7 Tesla on in vivo human brains for a high in-plane resolution of 390 um. In comparison to existing non-Cartesian sampling strategies, the proposed technique also yielded superior image quality. Conclusion: The proposed optimization-driven design of k-space trajectories is a versatile framework that is able to enhance MR sampling performance in the context of compressed sensing.

More information can be found in [14].

7.3. Benchmarking functional connectome-based predictive models for resting-state fMRI

Functional connectomes reveal biomarkers of individual psychological or clinical traits. However, there is great variability in the analytic pipelines typically used to derive them from rest-fMRI cohorts. Here, we consider a specific type of studies, using predictive models on the edge weights of functional connectomes, for which we highlight the best modeling choices. We systematically study the prediction performances of models

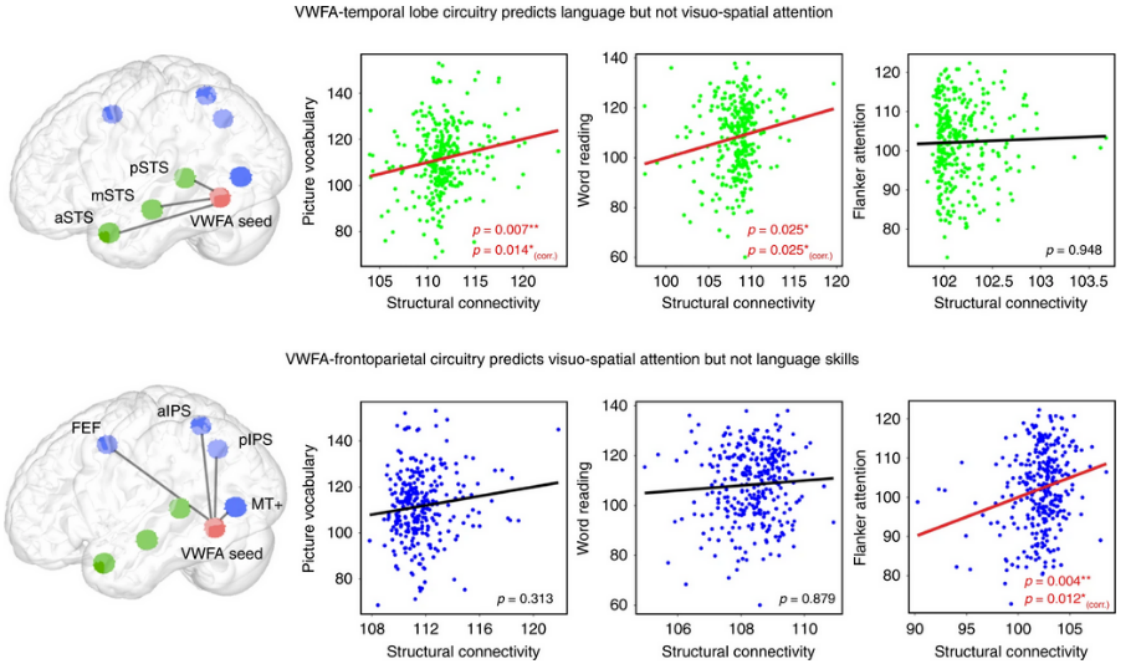


Figure 3. Structural connectivity of VWFA with STS nodes (anterior, middle, and posterior STS) was significantly correlated with individuals' performance on picture vocabulary and word reading tasks, but not on the Flanker visuo-spatial attention task. b Structural connectivity of VWFA with fronto-parietal attention network nodes (FEF, aIPS, pIPS, and MT+) was significantly correlated with individuals' performance on the Flanker attention task but not on the word reading or picture vocabulary tasks. Y-axis is age-adjusted performance scores on cognitive tasks and x-axis presents the predicted performance scores on cognitive tasks from the probability of structural connectivity of VWFA to either the language or the attention network ROIs.

in 6 different cohorts and a total of 2 000 individuals, encompassing neuro-degenerative (Alzheimer's, Post-traumatic stress disorder), neuro-psychiatric (Schizophrenia, Autism), drug impact (Cannabis use) clinical settings and psychological trait (fluid intelligence). The typical prediction procedure from rest-fMRI consists of three main steps: defining brain regions, representing the interactions, and supervised learning. For each step we benchmark typical choices: 8 different ways of defining regions –either pre-defined or generated from the rest-fMRI data– 3 measures to build functional connectomes from the extracted time-series, and 10 classification models to compare functional interactions across subjects. Our benchmarks summarize more than 240 different pipelines and outline modeling choices that show consistent prediction performances in spite of variations in the populations and sites. We find that regions defined from functional data work best; that it is beneficial to capture between-region interactions with tangent-based parametrization of covariances, a midway between correlations and partial correlation; and that simple linear predictors such as a logistic regression give the best predictions. Our work is a step forward to establishing reproducible imaging-based biomarkers for clinical settings.

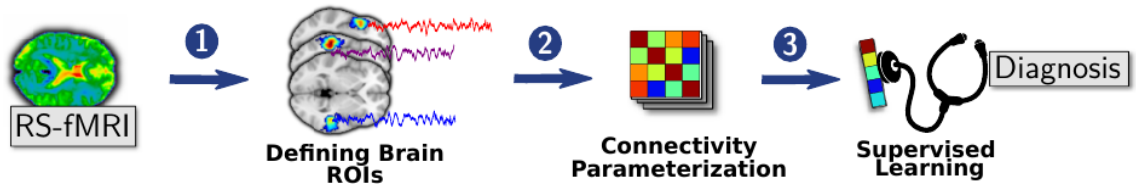


Figure 4. Functional connectome-prediction pipeline with three main steps: 1) definition of brain regions (ROIs) from rest-fMRI images or us- ing already defined reference atlases, 2) quantifying functional interactions from time series signals extracted from these ROIs and 3) comparisons of functional interactions across subjects using supervised learning.

More information can be found in [8].

7.4. Population shrinkage of covariance (PoSCE) for better individual brain functional-connectivity estimation

Estimating covariances from functional Magnetic Resonance Imaging at rest (r-fMRI) can quantify interactions between brain regions. Also known as brain functional connectivity, it reflects inter-subject variations in behavior and cognition, and characterizes neuropathologies. Yet, with noisy and short time-series, as in r-fMRI, covariance estimation is challenging and calls for penalization, as with shrinkage approaches. We introduce population shrinkage of covariance estimator (PoSCE) : a covariance estimator that integrates prior knowledge of covariance distribution over a large population, leading to a non-isotropic shrinkage. The shrinkage is tailored to the Riemannian geometry of symmetric positive definite matrices. It is coupled with a probabilistic modeling of the individual and population covariance distributions. Experiments on two large r-fMRI datasets (HCP n=815, Cam-CAN n=626) show that PoSCE has a better bias-variance trade-off than existing covariance estimates: this estimator relates better functional-connectivity measures to cognition while capturing well intra-subject functional connectivity.

More information can be found in [20].

7.5. Feature Grouping as a Stochastic Regularizer for High-Dimensional Structured Data

In many applications where collecting data is expensive , for example neuroscience or medical imaging, the sample size is typically small compared to the feature dimension. It is challenging in this setting to train

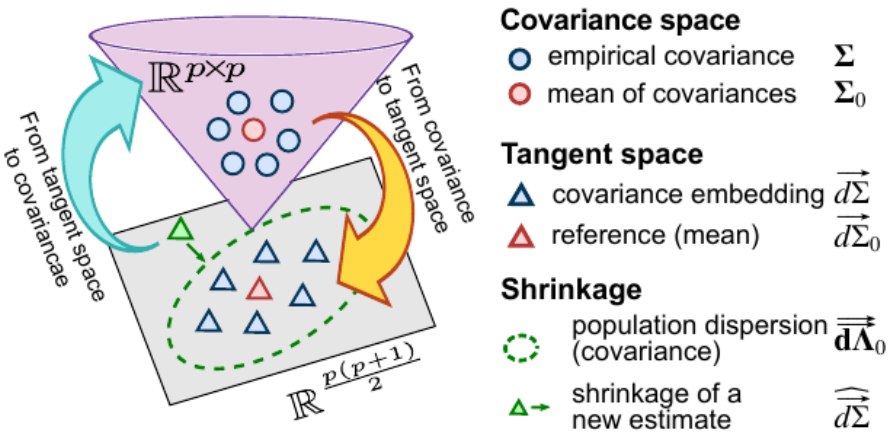


Figure 5. Tangent embedding and population prior modeling. Σ_0 is the mean covariance from a train set of covariances. It is the reference point in the tangent space. The population prior is defined as a Gaussian multivariate dis- tribution centered on $d\Sigma_0$. Λ_0 is the covariance dispersion over the population. The arrows depict the mapping between the non-Euclidean covariance space and the tangent space.

expressive, non-linear models without overfitting. These datasets call for intelligent regularization that exploits known structure, such as correlations between the features arising from the measurement device. However, existing structured regularizers need specially crafted solvers, which are difficult to apply to complex models. We propose a new regularizer specifically designed to leverage structure in the data in a way that can be applied efficiently to complex models. Our approach relies on feature grouping, using a fast clustering algorithm inside a stochas-tic gradient descent loop: given a family of feature groupings that capture feature covariations, we randomly select these groups at each iteration. We show that this approach amounts to enforcing a denoising regularizer on the solution. The method is easy to implement in many model archi-tectures, such as fully connected neural networks, and has a linear computational cost. We apply this regularizer to a real-world fMRI dataset and the Olivetti Faces datasets. Experiments on both datasets demonstrate that the proposed approach produces models that generalize better than those trained with conventional regularizers, and also improves convergence speed.

More information can be found in [25].

7.6. Manifold-regression to predict from MEG/EEG brain signals without source modeling

Magnetoencephalography and electroencephalography (M/EEG) can reveal neuronal dynamics non-invasively in real-time and are therefore appreciated methods in medicine and neuroscience. Recent advances in modeling brain-behavior relationships have highlighted the effectiveness of Riemannian geometry for summarizing the spatially correlated time-series from M/EEG in terms of their covariance. However , after artefact-suppression, M/EEG data is often rank deficient which limits the application of Riemannian concepts. In this article, we focus on the task of regression with rank-reduced covariance matrices. We study two Riemannian approaches that vectorize the M/EEG covariance between-sensors through projection into a tangent space. The Wasserstein distance readily applies to rank-reduced data but lacks affine-invariance. This can be overcome by finding a common sub-space in which the covariance matrices are full rank, enabling the affine-invariant geometric distance. We investigated the implications of these two approaches in synthetic generative models, which allowed us to control estimation bias of a linear model for prediction. We show that Wasserstein and geometric

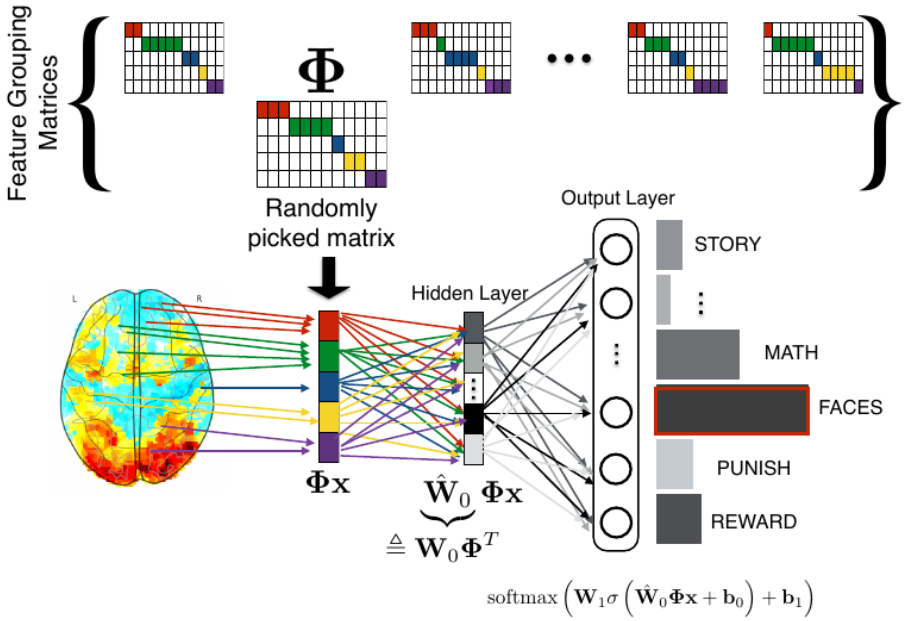


Figure 6. Illustration of the proposed approach: Forward propa- gation of a neural network with a single hidden layer using feature grouping during training. The parameters of the neural network to be estimated are $\mathbf{W}_0$, $\mathbf{b}_0$, $\mathbf{W}_1$, $\mathbf{b}1$. *A bank of feature grouping matrices are pre-generated where each matrix is calculated from a sub-sample of the training test. At each SGD iteration, a feature grouping matrix is sampled from the bank of pre-generated matri- ces. The gradient is then computed with respect to* $\widehat{\mathbf{W}}_0$ *to update* $\mathbf{W}_0$ *in backpropagation.*

distances allow perfect out-of-sample prediction on the generative models. We then evaluated the methods on real data with regard to their effectiveness in predicting age from M/EEG covariance matrices. The findings suggest that the data-driven Riemannian methods outperform different sensor-space estimators and that they get close to the performance of biophysics-driven source-localization model that requires MRI acquisitions and tedious data processing. Our study suggests that the proposed Riemannian methods can serve as fundamental building-blocks for automated large-scale analysis of M/EEG.

Figure 7. Age prediction on Cam-CAN MEG dataset for different methods, ordered by out-of-sample MAE. The y-axis depicts the projection method, with identity denoting the absence of projection. Colors indicate the subsequent embedding. The biophysics-driven MNE method (blue) performs best. The Riemannian methods (orange) follow closely and their performance depends little on the projection method. The non-Riemannian methods log-diag (green) perform worse, although the supervised projection clearly helps.

More information can be found in [47].

7.7. Stochastic algorithms with descent guarantees for ICA

Independent component analysis (ICA) is a widespread data exploration technique, where observed signals are modeled as linear mixtures of independent components. From a machine learning point of view, it amounts to a matrix factorization problem with a statistical independence criterion. Infomax is one of the most used ICA algorithms. It is based on a loss function which is a non-convex log-likelihood. We develop a new majorization-minimization framework adapted to this loss function. We derive an online algorithm for the streaming setting, and an incremental algorithm for the finite sum setting, with the following benefits. First, unlike most algorithms found in the literature, the proposed methods do not rely on any critical hyper-parameter like a step size, nor do they require a line-search technique. Second, the algorithm for the finite sum setting, although stochas-tic, guarantees a decrease of the loss function at each iteration. Experiments demonstrate progress on the state-of-the-art for large scale datasets, without the necessity for any manual parameter tuning.

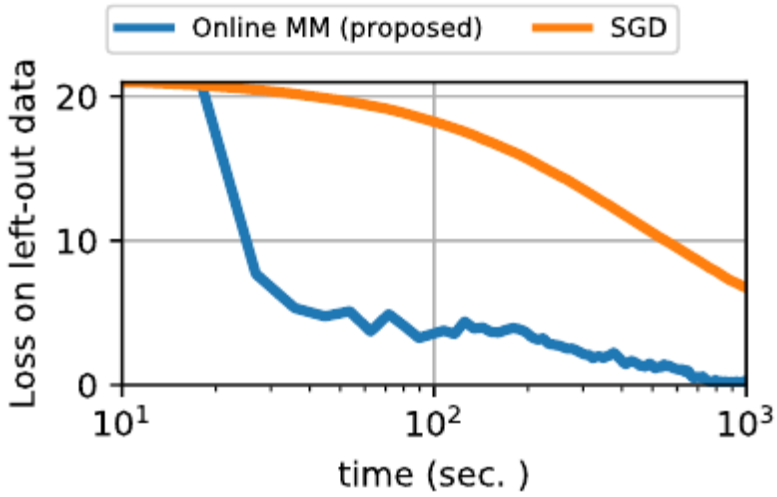


Figure 8. Online algorithms applied on a 32 GB real dataset with p = 100 and n = 4×10^7 . Time is in logarithmic scale. Values of the loss on left out data greater than its initial value are truncated.

More information can be found in [39].

7.8. Comparing distributions: $l1$ geometry improves kernel two-sample testing

Are two sets of observations drawn from the same distribution? This problem is a two-sample test. Kernel methods lead to many appealing properties. Indeed state-of-the-art approaches use the L2 distance between kernel-based distribution representatives to derive their test statistics. Here, we show that Lp distances (with $p \geq 1$) between these distribution representatives give metrics on the space of distributions that are well-behaved to detect differences between distributions as they metrize the weak convergence. Moreover, for analytic kernels, we show that the L1 geometry gives improved testing power for scalable computational procedures. Specifically, we derive a finite dimensional approximation of the metric given as the ℓ_1 norm of a vector which captures differences of expectations of analytic functions evaluated at spatial locations or frequencies (i.e, features). The features can be chosen to maximize the differences of the distributions and give interpretable indications of how they differs. Using an ℓ_1 norm gives better detection because differences between representatives are dense as we use analytic kernels (non-zero almost everywhere). The tests are consistent, while much faster than state-of-the-art quadratic-time kernel-based tests. Experiments on artificial and real-world problems demonstrate improved power/time tradeoff than the state of the art, based on ℓ_2 norms, and in some cases, better outright power than even the most expensive quadratic-time tests. More information can be found in [37].

7.9. Wasserstein regularization for sparse multi-task regression

We focus in this work on high-dimensional regression problems where each regressor can be associated to a location in a physical space, or more generally a generic geometric space. Such problems often employ sparse priors, which promote models using a small subset of regressors. To increase statistical power, the so-called multi-task techniques were proposed, which consist in the simultaneous estimation of several related models. Combined with sparsity assumptions, it lead to models enforcing the active regressors to be shared across models, thanks to, for instance L1 / Lq norms. We argue in this paper that these techniques fail to leverage the spatial information associated to regressors. Indeed, while sparse priors enforce that only a small subset

of variables is used, the assumption that these regressors overlap across all tasks is overly simplistic given the spatial variability observed in real data. In this paper, we propose a convex reg-ularizer for multi-task regression that encodes a more flexible geometry. Our regularizer is based on unbalanced optimal transport (OT) theory, and can take into account a prior geometric knowledge on the regressor variables, without necessarily requiring overlapping supports. We derive an efficient algorithm based on a regularized formulation of OT, which iterates through applications of Sinkhorn's algorithm along with coordinate descent iterations. The performance of our model is demonstrated on regular grids with both synthetic and real datasets as well as complex triangulated geometries of the cortex with an application in neuroimaging.

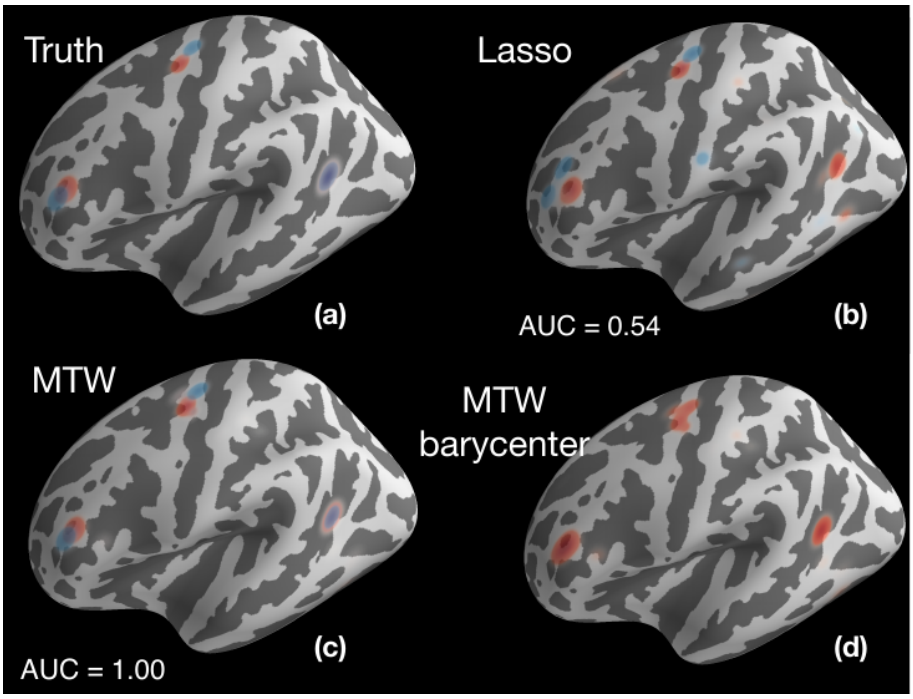


Figure 9. Each color corresponds to one of the two subjects (except for (d)). (a): True sources: one common feature (right side of the displayed hemisphere) and two non-overlapping sources. (b, c): Sources estimated by (b) Lasso and (c) MTW with the highest AUC score. (d) Shows the barycenter texttheta¯ associated with MTW model. In this figure, the displayed activations were smoothed for the sake of visibility.

More information can be found in [34].

7.10. Ensemble of Clustered Knockoffs for robust multivariate inference on MRI data

Continuous improvement in medical imaging techniques allows the acquisition of higher-resolution images. When these are used in a predictive setting, a greater number of explanatory variables are potentially related to the dependent variable (the response). Meanwhile, the number of acquisitions per experiment remains limited. In such high dimension/small sample size setting, it is desirable to find the explanatory variables that are truly related to the response while controlling the rate of false discoveries. To achieve this goal, novel multivariate inference procedures, such as knockoff inference, have been proposed recently. However, they require the feature covariance to be well-defined, which is impossible in high-dimensional settings. In this

paper, we propose a new algorithm, called Ensemble of Clustered Knockoffs, that allows to select explanatory variables while controlling the false discovery rate (FDR), up to a prescribed spatial tolerance. The core idea is that knockoff-based inference can be applied on groups (clusters) of voxels, which drastically reduces the problem's dimension; an ensembling step then removes the dependence on a fixed clustering and stabilizes the results. We benchmark this algorithm and other FDR-controlling methods on brain imaging datasets and observe empirical gains in sensitivity, while the false discovery rate is controlled at the nominal level.

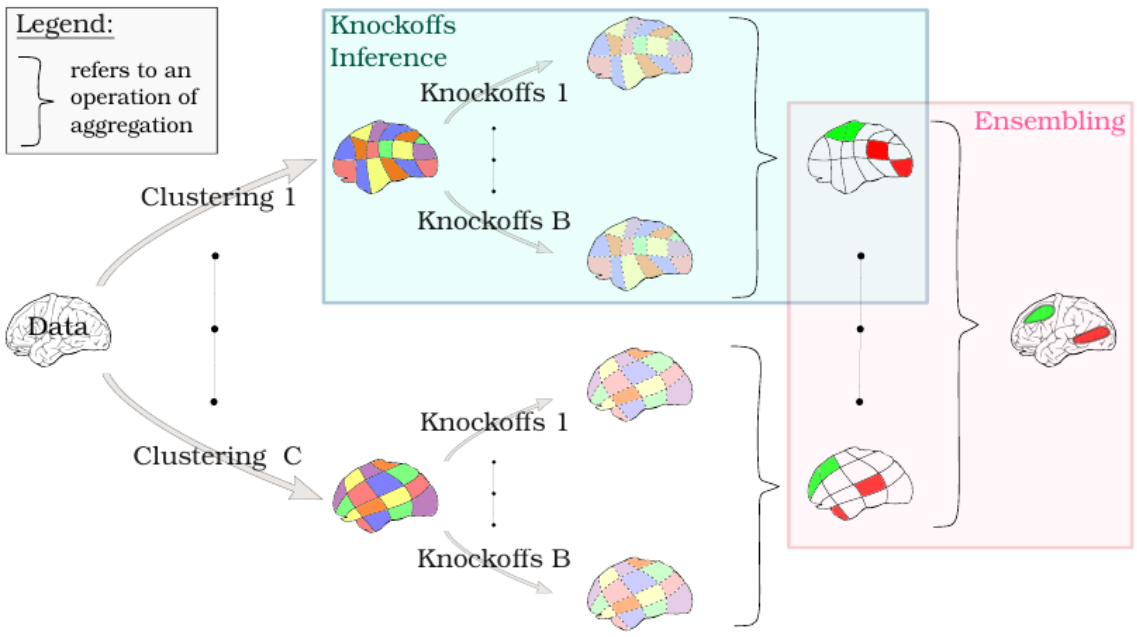


Figure 10. Representation of the ECKO algorithm. To create a stable inference result, we introduce ensembling steps both within each cluster level and at the voxel- level, across clusterings.

More information can be found in [46].

AIRSEA Project-Team

6. New Results

6.1. Modeling for Oceanic and Atmospheric flows

6.1.1. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Matthieu Brachet, Laurent Debreu, Emilie Duval, Christopher Eldred, Nicholas Kevlahan, Florian Lemarié, Gurvan Madec, 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). As an example, the stabilization of the coupling between barotropic (fast) and baroclinic (slow) modes in a three dimensional ocean model is a source of large numerical dissipation. This has been studied in details in [6].
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. The next COMMODORE meeting is planned for February 2020 and will take place in Hamburg. https://www.conferences.uni-hamburg.de/event/76.

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 favour the use of the non-hydrostatic compressible equations that removes the need for a 3D resolution at the price of reincluding acoustic waves [29].
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. In the context of the HEAT project, we worked on the analysis of dispersion analysis of continuous and discontinuous Galerkin methods of arbitrary degree of approximation ([31]), on compatible Galerkin schemes for shallow water model in 2D ([9]). 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 ([8]).

Accurate and stable implementation of bathymetry boundary conditions in ocean models remains a challenging problem. The dynamics of ocean flow often depend sensitively on satisfying bathymetry boundary conditions and correctly representing their complex geometry. Generalized (e.g.) terrain-following coordinates are often used in ocean models, but they require smoothing the bathymetry to reduce pressure gradient errors. Geopotential -coordinates are a common alternative that avoid pressure gradient and numerical diapycnal diffusion errors, but they generate spurious flow due to their “staircase” geometry. In [5], we introduce a new Brinkman volume penalization to approximate the no-slip boundary condition and complex geometry of bathymetry in ocean models. This approach corrects the staircase effect of -coordinates, does not introduce any new stability constraints on the geometry of the bathymetry and is easy to implement in an existing ocean model. The porosity parameter allows modelling subgrid scale details of the geometry. We illustrate the penalization and confirm its accuracy by applying it to three standard test flows: upwelling over a sloping bottom, resting state over a seamount and internal tides over highly peaked bathymetry features.

Figure (1) shows strong improvements obtained when the penalization method is used in comparison with traditional terrain following σ simulations. At 6km resolution, the penalization methods (Figure (1) d)), that takes into account details of bathymetry, allows to recover internal tide wave beams closed to the 3km simulation. (Figure (1) a)).

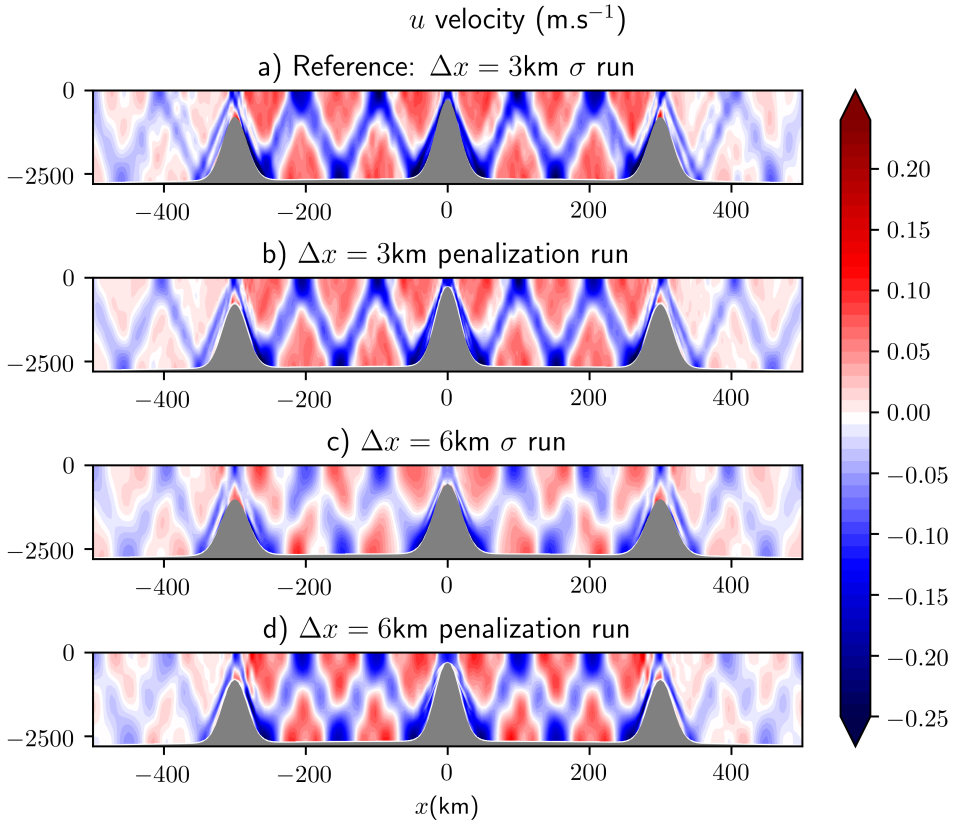


Figure 1. u velocity. Instantaneous solutions of the internal tide test case after 12 M2 tidal cycles of integration. (a) The reference σ coordinate run at 3 km resolution. (b) The penalized run at 3 km resolution. (c) The σ-coordinate run at 3 km resolution. (d) The penalized run at 6 km resolution.

6.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Florian Lemarié, Sophie Thery, Simon Clément.

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 [49]. 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 (see [21] for an overview). Our activities can be divided into four general topics

1. *Stability and consistency analysis of existing coupling methods*: in [49] 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 in [37] 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 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 [56]. 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 [57]. More recently we have started to work specifically on the discretization methods for the parameterization of planetary boundary layers in climate models [51] 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).This work will continue in the framework of the PhD-thesis of C. Simon.
3. *A simplified atmospheric boundary layer model for oceanic purposes*: Part of our activities within the IMMERSE project is dedicated to the development of a simplified model of the marine atmospheric boundary layer of intermediate complexity between a bulk parameterization and a full three-dimensional atmospheric model and to its integration to the NEMO general circulation model [24]. A constraint in the conception of such a simplified model is to allow an apt representation of the downward momentum mixing mechanism and partial re-energization of the ocean by the atmosphere while keeping the computational efficiency and flexibility inherent to ocean only modeling. A paper is in preparation and will be submitted in early 2020.
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. [50]) in large-scale realistic ocean-atmosphere coupled simulations [16], [53]. Moreover, within the ALBATROS project [23], 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 open discussion in Geoscientific Model Development journal on that topic [4].
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 [61] (paper in preparation). This work is done in the framework of S. Théry PhD (started in fall 2017). During the internship of C. Simon, efficient interface conditions have been derived at a (semi)-discrete level and can thus be systematically compared with the ones obtained from the continuous problem.

These topics are addressed through strong collaborations between the applied mathematicians and the climate community (Meteo-France, Ifremer, LMD, and LOCEAN). 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. In this context, a single-column version of the CNRM climate models has been designed and several coupling algorithms have been implemented (work done by S. Valcke, CERFACS). This model will be used to illustrate the relevance of our theoretical work in a semi-realistic context.

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 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.

Following this line of research, CASIS, a collaborative project with Mercator Océan started late 2017 until end of 2019 in order to extend these 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 [28], 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. Machine learning for parametrisation of the model dissipation.

Participants: Laurent Debreu, Eugene Kazantsev, Arthur Vidard, Olivier Zahm.

Artificial intelligence and machine learning may be considered as a potential way to address unresolved model scales and to approximate poorly known processes such as dissipation that occurs essentially at small scales. In order to understand the possibility to combine numerical model and neural network learned with the aid of external data, we develop a network generation and learning algorithm and use it to approximate nonlinear model operators. Beginning with a simple nonlinear equations like transport-diffusion and Burgers ones, we use artificially generated external data to learn the network by Adam algorithm [47]. Results show the possibility to approximate nonlinear, and even discontinuous dissipation operator with a quite good accuracy, however, several millions iterations are necessary to learn.

6.1.6. 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 is working on the design of methods to couple local nonhydrostatic models to larger scale hydrostatic ones (PhD started in Oct. 2018). 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). A prototype has been implemented, which allows for analytical solutions in simplified configurations and will make it possible to test different numerical approaches.

6.2. Assimilation of spatially dense observations

Participants: Elise Arnaud, François-Xavier Le Dimet, Arthur Vidard, Long Li, Emilie Rouzies.

6.2.1. 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 Jianwei Ma.

6.2.2. Observation error representation

Accounting for realistic observations errors is a known bottleneck in data assimilation, because dealing with error correlations is complex. Following a previous study on this subject, we propose to use multiscale modelling, more precisely wavelet transform, to address this question. In [3] we investigate the problem further by addressing two issues arising in real-life data assimilation: how to deal with partially missing data (e.g., concealed by an obstacle between the sensor and the observed system); how to solve convergence issues associated to complex observation error covariance matrices? Two adjustments relying on wavelets modelling are proposed to deal with those, and offer significant improvements. The first one consists in adjusting the variance coefficients in the frequency domain to account for masked information. The second one consists in a gradual assimilation of frequencies. Both of these fully rely on the multiscale properties associated with wavelet covariance modelling.

A collaborative project started with C. Lauvernet (IRSTEA) in order to make use of this kind of assimilation strategies on the control of pesticide transfer and it led to the co supervision of E. Rouzies PhD, starting in Dec 2019.

6.2.3. 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. This has been extended to water pollutant tracking as part of Long Li's PhD and published in [13]

6.3. Model reduction / multiscale algorithms

6.3.1. Parameter space dimension reduction and Model order reduction

Participants: Mohamed Reda El Amri, Arthur Macherey, Youssef Marzouk, Clémentine Prieur, Alessio Spantini, Ricardo Baptista, Daniele Bigoni, Olivier Zahm.

Numerical models describing the evolution of the system (ocean + atmosphere) contain a large number of parameters which are generally poorly known. The reliability of the numerical simulations strongly depends on the identification and calibration of these parameters from observed data. In this context, it seems important to understand the kinds of low-dimensional structure that may be present in geophysical models and to exploit this low-dimensional structure with appropriate algorithms. We focus in the team, on parameter space dimension reduction techniques, low-rank structures and transport maps techniques for probability measure approximation.

In [25], we proposed a framework for the greedy approximation of high-dimensional Bayesian inference problems, through the composition of multiple low-dimensional transport maps or flows. Our framework operates recursively on a sequence of "residual" distributions, given by pulling back the posterior through the previously computed transport maps. The action of each map is confined to a low-dimensional subspace that we identify by minimizing an error bound. At each step, our approach thus identifies (i) a relevant subspace of the residual distribution, and (ii) a low-dimensional transformation between a restriction of the residual onto this sub-space and a standard Gaussian. We prove weak convergence of the approach to the posterior distribution, and we demonstrate the algorithm on a range of challenging inference problems in differential equations and spatial statistics.

The paper [34] introduces a novel error estimator for the Proper Generalized Decomposition (PGD) approximation of parametrized equations. The estimator is intrinsically random: It builds on concentration inequalities of Gaussian maps and an adjoint problem with random right-hand side, which we approximate using the PGD. The effectivity of this randomized error estimator can be arbitrarily close to unity with high probability, allowing the estimation of the error with respect to any user-defined norm as well as the error in some quantity of interest. The performance of the error estimator is demonstrated and compared with some existing error estimators for the PGD for a parametrized time-harmonic elastodynamics problem and the parametrized equations of linear elasticity with a high-dimensional parameter space.

In the framework of Arthur Macherey's PhD, we have also proposed in [26] algorithms for solving high-dimensional Partial Differential Equations (PDEs) that combine a probabilistic interpretation of PDEs, through Feynman-Kac representation, with sparse interpolation. Monte-Carlo methods and time-integration schemes are used to estimate pointwise evaluations of the solution of a PDE. We use a sequential control variates algorithm, where control variates are constructed based on successive approximations of the solution of the PDE. We are now interested in solving parametrized PDE with stochastic algorithms in the framework of potentially high dimensional parameter space. In [36], we consider 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. We are now working on the nonlinear generalization of active subspaces. Reduced models are also developed In the framework of robust inversion. In [43], we have combined a new greedy algorithm for functional quantization with a Stepwise Uncertainty Reduction strategy to solve a robust inversion problem under functional uncertainties. An ongoing work aims at further reducing the number of simulations required to solve the same robust inversion problem, based on Gaussian process meta-modeling on the joint input space of deterministic control parameters and functional uncertain variable. These results are applied to automotive depollution. This research axis is conducted in the framework of the Chair OQUAIDO.

6.4. Sensitivity analysis

Participants: Elise Arnaud, Eric Blayo, Laurent Gilquin, Maria Belén Heredia, Adrien Hirvoas, François-Xavier Le Dimet, Henri Mermoz Kouye, Clémentine Prieur, Laurence Viry.

6.4.1. Scientific context

Forecasting geophysical systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

6.4.2. Global sensitivity analysis

Participants: Elise Arnaud, Eric Blayo, Laurent Gilquin, Maria Belén Heredia, Adrien Hirvoas, Alexandre Janon, Henri Mermoz Kouye, Clémentine Prieur, Laurence Viry.

6.4.2.1. Global sensitivity analysis with dependent inputs

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. Recently, the Shapley value, from econometrics, was proposed as an alternative to quantify the importance of random input variables to a function. Owen [54] derived Shapley value importance for independent inputs and showed that it is bracketed between two different Sobol' indices. Song et al. [60] recently advocated the use of Shapley value for the case of dependent inputs. In a recent work [55], 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 [30].

6.4.2.2. Extensions of the replication method for the estimation of Sobol' indices

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 [45]

In [11] 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 [15]. It has allowed to perform a global sensitivity analysis with input space dimension more than eighty, without any screening preliminary step.

6.4.2.3. Green sensitivity for multivariate and functional outputs

Participants: María Belén Heredia, Clémentine Prieur.

Another research direction for global SA algorithm starts with the report that most of the algorithms to compute sensitivity measures require special sampling schemes or additional model evaluations so that available data from previous model runs (e.g., from an uncertainty analysis based on Latin Hypercube Sampling) cannot be reused. One challenging task for estimating global sensitivity measures consists in recycling an available finite set of input/output data. Green sensitivity, by recycling, avoids wasting. These given data have been discussed, e.g., in [59], [58]. Most of the given data procedures depend on parameters (number of bins, truncation argument...) not easy to calibrate with a bias-variance compromise perspective. Adaptive selection of these parameters remains a challenging issue for most of these given-data algorithms. In the context of María Belén Heredia's PhD thesis, we have proposed a non-parametric given data estimator for agregated Sobol' indices, introduced in [48] and further developed in [44] for multivariate or functional outputs. This last work should be submitted soon.

6.4.2.4. Global sensitivity analysis for parametrized stochastic differential equations

Participants: Henri Mermoz Kouye, 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 (GSA) 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 [10]. The research on GSA for stochastic simulators is still ongoing, first in the context of the MATH-AmSud project FANTASTIC (Statistical inFerence and sensitivity ANalysis for models described by sTochASTIC differential equations) with Chile and Uruguay, secondly through the PhD thesis of Henri Mermoz Kouye, co-supervised by Clémentine Prieur, in collaboration with INRA Jouy.

6.5. Model calibration and statistical inference

6.5.1. Bayesian calibration

Participants: Maria Belén Heredia, Adrien Hirvoas, Clémentine Prieur.

Physically-based avalanche propagation models must still be locally calibrated to provide robust predictions, e.g. in long-term forecasting and subsequent risk assessment. Friction parameters cannot be measured directly and need to be estimated from observations. Rich and diverse data is now increasingly available from test-sites, but for measurements made along ow propagation, potential autocorrelation should be explicitly accounted for. In the context of María Belén Heredia's PhD, in collaboration with IRSTEA Grenoble, we have proposed in a comprehensive Bayesian calibration and statistical model selection framework with application to an avalanche sliding block model with the standard Voellmy friction law and high rate photogrammetric images. An avalanche released at the Lautaret test-site and a synthetic data set based on the avalanche were used to test the approach. Results have demonstrated i) the effciency of the proposed calibration scheme,

and ii) that including autocorrelation in the statistical modelling definitely improves the accuracy of both parameter estimation and velocity predictions. 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. In the context of this work, it is necessary to develop algorithms of identification quantifying and ranking all the uncertainty sources. This work in done in collaboration with IFPEN.

6.5.2. Non-Parametric statistical inference 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 [39] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [40] (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 [40], in collaboration with J.R. Leon (Montevideo, Uruguay) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [41], 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 [42], 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 [52].

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 [38], and would like to extend that results to hypo-elliptic diffusions.

6.6. Modeling and inference for extremes

Participants: Philomène Le Gall, Clémentine Prieur, Patricia Tencaliec.

In [19], we are 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 Anne-Catherine Favre (LGGE-Lab in Grenoble) and Philippe Naveau (LSCE, Paris).

In the context of Philomène Le Gall's PhD thesis, we are applying the aforementioned modeling of extreme precipitation with the aim of regionalizing extreme precipitation.

6.7. Land Use and Transport Models Calibration

Participants: 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 [46] and Thomas Capelle defended his in April 2017 and published his latest results in [2].

ANGE Project-Team

7. New Results

7.1. Numerical methods for fluid flows

7.1.1. PARAOPT: A parareal algorithm for optimality systems

Member: J. Salomon,
Coll.: Martin Gander, Felix Kwok

The time parallel solution of optimality systems arising in PDE constraint optimization could be achieved by simply applying any time parallel algorithm, such as Parareal, to solve the forward and backward evolution problems arising in the optimization loop. We propose in [21] a different strategy by devising directly a new time parallel algorithm, which we call ParaOpt, for the coupled forward and backward non-linear partial differential equations. ParaOpt is inspired by the Parareal algorithm for evolution equations, and thus is automatically a two-level method. We provide a detailed convergence analysis for the case of linear parabolic PDE constraints. We illustrate the performance of ParaOpt with numerical experiments both for linear and nonlinear optimality systems.

7.1.2. Dynamical Behavior of a Nondiffusive Scheme for the Advection Equation

Member: N. Aguillon,
Coll.: Pierre-Antoine Guihéneuf

In [16], we study the long time behaviour of a dynamical system strongly linked to the anti-diffusive scheme of Després and Lagoutiere for the 1-dimensional transport equation. This scheme is overcompressive when the Courant–Friedrichs–Levy number is 1/2: when the initial data is nondecreasing, the approximate solution becomes a Heaviside function. In a special case, we also understand how plateaus are formed in the solution and their stability, a distinctive feature of the Després and Lagoutiere scheme.

7.1.3. Convergence of numerical schemes for a conservation equation with convection and degenerate diffusion

Member: C. Guichard,
Coll.: Robert Eymard, Xavier Lhébrard

In [20], the approximation of problems with linear convection and degenerate nonlinear diffusion, which arise in the framework of the transport of energy in porous media with thermodynamic transitions, is done using a θ-scheme based on the centered gradient discretisation method. The convergence of the numerical scheme is proved, although the test functions which can be chosen are restricted by the weak regularity hypotheses on the convection field, owing to the application of a discrete Gronwall lemma and a general result for the time translate in the gradient discretisation setting. Some numerical examples, using both the Control Volume Finite Element method and the Vertex Approximate Gradient scheme, show the role of θ for stabilising the scheme.

7.1.4. Gradient-based optimization of a rotating algal biofilm process

Members: N. Aguillon, J. Sainte-Marie,
Coll.: Pierre-Olivier Lamare, Jérôme Grenier, Hubert Bonnefond, Olivier Bernard

Microalgae are microorganisms that have only very recently been used for bio-technological applications and more specifically for the production of bio-fuel. In the report [15] we focus on the shape optimization and optimal control of an innovative process where microalgae are fixed on a support. They are successively exposed to light and darkness. The resulting growth rate can be represented by a dynamic system describing the denaturation of key proteins due to excess light. A Partial Derivative Equation (PDE) model for the Rotary Algae Biofilm (RAB) is proposed. It represents the local growth of microalgae subjected to time-varying light. A gradient method based on the calculation of the model adjoint is proposed to identify the optimal (constant) folding of the process and the (time-varying) speed of the biofilm. Once this method is used in a realistic case, the optimization results in a configuration that significantly improves productivity compared to the case where the biofilm is fixed.

7.2. Modelling

7.2.1. Accurate steam-water equation of state for two-phase flow LMNC model with phase transition

Member: Y. Penel,
Coll.: Stéphane Dellacherie, Bérénice Grec, Gloria Faccanoni

The paper [9] is dedicated to the design of incomplete equations of state for a two-phase flow with phase transition that are specific to the low Mach number regime. It makes use of the fact that the thermodynamic pressure has small variations in this regime. These equations of state supplement the 2D LMNC model introduced in previous works. This innovative strategy relies on tabulated values and is proven to satisfy crucial thermodynamic requirements such as positivity, monotonicity, continuity. In particular, saturation values are exact. This procedure is assessed by means of analytical steady solutions and comparisons with standard analytical equations of state, and shows a great improvement in accuracy.

7.2.2. Numerical simulations of Serre - Green-Naghdi type models for dispersive free surface flows

Members: Y. Penel, J. Sainte-Marie
Coll.: Enrique D. Fernandez-Nieto, Tomas Morales de Luna, Cipriano Escalante Sanchez

The Serre - Green-Nagdhi equations are simulated under their non-hydrostatic formulation by means of a projection-correction method. This is then extended to the layerwise discretisation of the Euler equations with a special care to the computational cost. An original alternating direction method is used and relies on the tools designed for the monolayer case.

7.2.3. Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow

Member: M. Parisot

The work [12] is devoted to the numerical resolution in the multidimensional framework of a hierarchy of reduced models of the water wave equations, such as the Serre-Green-Naghdi model. A particular attention is paid to the dissipation of mechanical energy at the discrete level, that act as a stability argument of the scheme, even with source terms such space and time variation of the bathymetry. In addition, the analysis leads to a natural way to deal with dry areas without leakage of energy. To illustrate the accuracy and the robustness of the strategy, several numerical experiments are carried out. In particular, the strategy is capable of treating dry areas without special treatment.

7.2.4. Congested shallow water model: on floating body

Members: E. Godlewski, M. Parisot, J. Sainte-Marie, F. Wahl

In [22], we are interested in the numerical modeling of body floating freely on the water such as icebergs or wave energy converters. The fluid-solid interaction is formulated using a congested shallow water model for the fluid and Newton's second law of motion for the solid. We make a particular focus on the energy transfer between the solid and the water since it is of major interest for energy production. A numerical approximation based on the coupling of a nite 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 verify the method and to show the feasibility of extending it to more complex cases.

7.2.5. Pseudo-compressibility, dispersive model and acoustic waves in shallow water flows

Members: E. Godlewski, M-O. Bristeau, J. Sainte-Marie

In this paper we study a dispersive shallow water type model derived from the compressible Navier-Stokes system. The compressible effects allow to capture the acoustic waves propagation and can be seen as a relaxation of an underlying incompressible model. Hence, the pseudo-compressibility terms circumvent the resolution of an elliptic equation for the non-hydrostatic part of the pressure. For the numerical approximation of shallow water type models, the hy- perbolic part, often approximated using explicit time schemes, is constrained by a CFL condition. Since the approximation of the dispersive terms – im- plicit in time – generally requires the numerical resolution of an elliptic equation, it is very costly. In this paper, we show that when considering the pseudo-compressibility terms a fully explicit in time scheme can be derived. This drastically reduces the cost of the numerical resolution of dispersive models especially in 2d and 3d.

7.2.6. Some quasi-analytical solutions for propagative waves in free surface Euler equations

Members: B. Di Martino, M-O. Bristeau, J. Sainte-Marie, A. Mangeney, F. Souillé

This note describes some quasi-analytical solutions for wave propagation in free surface Euler equations and linearized Euler equations. The obtained solutions vary from a sinusoidal form to a form with singularities. They allow a numerical validation of the free-surface Euler codes.

7.2.7. Challenges and prospects for dynamical cores of oceanic models across all scales

Members: E. Audusse, J. Sainte-Marie
Review paper, more than 30 co-authors

The paper [11] provides an overview of the recent evolution and future challenges of oceanic models dynamical cores used for applications ranging from global paleoclimate scales to short-term prediction in estuaries and shallow coastal areas. The dynamical core is responsible for the discrete approximation in space and time of the resolved processes, as opposed to the physical parameterizations which represent unresolved or under-resolved processes. The paper reviews the challenges and prospects outlined by the modeling groups that participated to the Community for the Numerical Modeling of the Global, Regional, and Coastal Ocean (COMMODORE) workshop. The topics discussed in the paper originate from the experience acquired during the development of 16 dynamical cores representative of the variety of numerical methods implemented in models used for realistic ocean simulations. The topics of interest include the choice of model grid and variables arrangement, vertical coordinate, temporal discretization, and more practical aspects about the evolution of code architecture and development practices.

7.2.8. The Navier-Stokes system with temperature and salinity for free surface flows Part I: Low-Mach approximation & layer-averaged formulation

Members: M-O. Bristeau, L. Boittin, A. Mangeney, J. Sainte-Marie, F. Bouchut

In this paper, we are interested in free surface flows where density variations coming e.g. from temperature or salinity differences play a significant role in the hydrodynamic regime. In water, acoustic waves travel much faster than gravity and internal waves, hence the study of models arising in compressible fluid mechanics often requires a decoupling between these waves. Starting from the compressible Navier-Stokes system, we derive the so-called Navier-Stokes-Fourier system in an incompressible context (the density does not depend on the fluid pressure) using the low-Mach scaling. Notice that a modified low-Mach scaling is necessary to obtain a model with a thermo-mechanical compatibility. The case where the density depends only on the temperature is studied first. Then the variations of the fluid density with respect to the temperature and the salinity are considered. We give a layer-averaged formulation of the obtained models in an hydrostatic context. Allowing to derive numerical schemes endowed with strong stability properties – that are presented in a companion paper – the layer-averaged formulation is very useful for the numerical analysis and the numerical simulations of the models. Several stability properties of the layer-averaged Navier-Stokes-Fourier system are proved.

7.2.9. The Navier-Stokes system with temperature and salinity for free surface flows - Part II: Numerical scheme and validation

Members: M-O. Bristeau, L. Boittin, A. Mangeney, J. Sainte-Marie, F. Bouchut

In this paper, we propose a numerical scheme for the layer-averaged Euler with variable density and the Navier-Stokes-Fourier systems presented in part I.These systems model hydrostatic free surface flows with density variations. We show that the finite volume scheme presented is well balanced with regards to the steady state of the lake at rest and preserves the positivity of the water height. A maximum principle on the density is also proved as well as a discrete entropy inequality in the case of the Euler system with variable density. Some numerical validations are finally shown with comparisons to 3D analytical solutions and experiments.

7.3. Functional analysis of PDE models in Fluid Mechanics

7.3.1. On the rigid-lid approximation of shallow water Bingham model

Member: J. Sainte-Marie
Coll.: Bilal Al Taki, Khawla Msheik

The paper [17] discusses the well posedness of an initial value problem describing the motion of a Bingham fluid in a basin with a degenerate bottom topography. A physical interpretation of such motion is discussed. The system governing such motion is obtained from the Shallow Water-Bingham models in the regime where the Froude number degenerates, i.e taking the limit of such equations as the Froude number tends to zero. Since we are considering equations with degenerate coefficients, then we shall work with weighted Sobolev spaces in order to establish the existence of a weak solution. In order to overcome the difficulty of the discontinuity in Bingham's constitutive law, we follow a similar approach to that introduced in [G. Duvaut and J.-L. Lions, Springer-Verlag, 1976]. We study also the behavior of this solution when the yield limit vanishes. Finally, a numerical scheme for the system in 1D is furnished.

7.3.2. Global bmo-1($\mathbb{R}^N$) radially symmetric solution for compressible Navier-Stokes equations with initial density in $\mathbb{L}^\infty(\mathbb{R}^N)$

Member: B. Haspot

In [24], we investigate the question of the existence of global weak solutionfor the compressible Navier Stokes equations provided that the initial momentum belongs to bmo-1($\mathbb{R}^N$) with $N = 2, 3$ and is radially symmetric. We prove then a equivalent of the so-called Koch-Tataru theorem for the compressible Navier-Stokes equations. In addition we assume that the initial density is only bounded in $\mathbb{L}^\infty(\mathbb{R}^N)$, it allows us in particular to consider initial density admitting shocks. 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 regularizinginasmuch as the density becomes Lipschitz. To finish we prove the global existence of strong solution for large initial data pro-vided that the initial data are radially symmetric and sufficiently regular in dimension $N = 2, 3$ for $\gamma-$law pressure.

7.3.3. New effective pressure and existence of global strong solution for compressible Navier-Stokes equations with general viscosity coefficient in one dimension

Member: Boris Haspot
Coll.: Cosmin Burtea

In this paper we prove the existence of global strong solution for the Navier-Stokesequations with general degenerate viscosity coefficients. The cornerstone of the proofis the introduction of a new effective pressure which allows to obtain an Oleinik-typeestimate for the so called effective velocity. In our proof we make use of additionalregularizing effects on the velocity which requires to extend the technics developedby Hoff for the constant viscosity case.

7.4. Assessments of models by means of experimental data and assimilation

7.4.1. Metamodeling corrected by observational data

Members: V. Mallet, J. Hammond

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, and then corrected this metamodel with observational data. Novel work was dedicated to the error modeling for a more balanced integration of the observations. The work was also applied to air quality simulation over Paris using several months of data.

7.4.2. Metamodeling of a complete air quality simulation chain

Members: A. Lesieur, V. Mallet
Coll.: Ruiwei Chen

With the objective of uncertainty quantification, we worked 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.

7.4.3. Artificial neural networks for the modeling of air pollution

Member: V. Mallet

Air quality simulations at national, continental or global scales are subject to large uncertainties which are typically mitigated by data assimilation techniques. Another approach to improve the forecasts is to design an error model, learning from historical discrepancies between simulations and observations. Such a model was built using an artificial neural network trained with many meteorological and geographical data. Further studies showed that the technique could successfully generate not only an error model (to improve pre-existing simulations), but also a complete model (without the need for pre-existing simulations) whose forecasts are more accurate than those of traditional models.

7.4.4. Uncertainty quantification in atmospheric dispersion of radionuclides

Members: V. Mallet,
Coll.: Irène Korsakissok

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 carried out the calibration of the metamodel input distributions using Markov chain Monte Carlo.

7.4.5. Meta-modeling for urban noise mapping

Members: A. Lesieur, V. Mallet
Coll.: Pierre Aumond, Arnaud Can

Noise computing software can require several hours to produce a map over an urban center for a given set of input data. This computational cost makes the models unsuitable for applications like uncertainty quantification or data assimilation where thousands of simulations, or more, can be required. One solution is to replace the physical model with a meta-model which is very fast and yet fairly reproduces the results of the physical model. The strategy is first to reduce the dimension of both inputs and outputs of the physical model, which leads to a reduced model. This reduced model is then replaced by a statistical emulator. The emulator is trained with calls to the reduced model for a set of chosen inputs. The emulator relies on the interpolation between the training output values.

7.4.6. Data assimilation for urban noise maps generated with a meta-model

Members: A. Lesieur, V. Mallet
Coll.: Pierre Aumond, Arnaud Can

In an urban area, it is increasingly common to have access to both a simulated noise map and a sensor network. A data assimilation algorithm is developed to combine data from both a noise map simulator and a network of acoustic sensors. One-hour noise maps are generated with a meta-model fed with hourly traffic and weather data. The data assimilation algorithm merges the simulated map with the sound level measurements into an improved noise map. The performance of this method relies on the accuracy of the meta-model, the input parameters selection and the model of the error covariance that describes how the errors of the simulated sound levels are correlated in space. The performance of the data assimilation is obtained with a leave-one-out cross-validation method.

7.4.7. Uncertainty quantification in wildland fire propagation

Members: F. Allaire, V. Mallet
Coll.: Jean-Baptiste Filippi

We worked further on the Monte Carlo simulation of wildland fires. We calibrated the input distributions that represent the uncertainties in the inputs of our fire spread predictions by using the observations of the final contours for a number of fire cases. We used a new metric to measure the dissimilarity between two burned sufaces that relies on the Wasserstein distance. We designed a metamodel and carried out the calibration of the model input distributions using Markov chain Monte Carlo.

7.4.8. A non-intrusive reduced order data assimilation method applied to the monitoring of urban flows

Member: J. Hammond
Coll.: R. Chakir

In [13], we investigate a variational data assimilation method to rapidly estimate urban pollutant concentration around an area of interest using measurement data and CFD based models in a non-intrusive and computationally efficient manner. In case studies presented here, we used a sample of solutions from a dispersion model with varying meteorological conditions and pollution emissions to build a Reduced Basis approximation space and combine it with concentration observations. The method allows to correct for unmodeled physics, while significantly reducing online computational time.

7.5. Software Devlopments

Members: C., A. El Baz, J. Sainte-Marie

Several improvements of FreshKiss3D software have been made:

1. the code can now be used on RPM-based systems (Fedora, Redhat) thanks to a user contribution (Julien Jerphanion);
2. the conda-based installation steps are now automatically tested by means of a continuous integration process performed on Linux and Mac virtual machines provided by the Inria platform https://ci.inria.fr/;
3. third-party libraries have been updated so as to benefit from their very latest features;
4. software quality is now monitored by static analysis via the Inria SonarQube platform (https://sonarqube.inria.fr/);
5. code optimization including parallelization with MPI is currently under development.

CASTOR Project-Team

4. New Results

4.1. On the identification of the electron temperature profile from polarimetry Stokes vector measurements in Tokamak free-boundary equilibrium reconstruction

Participants: Blaise Faugeras, Francesco Orsitto.

This paper reports numerical investigations on the identification of the electron temperature profile Te from interferometry and polarimetry Stokes vector measurements with the equilibrium code NICE (Newton direct and Inverse Computation for Equilibrium). This latter enables the consistent resolution of the inverse equilibrium reconstruction problem in the framework of nonlinear free-boundary equilibrium coupled to the Stokes model equation for polarimetry. We find that for ITER plasma with high Ip, Ne and Te the identification from noisy measurements is possible (Project EUROfusion / WP01Jet Campaigns (WPJET1) .

4.2. Equilibrium reconstruction for the JT-60SA tokamak

Participant: Blaise Faugeras.

Twin experiments were performed with this tokamak geometry in the framework of the EUROfusion / WP10 JT-60SA (WPSA) project.

4.3. Plasma boundary reconstruction for the ISTTOK tokamak

Participants: Blaise Faugeras, Rui Coelho, R. Santos.

Plasma boundary reconstruction is one of the main tools to provide a reliable control and tokamak performance. We explore the feasibility for the ISTTOK tokamak (Portugal) of a reconstruction method based on calculated vacuum magnetic flux map and plasma intersection with the wall. We show that via square wave input response curves and pre-processing of the poloidal field coil currents, it is possible to build for ISTTOK a simple scaling model for the effective equilibrium magnetic fields, and perform plasma boundary reconstruction using the algorithm VacTH. This algorithm, included in the NICE numerical code suite, relies on the decomposition of the poloidal flux in toroidal harmonics. The reconstructed plasma boundary is shown for a given discharge and its shape and position are shown to evolve consistently with the typical timescale evolution of ISTTOK discharges. This provides an opportunity of using this plasma boundary reconstruction method as a diagnostic tool for ISTTOK.

4.4. Implementation of a method enabling error bar computations for all reconstructed equilibrium quantitites

Participant: Blaise Faugeras.

Error bars on control variables are directly given by the inverse of the Hessian of the minimized cost function. This is not the case for other quantities such as the safety factor profile for example, and the computation of error bars for these important output quantities necessitate the non-trivial computation of their (discrete) derivatives with respect to the control variables as wall as the state variables.

4.5. New developments on the code NICE

Participant: Blaise Faugeras.

Developments have been done on the code NICE:

- Implementation of a mode 'without plasma' for magnetostatic computations.
- Implementation of pressure constraints in NICE for IMAS tested on JET data.
- Regular updates of the relased NICE actor in IMAS (EUROfusion / WP13 Code Development for Integrated Modeling)
- A Matlab interface has been developped to run the free boundary direct, evolutive, and inverse static modes of NICE.

4.6. Automatic identification of the plasma equilibrium operating space in tokamaks

Participants: Blaise Faugeras, Xia Song, Eric Nardon, Holger Heumann.

In order to identify the plasma equilibrium operating space for future tokamaks, a new objective function is introduced in the inverse static free-boundary equilibrium code FEEQS.M. This function comprises terms which penalize the violation of the central solenoid and poloidal field coils limitations (currents and forces). The penalization terms do not require any weight tuning. Hence, this new approach automates to a large extent the identification of the operating space. As an illustration, the new method is applied on the ITER 15 and 17MA inductive scenarios, and similar operating spaces compared to previous works are found. These operating spaces are obtained within a few ($\sim$3) hours of computing time on a single standard CPU.

4.7. Automating the design of Tokamak experiment scenarios

Participants: Jacques Blum, Holger Heumann.

The real-time control of plasma position, shape and current in a tokamak has to been sured 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 HL-2M highlight the flexibility and broad scope of the proposed optimal control formulation.

4.8. Coupling NICE-METIS

Participants: Jean François Artaud, Jacques Blum, Cédric Boulbe, Blaise Faugeras.

The free boundary equilibrium code NICE has been coupled to the fast transport solver METIS in a Matlab worflow. A first test case has been proposed on ITER. This work has been done for the project Eurofusion WPCD.

4.9. Advances in high order mixed finite elements for Maxwell's equations

Participant: Francesca Rapetti.

The implementation of high order curl- or div-conforming finite element spaces is quite delicate, especially in the three-dimensional case. I have worked on an implementation strategy, which has been applied in the open source finite element software FreeFem++. In particular, I have used the inverse of a generalized Vandermonde matrix to build a basis of generators in duality with the degrees of freedom, which then provides in FreeFem++ an easy-to-use but powerful interpolation operator. With Marcella Bonazzoli, now at the Inria Team DEFI in Saclay, I have carefully addressed the problem of applying the same Vandermonde matrix to possibly differently oriented tetrahedra of the mesh over the computational domain. [17]

High order mixed finite element spaces generally lack natural choices of bases but they do have spanning families. I have worked on these FEs for simplicial meshes and proven theoretically their effectiveness. I have also commented on some aspects of a new set of degrees of freedom, the so-called weights on the small simplices, to represent discrete functions in these spaces [11].

4.10. Construction of divergence-free bases

Participants: Francesca Rapetti, Ana Alonso Rodriguez.

I have worked to propose and analyze an efficient algorithm for the computation of a basis of the space of divergence-free Raviart-Thomas finite elements. The algorithm is based on graph techniques. The key point is to realize that, with very natural degrees of freedom for fields in the space of Raviart-Thomas finite elements of degree r + 1 and for elements of the space of discontinuous piecewise polynomial functions of degree $r \geq 0$, the matrix associated with the divergence operator is the incidence matrix of a particular graph. By choosing a spanning tree of this graph, it is possible to identify an invertible square submatrix of the divergence matrix and to compute easily the moments of a field in the space of Raviart-Thomas finite elements with assigned divergence. The analyzed approach is then used to construct a basis of the space of divergence-free Raviart-Thomas finite elements. The numerical tests show that the performance of the algorithm depends neither on the topology of the domain nor or the polynomial degree r [16].

4.11. First steps to polytopal/polyhedral meshes

Participant: Francesca Rapetti.

Merging ideas from compatible discretisations and polyhedral methods, I have worked with D. Di Pietro and J. Droniou to construct novel fully discrete polynomial de Rham sequences of arbitrary degree on polygons and polyhedra. The spaces and operators that appear in these sequences are directly amenable to computer implementation. Besides proving exactness, we have shown that the usual sequence of Finite Element spaces forms, through appropriate interpolation operators, a commutative diagram with other proposed sequence, which ensures suitable approximation properties. A discussion on reconstructions of potentials and discrete L2-products completes the work [14].

4.12. C^1 finite elements on triangular meshes

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. However due to the appearence of fourth order terms in the PDE systems that we are interested in, pure Galerkin methods require the use of finite element methods with C^1 continuity. The PhD thesis of Ali Elarif that has begun in october 2017 is devoted to the study of these methods for complex PDE models encountered in plasma physics. Relying on the work previously done on steady elliptic PDE, this year we applied these finite element methods to some evolution problems like the incompressible Navier-Stokes and MHD equations in stream-function formulation. Error estimates in H^2 norms have been obtained using standard finite element techniques. The simulation of some instabilities encountered in plasma physics have been done with very satisfactory results.

4.13. Modelling of acoustic streaming

Participants: Hervé Guillard, Argyris Delis [TUC, Crete].

Acoustic streaming is a particularly interesting example of the interaction of phenomena occuring on two different time scales. From a practical point of view, it is mainly used to generate a slow motion in micro-fluidic devices by means of high frequency acoustic sources. Modelling of these interaction is a challenge : taking into account the high frequency phenomena is prohibitively expensive but on the other hand, there is no universal agreement on existing averaged models. In order to have reference simulations, we have constructed a numerical code solving the compressible Navier-Stokes equations with high-order accuracy using compact schemes. Comparison with asymptotic analytical results has been done and shows that the code is able to simulate acoustic waves propagation in a stable way on long time scale, a property that is essential for the study of this phenomenon.

4.14. Mortar finite element methods

Participants: Hervé Guillard, Francesca Rapetti.

Hermite-Bezier finite element modeling is the standard method used to discretize the MHD equations in codes such as JOREK. This finite element family allows for an accurate description of the magnetic topology using flux aligned grids where the iso-parametric curved elements match the magnetic flux level sets. However, the description of complex material geometries is difficult with this family of finite element. We have begun to study the use of discretization methods using overlapping meshes where one mesh is composed of quadrangular Hermite-Bezier finite element while the second one is made of triangular elements.

4.15. Collisions in gyrokinetic equation

Participants: Afeintou Sangam, Vladimir T. Tikhonchuk.

Charged particles in plasma in strong magnetic fields undergo a complicated motion, which is a combination of a fast cyclotron gyration around the magnetic field lines and a relatively slow dynamics along and across the magnetic field lines. Gyrokinetic equations, devised to describe plasma under such conditions, eliminate the fast cyclotron gyration from the equation of motion, thus reducing the space-velocity phase space dimension from six to five.

Originally, the gyrokinetic formulation was devised for a collisionless plasmas. The quest for retaining collisions in gyrokinetic equations is ongoing. Collisions are important if one wants to describe the transport properties of a magnetized plasma on a macroscopic level. A description of the transport of energy and momentum was proposed in Refs. [18], [20], [19], [22], [21]. However, mathematical description of collisions in these works is too complicated for numerical implementation. We develop a simplified description of collision operators in the gyrokinetic formulation that preserve the pertinent conservation features and suitable for numerical modeling. A comparison of these operators with several test cases is under investigation.

4.16. Singular solutions of dispersive systems

Participants: S. Gavrilyuk, B. Nkonga, K-M Shyue, L. Truskinovsky.

We study a dispersive regularization of p-system. The governing equations are the Euler- Lagrange equations for a Lagrangian depending not only on the velocity and density, but also on the first material derivative of density. Such regularization arises, in particular, in the modeling of waves in solids, in bubbly fluids as well as in the theory of water waves. We show that such terms are not always regularizing. The solution can develop shocks even in the presence of dispersive terms. In particular, we construct such a shock solution that connects a constant state to a periodic wave train. The corresponding shock speed coincides with the velocity of the wave train. The generalized Rankine-Hugoniot relations (jump relations) are also obtained. The numerical evidence of the existence of such shocks is demonstrated in the case of the Serre-Green-Naghdi equations describing long surface water waves. In particular, it has been shown that such waves can dynamically be formed

4.17. A path conservative finite volume method for a shear shallow water model

Participants: P. Chandrashekar, B. Nkonga, A. K. Meena, A. Bhole.

The shear shallow water model provides a higher order approximation for shallow water flows by including the effect of vertical shear in the model. This model can be derived from the depth averaging process by including the second order velocity fluctuations, which are neglected in the classical shallow water approximation. The resulting model has a non-conservative structure, which resembles the 10-moment equations from gas dynamics. This structure facilitates the development of path conservative schemes and we construct HLL, 3-wave and 5-wave HLLC-type solvers. An explicit and semi-implicit MUSCL-Hancock type second order scheme is proposed for the time integration. Several test cases including roll waves show the performance of the proposed modeling and numerical strategy.

4.18. Full MHD Modeling of Shattered Pellet Injection

Participants: B. Nkonga, P. Chandrashekar, A. Bhole.

To avoid disruptions, the first thing to do is to operate as far as possible from disruptions operational limits. It means that plasma scenarios must be designed taking these limits into account. The challenge is to deal with peeling-ballooning instabilities called Edge Localized Modes (ELMs) which are characterized by the quasi-periodic relaxation of the pressure pedestal profile which results in the expelling of particles and energy from the bulk plasma to the edge. Injecting of impurities is one of the solutions to change the pedestal profile and mitigate MHD instabilities. The current design of the ITER DMS(Disruption Mitigation System) is a hybrid system using Massive Gas Injection (MGI) and Shattered Pellet Injection (SPI), methods which have demonstrated their efficiency on current tokamaks (JET, DIIID, ...). Considering that the plasma is composed of impurities, main ion core and set of electrons, premixed “Full MHD” formulation has been proposed. This model assumes that, for any control volume, the plasma is locally neutral and at the thermal and coronal equilibrium. Properties of this model are under analysis, according to the tabulated equation of state. A numerical approximation in the Jorek Code is also under progress. This work has been done in the context of the JET program 2019.

COFFEE Project-Team (section vide)

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 [23]. This study has been performed within the PhD thesis of Musaab Khalid 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 [39] and applied on a laboratory fume-hood [26], [30], [43]. This study demonstrated the feasibility of conducting on-site large-scale image-based measurements for indoor airflows characterization. The technique was also assessed in an outdoor flow

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 via ensemble assimilation scheme. We have proposed a novel method for volumetric velocity reconstruction exploring the locality of 3D object space. Under this formulation the velocity of local patch was sought to match the projection of the particles within the local patch in image space to the image recorded by camera. The core algorithm to solve the matching problem is an instance-based estimation scheme that can overcome the difficulties of optimization originated from the nonlinear relationship between the imageintensity residual and the volumetric velocity. The proposed method labeled as Lagrangian Particle ImageVelocimetry (LaPIV) is quantitatively evaluated with synthetic particle image data. The promising results indicated the potential application of LaPIV to a large variety of volumetric velocity reconstruction problems [44].

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 [21]. 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 [29].

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 is pursued within a strong collaboration with the industrial partner: SCALIAN

6.2.4. Learning of the dynamics of large scale geophysical systems using semi-group theory for data assimilation

Participants: Etienne Mémin, Gilles Tissot.

The goal of this study is to propose new ensemble data assimilation methodologies to estimate oceanic and turbulent flows. In classical methods, from a distribution of initial conditions, an ensemble of simulations are computed and used for estimation. Ideally, from this solution, a new ensemble has to be generated to refine the estimation. However, due to large numerical costs and operational constraints, this iterative procedure is in practice intractable. In order to improve actual performances, we propose to take these limitations into account and to develop new methodologies able to better take advantage of the information contained in the ensemble and in the dynamical model. More precisely, we propose to learn the non-linear dynamical features of the system and to be able to reproduce it without having to run a new simulation. The formalism is based on two concepts: i) the reproducing kernel Hilbert spaces (RKHS) that are a basis of smooth functions in the phase space giving interpolatory properties ii) the Koopman operator, that is an infinite-dimensional operator able to propagate in time any observable of the phase space. These two elements allow to define a rigorous framework in which hypothesis classically done in ensemble methods appear naturally. Thus, classical methods enter in a special case of this new formalism, that allows us to generalise them in a way to improve the learning of the non-linear dynamical system. Numerical tests are performed using the Ginzburg-Landau equation and a quasi-geostrophic flow model.

6.2.5. 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, Pranav Chandramouli, Long Li, Etienne Mémin.

In this research axis we have devised a principle to derive representation of flow dynamics under location 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. Rigorously derived from a stochastic version of the Reynolds transport theorem [9], this framework, referred to as modeling under location uncertainty (LU), 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 LU 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 LU leads to remarkable estimation of the unresolved errors opposite 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). For a wind forced QG model in a square box, which is an idealized model of north-Atlantic circulation, we have shown that for different versions of the noise the QG LU model leads to improve long-terms statistics when compared to classical large-eddies simulation strategies. For a QG model we have demonstrated that the LU model allows conserving the global energy. We have also shown numerically that Rosby waves were conserved and that inhomogeneity of the random component triggers secondary circulations. This feature enabled us to draw a formal bridge between a classical system describing the interactions between the mean current and the surface waves and the LU model in which the turbophoresis advection term plays the role of the classical Stokes drift.

Supported by funding from Inria-Mitacs Globalink, we hosted Ruediger Brecht, PhD student at Memorial University of Newfoundland, Canada, for a period of 3 months (May to August) in the Fluminance group. During his stay, Ruediger Brecht worked on the incorporation of a stochastic representation of the small-scale velocity component of a fluid flow in a variational integrator for the rotating shallow-water equations on the sphere, already developed within the first part of its PhD work. This work was based on an ongoing study in the group on a stochastic Quasi-geostrophic model and followed a series of works performed in the Fluminance group to define stochastic geophysical flow dynamics.

6.3.2. Large eddies simulation models under location uncertainty

Participants: Mohamed Yacine Ben Ali, Pranav Chandramouli, Dominique Heitz, Etienne Mémin, Gilles Tissot.

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 3900 and wall-bounded flows. 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 conservation laws of the stochastic fluid dynamics investigated. 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. Stochastic compressible fluid dynamics

Participants: Etienne Mémin, Gilles Tissot.

Some work has been performed to extend the stochastic formulation under location uncertainty to compressible flows. The interest is to extend the formulation on the one hand to compressible fluids (for instability mechanisms involved in areoacoustics for instance, or for thermal effects in mixing layers) and on the other hand to geophysical flows where the Boussinesq equation is not valid anymore (density variations due to temperature or salinity gradients). A theoretical study has been performed that opens the door to numerical validations. In particular a baroclinic torque term has been identified that could have major effects in some situations.

6.3.5. Stochastic hydrodynamic stability under location uncertainty

Participants: Etienne Mémin, Gilles Tissot.

In order to predict instability waves propagating within turbulent flows, eigenmodes of the linearised operator is not well suited since it neglects the effect of turbulent fluctuations on the wave dynamics. To cope this difficulty, resolvent analysis has become popular since it represents the response of the linearised operator to any forcing representing the generalised stress tensors. The absence of information on the non-linearity is a strong limitation of the method. In order to refine these models, we propose to consider a stochastic model under location uncertainty expressed in the Fourier domain, to linearise it around the corrected mean-flow and to study resulting eigenmodes. The stochastic part represents the effect of the turbulent field onto the instability wave. It allows to specify a structure of the noise and then to improve existing models. Improvements compared to the resolvent analysis have been found for turbulent channel flow data at $\mathfrak{R}_\tau = 180$. This work is in collaboration with André Cavalieri (Instituto Tecnologico de Aeronautica, SP, Brésil).

6.3.6. 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 published in JMAA [24]

In the framework of the collaboration with the University of Pisa (Italy), namely with Luigi Berselli 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. This is a very nice generalisation to non stationnary source terms of a famous results by Foais. IN the same work, we also 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. These results have been published in Nonlinearity [19]. These results have been extended in the framework of POD for reduced models in [18].

In [55] we have shown the existence of a solution to a 1D Reynolds Averaged Navier-Stokes vertical model suitable in the atmospheric boundary layer, under suyitable assumption on the data. The paper is received for publication in thje journal Pure and Applied Functional Analysis (PAFA).

We also have introduced a turbulence model including a backscatter term, which has the same structure as the Voigt model. The additional term is derived in certain specific regimes of the flow, such as the convergence to stable statistical states. We get estimates for the velocity v in $L^\infty_t H^1_x \cap W^{1,2}_t H^{1/2}_x$, that allow us to prove the existence and uniqueness of a regular-weak solutions (v, p) to the resulting system, for a given fixed eddy viscosity. We then prove a structural compactness result that highlights the robustness of the model. This allows us to pass to the limit in the quadratic source term in the equation for the turbulent kinetic energy k, which yields the existence of a weak solution to the corresponding Reynolds Averaged Navier-Stokes system satisfied by (v, p, k). These results are writen in [47], a paper which is under revision in Non Linear Analysis.

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 l = l(Re6, z), Re6 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.

6.3.7. 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 [25][17].

6.3.8. 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. The third challenge was to carry out a time resolved tomoPIV experiments in a turbulent wake flow. This work has been submitted to the Journal of Computational Physics.

6.3.9. 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, enables to 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 published in the journal "Experiments in Fluids" [21].

6.4. Visual servoing approach for fluid flow control

6.4.1. A state space representation for the closed-loop control of shear flows

Participants: Johan Carlier, Christophe Collewet.

The goal of this study is to develop a generic state representation for the closed-loop control of shear flows. We assume that the actuator acts at the boundaries. Our approach is based on a linearization of the Navier-Stokes equations around the desired state. Particular care was paid to the discrete approximation of the linear model to design a well-conditioned and accurate state matrix describing time evolution of disturbances evolving in parallel shear flow as long as these disturbances remain sufficiently small. A state matrix representation is obtained for the periodic channel flow and the spatially developing mixing layer flow. This approach has been validated through the representativity of our model in terms of linear stability. This work has been presented to the French Mechanics Congress CFM'2019 (https://hal.inria.fr/hal-02283161) [36].

6.4.2. Closed-loop control of a spatially developing shear layer

Participants: Christophe Collewet, Johan Carlier.

This study aims at controling one of the prototypical flow configurations encountered in fluid mechanics: the spatially developing turbulent shear layer occuring between two parallel incident streams with different velocities. Our goal is to maintain the shear-layer in a desired state and thus to reject upstream perturbations. 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, in term of magnitude as well as in term of spatial dispersion. 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 (magnitude and spatial dispersion) has been developed to lead to an adaptive control law. Simple conditions to ensure the local asymptotic stability of the whole scheme have been derived. This work has been also presented to the French Mechanics Congress CFM'2019 (https://hal.archives-ouvertes.fr/hal-02189111) [37].

6.4.3. Design of a DBD plasma actuator for closed-loop control

Participants: Johan Carlier, Christophe Collewet.

The goal of this study is to design a DBD plasma actuator for closed-loop control. This kind of actuator is widely used in the flow control community however, it is more appropriate to force a flow than to control it. Indeed, to control a flow under a closed-loop fashion, the action must be proportional to the control signal provided by the control law. It is unfortunately not the case with these actuators. We have modified the classical DBD plasma actuator so that the action is almost a linear fonction of the control signal. Our approach have been validated by a prototype and by first experiments.

6.5. Coupled models in hydrogeology

6.5.1. Reactive transport in multiphase flow

Participant: Jocelyne Erhel.

Groundwater resources are essential for life and society, and should be preserved from contamination. Pollutants are transported through the porous medium and a plume can propagate. Reactive transport models aims at simulating this dynamic contamination by coupling advection dispersion equations with chemistry equations. If chemistry is at thermodynamic equilibrium, then the system is a set of partial differential and algebraic equations (PDAE). Space discretization leads to a semi-discrete DAE system which should be discretized in time. An explicit time scheme allows an easy decoupling of transport and chemistry, but very small timesteps should be taken, leading to a very large CPU time. Therefore, an implicit time scheme is preferred, coupling transport and chemistry in a nonlinear system. The special structure of linearized systems can be used in preconditioned Newton-Krlyov methods in order to improve efficiency. Some experiments illustrate the methodology and show also the need for an adaptive timestep and a control of convergence in Newton's iterations.

This work was presented at a workshop [31].

6.5.2. Characterizations of Solutions in Geochemistry at equilibrium

Participant: Jocelyne Erhel.

Geochemistry at thermodynamic equilibrium involves aqueous reactions and mineral precipitation or dissolution. Quantities of solute species are assumed to be strictly positive, whereas those of minerals can vanish. The mathematical model is expressed as the minimization of Gibbs energy subject to positivity of mineral quantities and conservation of mass. Optimality conditions lead to a complementarity problem. We show that, in the case of a dilute solution, this problem can also be considered as optimality conditions of another minimization problem, subject to inequality constraints. This new problem is easier to handle, both from a theoretical and a practical point of view. Then we define a partition of the total quantities in the mass conservation equation. This partition builds a precipitation diagram such that a mineral is either precipitated or dissolved in each subset. We propose a symbolic algorithm to compute this diagram. Simple numerical examples illustrate our methodology.

This work was published in the journal Computational Geosciences [22] and presented at an international conference [38].

6.5.3. Mathematical models of kinetic reactions in geochemistry

Participants: Jocelyne Erhel, Bastien Hamlat.

In geochemistry, kinetic reactions can lead to the appearance or disappearance of minerals or gas. We defined two mathematical models based first on a differential inclusion system and second on a projected dynamical system. We proposed a regularization process for the first model and a projection algorithm for the second one.

This work, supported by IFPEN, was presented at a conference [39] and a workshop [32].

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 for 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) [31].

LEMON Project-Team

6. New Results

6.1. Inland flow processes

6.1.1. Shallow water models with porosity

We propose in [10] a discussion on the publication 'Dam break in rectangular channels with different upstream-downstream widths' (Valiani and Caleffi, 2019). The authors consider an augmented shallow water system for modelling the dam-break problem in a channel with discontinuous width and present its analytical solutions depending on the upstream-downstream water depth and channel's width ratios. In this discussion we contest the conservation of the hydraulic head through the width's discontinuity, which is stated by the authors, and we exemplify it by performing 2D Shallow water simulations reproducing some test cases presented in the paper.

6.1.2. Forcing

A book chapter entitled *Space-time simulations of extreme rainfall : why and how ?* involving among others two members of the team, Vincent Guinot and Gwladys Toulemonde has been written and accepted for publication [6]. The book whose title is *Mathematical Modeling of Random and Deterministic Phenomena* will be published by Wiley. This chapter aims to present practical interest of doing space-time simulations of extreme rainfall and to propose a state-of-art about that.

6.2. Marine and coastal systems

6.2.1. Numerical Modelling of Hydrokinetic Turbines

Recent studies have pointed out the potential of several coastal or river areas to provide significant energy resources in the near future. However, technological processes for extracting energy using Marine Current Energy Converters (MCEC) are not generically "field-ready" and still require significant research to be set up. The book chapter [8] comes within this framework: we develop the numerical model OceaPoS, useful to carry out a comprehensive description of turbulent flow patterns past MCEC and forward optimize the turbine arrays configurations and evaluate their environmental effects. The OceaPos model consists in describing the fluid as an ensemble of Lagrangian particles ruled by a Stochastic process. OceaPos follows the same methodology than SDM-WindPoS model for wind farm simulations and adapts the Lagrangian stochastic downscaling method (SDM) to the tidal and oceanic boundary layer. We also introduce a Lagrangian version of actuator discs to take account of one or several MCEC's devices and their effects on the flow dynamics. Several benchmarks are presented, and numerical predictions are compared to experimental results.

6.2.2. Multi-scale ocean modeling

In [3], we derive 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 use the derived transparent boundary conditions as interface conditions in a domain decomposition method, where they become local in time. We analyze numerically their efficiency thanks to comparisons made with other interface conditions.

In Cemracs 2019 in Marseille, Joao CALDAS was enrolled in the project "Model analysis for tsunami generation by landslides", with Louis EMERALD (PhD student, Université de Rennes), Emmanuel AUDUSSE (maître de conférences, Université Paris 13), Martin PARISOT (chargé de recherche, Inria Bordeaux, CARDAMOM team), Philippe HEINRICH (researcher, CEA) and Alexandre PARIS (PhD student, CEA). The project, funded by CEA, aims to study and compare different fluid mechanics models (Navier-Stokes, Boussinesq, Shallow Water equations) in the simulation of waves generated by landslides. The observed behaviour of the models is correlated to the amount of energy transferred from the sediments to the fluid, both in the wave generation zone (next to the landslide) and in the wave propagation zone (far away from it). An inverse problem is proposed for recovering the landslide from a given evolution of the free surface elevation. A publication will appear in the proceedings of CEMRACS 2019.

6.3. Stochastic models for extreme events

6.3.1. Hierarchical space-time modeling of exceedances

This novel approach is presented in this subsection but it is important to note that it also could has been presented in the subsection *Forcing* because the proposed method could be used as rainfall forcing and because it answers to some mentionned challenges.

The statistical modeling of space-time extremes in environmental applications is a valuable approach to understand complex dependencies in observed data and to generate realistic scenarios for impact models. Motivated by hourly rainfall data in Southern France presenting asymptotic independence, we propose in a joint work (J.N. Bacro, C. Gaetan, T. Opitz and G. Toulemonde) published in the Journal of the ASA [2] a novel hierarchical model for high threshold exceedances 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 also been presented by Gwladys Toulemonde in 2019 in two invited talks (EVA, Zagreb 2019; CMStatistics, ERCIM, London 2019), in one contributed international conference (ISI, Kuala Lumpur, 2019) and one seminar organized by the "Ecole Polytechnique" (Paris, 2019).

6.3.2. Extension of the XGumbel copula to the spatial framework

An extension of the XGumbel copula to the spatial framework has been developped. This work has been presented in the international conference Extreme Value Analysis (EVA 2019, Zagreb) and is currently under review [9]. The XGumbel copula combines two Gumbel copulas with weight parameters, termed extra-parameters, taking values in the unit hyper-cube. In a multisite study, the copula dimension being the number of sites, the XGumbel copula quickly becomes over-parametrized. In addition, interpolation to ungauged locations is not easily achieved. We propose a spatial model for maxima that combines a spatial regression for GEV marginals built with a vector generalized linear model and the spatialized XGumbel copula defined thanks to a spatial mapping for the extra-parameters. The mapping is designed shaped as a disk according to bivariate properties of the XGumbel copula. An Approximate Bayesian Computation (ABC) scheme that seeks to reproduce upper tail dependence coefficients for distance classes is used to infer the parameters. The extension of the XGumbel copula to the spatial framework has been used to study annual maxima of daily precipitation totals at 177 gauged stations over a 57 year period in the French Mediterranean.

MAGIQUE-3D Project-Team

7. New Results

7.1. High-order numerical methods for modeling wave propagation in complex media: development and implementation

7.1.1. High order discretization of seismic waves-problems based upon DG-SE methods

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 DG-SEm numerical method for solving time-dependent elasto- acoustic wave problems. We consider the first-order coupled formulation for which we propose a DG-SEm formulation which turns out to be stable.

While the 2D case is almost direct, the 3D case requires a particular attention on the coupling boundary on which it is necessary to manage the possible positions of the faces of the tetrahedrons with respect to that of the neighboring hexaedra.

In the framework of this DG-SEm coupling, we are also interested in the Perfectly Matched Layer (PML) in particular the use of the SEm inside it to stabilize it in cases where the use of DGm leads to instabilities.

These results have been obtained in collaboration with Henri Calandra (TOTAL) and Christian Gout (INSA Rouen) and have been presented at Journées Ondes Sud-Ouest (JOSO) in Le Barp, the 14th International Conference on Mathematical and Numerical Aspects of Wave Propagation (WAVES) in Vienna (Austria) and MATHIAS conference in Paris [21], [26]

7.1.2. Isogeometric analysis of sharp boundaries in full waveform inversion

Participants: Hélène Barucq, Julien Diaz, Stefano Frambati.

Efficient seismic full-waveform inversion simultaneously demands a high efficiency per degree of freedom in the solution of the PDEs, and the accurate reproduction of the geometry of sharp contrasts and boundaries. Moreover, it has been shown that the stability constant of the FWI minimization grows exponentially with the number of unknowns. Isogeometric analysis has been shown to possess a higher efficiency per degree of freedom, a better convergence in high energy modes (Helmholtz) and an improved CFL condition in explicit-time wave propagation, and it seems therefore a good candidate for FWI.

In the first part of the year, we have focused on a small-scale one-dimensional problem, namely the inversion over a multi-step velocity model using the Helmholtz equation. By exploiting a relatively little-known connection between B-splines ad Dirichlet averages, we have added the knot positions as degrees of freedom in the inversion. We have shown that arbitrarily-placed discontinuities in the velocity model can be recovered using a limited amount of degrees of freedom, as the knots can coalesce at arbitrary positions, obviating the need for a very fine mesh and thus improving the stability of the inversion.

In order to reproduce the same results in two and three dimensions, the usual tensor-product structure of B-splines cannot be used. We have therefore focused on the construction of (unstructured) multivariate B-spline bases. We have generalized a known B-spline basis construction through the language of oriented matroids, showing that multivariate spline bases can be easily constructed with repeated knots and that the construction algorithm can be extended to three dimensions. This gives the freedom to locally reduce the regularity of the basis functions and to place internal boundaries in the domain. The resulting mass matrix is block-dagonal, with adjustable block size, providing an avenue for a simple unstructured multi-patch DG-IGA scheme that is being investigated. With this goal in mind, more efficient quadrature schemes for multivariate B-splines exploiting the connection to oriented matroids are also being investigated.

A research report is in preparation.

7.1.3. 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 point source for core-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 (LDV). The point-source-LDV configuration is convenient to model numerically. It can also favor the incertitude estimate of the source and receiver locations. 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 high-order FEM schemes, opening the possibility of performing FWI on dense, high frequency and large band-width data.

This work is in collaboration with Clarisse Bordes, Daniel Brito, Federico Sanjuan and Deyuan Zhang (LFCR, UPPA) and with Stéphane Garambois (ISTerre). Ii is one of the topic of the PhD. thesis of Chengyi Shen.

7.1.4. Simulation of electro-seismic waves using advanced numerical methods

Participants: Hélène Barucq, Julien Diaz, Ha Howard Faucher, Rose-Cloé Meyer.

We study time-harmonic waves propagation in conducting poroelastic media, in order to obtain accurate images for complex media with high-order methods. In these kind of media, we observe the coupling between electromagnetic and seismic wave fields, which is called seismokinetic effect. The converted waves are very interesting because they are heavily sensitive to the medium properties, and the modeling of seismo-electric conversion can allow to detect interfaces in the material where the seismic field would be blind. To the best of our knowledge, the numerical simulation of this phenomenon has never been achieved with high-order finite element methods. Simulations are difficult to perform in time domain, because the time step and the mesh size have to be adapted to the huge variations of wave velocities. To ease the numerical implementation, we work in the frequency domain. We can then include physical parameters that depend non-linearly on the frequencies. Then, we have developed a new Hybridizable Discontinuous Galerkin method for discretizing the equations. This allows us to reduce the computational costs by considering only degrees of freedom on the skeleton of the mesh. We have validated the numerical method thanks to comparison with analytical solutions. We have obtained numerical results for 2D realistic poroelastic media and conducting poroelastic media is under investigation.

Results on analytical solutions for poroelasticity are presented in the research report [44].

7.1.5. Quasinormal mode expansion of electromagnetic Lorentz dispersive materials

Participants: Marc Duruflé, Alexandre Gras.

We have studied the electromagnetic scattering of optical waves by dispersive materials governed by a Drude-Lorentz model. The electromagnetic fields can be decomposed onto the eigenmodes of the system, known as quasinormal modes. In [51], a common formalism is proposed to obtain different formulas for the coefficients of the modal expansion. In this paper, it is also explained how to handle dispersive Perfectly Matched Layers and degenerate eigenvectors. Lately, we have investigated the use of an interpolation method in order to compute quickly the diffracted field for a large number of frequencies.

7.1.6. A Hybridizable Galerkin Discontinuous formulation for elasto-acoustic coupling in frequency-domain

Participants: Hélène Barucq, Julien Diaz, Vinduja Vasanthan.

We are surrounded by many solid-fluid interactions, such as the seabed or red blood cells. Indeed, the seabed represents the ocean floors immersed in water, and red blood cells are coreless hemoglobin-filled cells. Hence, when wanting to study the propagation of waves in such domains, we need to take into account the interactions at the solid-fluid interface. Therefore, we need to implement an elasto-acoustic coupling. Many methods have already tackled with the elasto-acoustic coupling, particularly the Discontinuous Galerkin method. However, this method needs a large amount of degrees of freedom, which increases the computational cost. It is to overcome this drawback that the Hybridizable Discontinuous Galerkin (HDG) has been introduced. The implementations of HDG for the elastic wave equations, as well as partially for the acoustic ones, have been done previously. Using these, we have performed in this work the elasto-acoustic coupling for the HDG methods in 1D, 2D and 3D. The results are presented in Vinduja Vasanthan's master's thesis [54].

7.1.7. Absorbing Radiation Conidition in elongated domains

Participants: Hélène Barucq, Sébastien Tordeux.

We develop and analyse a high-order outgoing radiation boundary condition for solving three-dimensional scattering problems by elongated obstacles. This Dirichlet-to-Neumann condition is constructed using the classical method of separation of variables that allows one to define the scattered field in a truncated domain. It reads as an infinite series that is truncated for numerical purposes. The radiation condition is implemented in a finite element framework represented by a large dense matrix. Fortunately, the dense matrix can be decomposed into a full block matrix that involves the degrees of freedom on the exterior boundary and a sparse finite element matrix. The inversion of the full block is avoided by using a Sherman–Morrison algorithm that reduces the memory usage drastically. Despite being of high order, this method has only a low memory cost. This work has been published in [13].

7.1.8. Discontinuous Galerkin Trefftz type method for solving the Maxwell equations

Participants: Margot Sirdey, Sébastien Tordeux.

Trefftz type methods have been developed in Magique 3D to solve Helmholtz equation and it has been presented in [25].These 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 is the topic of the PhD thesis of Margot Sirdey.

7.1.9. Reduced models for multiple scattering of electromagnetic waves

Participants: Justine Labat, Victor Péron, Sébastien Tordeux.

In this project, we develop fast, accurate and efficient numerical methods for solving the time-harmonic scattering problem of electromagnetic waves by a multitude of obstacles for low and medium frequencies in 3D. First, we consider a multi-scale diffraction problem in low-frequency regimes in which the characteristic length of the obstacles is small compared to the incident wavelength. We use the matched asymptotic expansion method which allows for the model reduction. Then, small obstacles are no longer considered as geometric constraints and can be modelled by equivalent point-sources which are interpreted in terms of electromagnetic multipoles. Second, we justify the Generalized Multiparticle Mie-solution method (Xu, 1995) in the framework of spherical obstacles at medium-frequencies as a spectral boundary element method based on the Galerkin discretization of a boundary integral equation into local basis composed of the vector spherical harmonics translated at the center of each obstacle. Numerically, a clever algorithm is implemented in the context of periodic structures allowing to avoid the global assembling of the matrix and so, reduce memory usage. The reduced asymptotic models of the first problem can be adapted for this regime by incorporating non-trivial corrections appearing in the Mie theory. Consequently, a change in variable between the two formulations can be made explicit, and an inherent advantage of the asymptotic formulation is that the basis and the shape can be separated with a semi-analytical expression of the polarizability tensors. A comparison of these different methods in terms of their accuracy has been carried out. Finally, for both methods and in the context of large numbers of obstacles, we implement an iterative resolution with preconditionning in a GMRES framework.

These results have been presented at Journées Ondes Sud-Ouest (JOSO) in Le Barp (France) and the 14th International Conference on Mathematical and Numerical Aspects of Wave Propagation (WAVES) in Vienna (Austria), see [22], [39]. Part of this work has been published in Wave Motion [17].

7.1.10. 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 published in COMPEL - The international journal for computation and mathematics in electrical and electronic engineering, [16].

7.1.11. Asymptotic Models and Impedance Conditions for Highly Conductive Sheets in the Time-Harmonic Eddy Current Model

Participant: Victor Péron.

This work is concerned with the time-harmonic eddy current problem for a medium with a highly conductive thin sheet. We present asymptotic models and impedance conditions up to the second order of approximation for the electromagnetic field. The conditions are derived asymptotically for vanishing sheet thickness where the skin depth is scaled like thickness parameter. The first order condition is the perfect electric conductor boundary condition. The second order condition turns out to be a Poincaré-Steklov map between tangential components of the magnetic field and the electric field. This work has been published in SIAM Journal on Applied Mathematics, [18].

7.2. Understanding the interior of the Earth and the Sun by solving inverse problems

7.2.1. Time-Domain Full Waveform Inversion based on high order discontinuous numerical schemes

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.

The adjoint state can be both defined as the adjoint of the continuous equation or the discrete problem. This choice is still under study and complementary results has been presented at WAVES 2019 conference in Vienna [38].

The FWI has been largely developed for time-harmonic wave problems essentially because of computational time which is clearly below the one of corresponding time-dependent problems. However, the memory cost in large 3D domain is overflowing the computer capabilities, which motivates us to develop a FWI algorithm in the time-domain. To fully exploit the information from the seismic traces, while preserving the computational cost, it is important to use an accurate and flexible discretization. For that purpose we study several time schemes such as Runge-Kutta 2/4 or Adam-Bashforth 3 and regarding the space discretization, we employ Discontinuous Galerkin (DG) elements which are well-known not only for their h and p adaptivities but also for their massively parallel computation properties.

In the work-flow of DIP a Reverse Time Migration (RTM) code has been developed in collaboration with Total using their Galerkin Discontinuous acoustic time domain solver. Then this code served as a prototype of the time domain FWI code called utFWI (Unstructured Time-Domain Full Waveform Inversion) (https://bil.inria.fr/fr/software/view/3740/tab). Thanks to this code, 2D acoustic multi-scale reconstructions has been performed. Several optimizers such as gradient descent, non linear conjugate gradient and limited BFGS have also been developed.

This work is a collaboration with Henri Calandra (TOTAL). The time domain FWI results has been presented at Total conference MATHIAS 2019 in Paris [37] and also during the Fall Meeting 2019 AGU in San Francisco [47].

7.2.2. *Box-Tomography imaging in the deep mantle*

Participant: Yder Masson.

Box Tomography is a seismic imaging method (Masson and Romanowicz, 2017) that allows the imaging of localized geological structures buried at arbitrary depth inside the Earth, where neither seismic sources nor receivers are necessarily present. The big advantage of box-tomography over standard tomographic methods is that the numerical modelling (i.e. the ray tracing in travel time tomography and the wave propagation in waveform tomography or full waveform inversion) is completely confined within the small box-region imaged. Thus, box tomography is a lot more efficient than global tomography (i.e. where we invert for the velocity in the larger volume that encompasses all the sources and receivers), for imaging localized objects in the deep Earth. Following a successful, yet partial, application of box tomography to the imaging of the North American continent (i.e. Clouzet et al, 2018), together with Barbara Romanowicz and Sevan Adourian at the Berkeley Seismological Laboratory, we finished implementing the necessary tools for imaging localized structure in the Earth's lower mantle. The following tasks have been completed:

- Modify the global wave propagation solver Specfem_3D_globe in order to compute Green's functions in our current reference Earth model (SEMUCB).
- Modify the local wave propagation solver RegSEM_globe so that the wavefield can be recorded and stored at the surface of the modeling domain.
- Implement real-time compression for an improved management of computed data.

Preliminary results have been presented at the American geophysical union fall meeting 2019 [24]. In the near Future, the latest implementation of Box-Tomography will be deployed to investigate the deep structure under the Yellowstone hotspot down to 20 seconds period.

7.2.3. *Wave-propagation modeling using the distributional finite difference method (DFD).*

Participant: Yder Masson.

In the last decade, the Spectral Element Method (SEM) has become a popular alternative to the Finite Difference method (FD) for modeling wave propagation in heterogeneous geological media. Though this can be debated, SEM is often considered to be more accurate and flexible than FD. This is because SEM has exponential convergence, it allows to accurately model material discontinuities, and complex structures can be meshed using multiple elements. In the mean time, FD is often thought to be simpler and more computationally efficient, in particular because it relies on structured meshed that are well adapted to computational architectures. This motivated us to develop a numerical scheme called the Distributional Finite Difference method (DFD), which combines the efficiency and the relative simplicity of the finite difference method together with an accuracy that compares to that of the finite/spectral element method. Similarly to SEM, the DFD method divides the computational domain in multiple elements but their size can be arbitrarily large. Within each element, the computational operations needed to model wave propagation closely resemble that of FD which makes the method very efficient, in particular when large elements are employed. Further, large elements may be combined with smaller ones to accurately mesh certain regions of space having complex geometry and material discontinuities, thus ensuring higher flexibility. An exploratory code allowing to model 2D and 3D wave propagation in complex media has been developed and demonstrated the interest of the proposed scheme. We presented numerical examples showing the accuracy and the interest of the DFD method for modeling wave propagation through the Earth at the EGU general assembly, 2019 and at the AGU fall meeting 2019 [41].

7.2.4. Time-harmonic seismic inverse problem

Participants: Hélène Barucq, Florian Faucher.

We study the inverse problem associated with the propagation of time-harmonic wave. In the seismic context, the available measurements correspond with partial reflection data, obtained from one side illumination (only the Earth surface is available). The inverse problem aims at recovering the subsurface Earth medium parameters and we employ the Full Waveform Inversion (FWI) method, which employs an iterative minimization algorithm of the difference between the measurement and simulation.

In particular, we investigate the use of new misfit functionals, based upon the *reciprocity-gap*. The use of such functional is only possible when specific measurements are available, and relates to Green's identity. The feature of the cost function is to allow a separation between the observational and numerical sources. In fact, 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.

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).

This work has given rise to a publication [11] and a preprint [53]. It has also been presented in several conferences [32], [35], [33], [34].

7.2.5. Convergence analysis for the seismic inverse problem

Participants: Hélène Barucq, Florian Faucher.

The determination of background velocity by Full Waveform Inversion (FWI) is known to be hampered by the local minima of the data misfit caused by the phase shifts associated with background perturbations. Attraction basins for the underlying optimization problems can be computed around any nominal velocity model and guarantee that the misfit functional has only one (global) minimum. The attraction basins are further associated with tolerable error levels which represent the maximal allowed distance between the (observed) data and the simulations (i.e., the acceptable noise level). The estimates are defined a priori, and only require the computation of (possibly many) first and second order directional derivatives of the (model to synthetic) forward map. The geometry of the search direction and the frequency influence the size of the attraction basins, and complex frequency can be used to enlarge the basins. The size of the attraction basins for the perturbation of background velocities in the classical FWI (global model parametrization) and the data space reflectivity (MBTT) reformulation are compared: the MBTT reformulation increases substantially the size of the attraction basins. Practically, this reformulation compensates for the lack of low frequency data.

This work is a collaboration with Guy Chavent (Inria Rocquencourt) and Henri Calandra (TOTAL). The results have been published in Inverse Problems [12] and in the Research Report [43].

7.2.6. Eigenvector representation for the seismic inverse problem

We study the seismic inverse problem for the recovery of subsurface properties in acoustic media. In order to reduce the ill-posedness of the problem, the heterogeneous wave speed parameter is represented using a limited number of coefficients associated with a basis of eigenvectors of a diffusion equation, following the regularization by discretization approach. We compare several choices for the diffusion coefficient in the partial differential equations, which are extracted from the field of image processing. We first investigate their efficiency for image decomposition (accuracy of the representation with respect to the number of variables and denoising). Next, we implement the method in the quantitative reconstruction procedure for seismic imaging, following the Full Waveform Inversion method.
This work is a collaboration with Otmar Scherzer (University of Vienna) and the results have been documented in [49].

7.2.7. Outgoing solutions for the scalar wave equation in Helioseismology

Participants: Hélène Barucq, Florian Faucher, Ha Pham.

We study the construction and uniqueness of physical solutions for the time-harmonic scalar wave equation arising in helioseismology. The definition of outgoing solutions to the equation in consideration or their construction and uniqueness has not been discussed before in the context of helieoseismology. In our work, we use the Liouville transform to conjugate the original equation to a potential scattering problem for Schrödinger operator, with the new problem containing a Coulomb-type potential. Under assumptions (in terms of density and background sound speed) generalizing ideal atmospheric behavior, we obtain existence and uniqueness of variational solutions.

Solutions obtained in this manner are characterized uniquely by a Sommerfeld-type radiation condition at a new wavenumber. The appearance of this wavenumber is only clear after applying the Liouville transform. Another advantage of the conjugated form is that it makes appear the Whittaker special functions, when ideal atmospheric behavior is extended to the whole domain. This allows for the explicit construction of the outgoing Green kernel and the exact Dirichlet-to-Neumann map and hence reference solutions and radiation boundary condition.

This work has given rise to a report of 135 pages, [45]. Some part have been extracted for a publication which has recently been accepted in ESAIM: M2AN.

Consequently to this work, ongoing research includes the fast construction of the Green's kernel, which is possible thanks to the family of special functions obtained from the previous analytical study, [36]. As part of the ANTS associate team, applications to helioseismology is also ongoing. This work is a collaboration with Damien Fournier and Laurent Gizon (Max Planck Institute at Göttingen). The "Ants workshop on computational helioseismology" has also been organized in this context.

7.2.8. Modeling the propagation of acoustic wave in the Sun

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé, Nathan Rouxelin.

We study time-harmonic propagation of acoustic waves in the Sun in the presence of gravity forces.

Galbrun's equation, a Lagrangian description of acoustic wave propagation, is usually used in helioseismology. However, the discretization of this equation with high-order discontinuous Galerkin methods leads to poor numerical results for solar-like background flow and geometries. As better numerical results were obtained by using another model, the Linearized Euler's Equations, we investigate the equivalence between those two models.If compatible boundary conditions are chosen, it should be possible to reconstruct the solution of Galbrun's equation by solving the Linearized Euler's Equations and then a vectorial transport equation. It turns out that finding those boundary conditions is quite difficult and not always possible.

We also have constructed a reduced model for acoustic wave propagation in the presence of gravity. Under some additional assumptions on the background flow and for high frequencies, the Linearized Euler's Equations can be reduced to a scalar equation on the pressure perturbation. This equation is well-posed in a usual variational framework and it will provide a useful reference solution to validate our numerical methods. It also seems that a similar process could be used in more realistic cases.

7.2.9. Equivalent boundary conditions for acoustic media with exponential densities. Application to the atmosphere in helioseismology

Participants: Juliette Chabassier, Marc Duruflé, Victor Péron.

We present equivalent boundary conditions and asymptotic models for the solution of a transmission problem set in a domain which represents the sun and its atmosphere. This problem models the propagation of an acoustic wave in time-harmonic regime. The specific non-standard feature of this problem lies in the presence of a small parameter δ which represents the inverse rate of the exponential decay of the density in the atmosphere. This problem is well suited for the notion of equivalent conditions and the effect of the atmosphere on the sun is as a first approximation local. This approach leads to solve only equations set in the sun. We derive rigorously equivalent conditions up to the fourth order of approximation with respect to δ for the exact solution u. The construction of equivalent conditions is based on a multiscale expansion in power series of δ for u. Numerical simulations illustrate the theoretical results. Finally we measure the boundary layer phenomenon by introducing a characteristic length that turns out to depend on the mean curvature of the interface between the subdomains. This work has been published in Applied Mathematics and Computation [15].

7.3. Hybrid time discretizations of high-order

7.3.1. Construction and analysis of a fourth order, energy preserving, explicit time discretization for dissipative linear wave equations.

Participants: Juliette Chabassier, Julien Diaz.

A paper was accepted in M2AN [14]. 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).

7.3.2. Space-Time Discretization of Elasto-Acoustic Wave Equation in Polynomial Trefftz-DG Bases

Participants: 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 [28] have been presented during the ICIAM conference (Valencia, July 15-19).

7.4. Modeling and design of wind musical instruments

7.4.1. Full Waveform inversion for bore reconstruction of woodwind instruments

Participants: Juliette Chabassier, Augustin Ernoult.

Several techniques can be used to reconstruct the internal geometry of a wind instrument from acoustics measurements. One possibility is to simulate the passive linear acoustic response of an instrument and to use an optimization process to fit the simulation to the measurements. This technique can be seen as a first step toward the design of wind instruments, where the targeted acoustics properties come no more longer from measurements but are imposed by the designer. We applied the FWI methodology, along with 1D spectral finite element discretization in space [19], to the woodwind instruments (with tone holes, losses and radiation). The algorithm have been implemented in Pyhton3 and is know operational to reconstruct the bore of real instrument. This functionality will be available in an upcoming version of the toolbox OpenWind.

7.4.2. Computation of the entry impedance of a wind instrument with toneholes and radiation

Participants: Guillaume Castera, Juliette Chabassier, Augustin Ernoult, Alexis Thibault, Robin Tournemenne.

Modeling the entry impedance of wind instruments pipes is essential for sound synthesis or instrument qualification. Based on a one-dimensional model of acoustic propagation (“telegraphist’s equations”) we find approximate solutions using a high-order Finite Element Method (FEM1D). Contrary to the more standard semi-analytic Transfer Matrix Method (TMM), the FEM1D can take into account arbitrarily complex and variable coefficients [19]. It is therefore well-suited for the realistic cases involving boundary losses, smooth waveguide geometry, and possibly even a temperature gradient along the instrument’s bore. We model toneholes as junctions of one-dimensional waveguides, and an acoustic radiation impedance models the radiation of all open tube ends. A global matrix is assembled to connect all these elements together, and solved for each frequency to compute the impedance curve. Source code is available in our Python3 toolbox OpenWind.

7.4.3. Time-domain simulation of a reed instrument with toneholes

Participants: Juliette Chabassier, Augustin Ernoult, Alexis Thibault, Robin Tournemenne.

As part of the project aiming at providing practical tools for instrument design, we have been developing a sound synthesis module for reed instruments. We model a reed music instrument, such as the oboe or the bassoon, as a coupled system composed of a nonlinear source mechanism (the reed), and a linear resonating part (the air within the instrument’s bore). Acoustic wave propagation inside of the instrument is reduced to a one-dimensional model, on which a variational approximation is performed, yielding high-order finite elements in space. Tone holes on the side of the instrument are taken into account using junctions of one-dimensional waveguides. The acoustic radiation impedance is written as a positive Padé approximation, so that it leads to a stable time domain model even when opening and closing holes during the simulation. For the reed, a one-degree-of-freedom lumped model is used, in which the reed opening follows a second order ODE and acts as a valve, modulating the flow that enters the pipe based on the pressure difference between the player’s mouth and the inside of the instrument. Energy-consistent time discretization schemes have been derived for each component, so that it is possible to simulate instruments with an arbitrary geometry with good numerical accuracy and stability. The simulations have been implemented in Python3 and will be made available in an upcoming version of the toolbox OpenWind.

7.4.4. Time-domain simulation of 1D acoustic wave propagation with boundary layer losses

Participants: Juliette Chabassier, Augustin Ernoult, Alexis Thibault.

Energy dissipation effects are of critical importance in musical acoustics. Boundary layer losses occurring in acoustic waveguides are usually modeled in the frequency domain, leading to slowly-decreasing kernels in the time domain similar to fractional derivatives. We have developed an energy-consistent time-domain model based on positive fraction approximation of the dissipative operators, leading to the use of $2N + 1$ additional variables per degree of freedom. Coefficients for these new variables depend only on N so that they can be tabulated without any prior knowledge of the waveguide geometry. They are found with an optimization procedure. This model can be discretized using 1D finite elements [19], and an energy-consistent time-stepping scheme can be found. The resulting numerical scheme has been implemented numerically, and source code will be made available in an upcoming version of the toolbox OpenWind. An article is being written and will be submitted soon to JASA.

7.4.5. 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.

SERENA Project-Team

7. New Results

7.1. Hybrid high-order methods for nonlinear mechanics

Participants: Alexandre Ern, Nicolas Pignet.

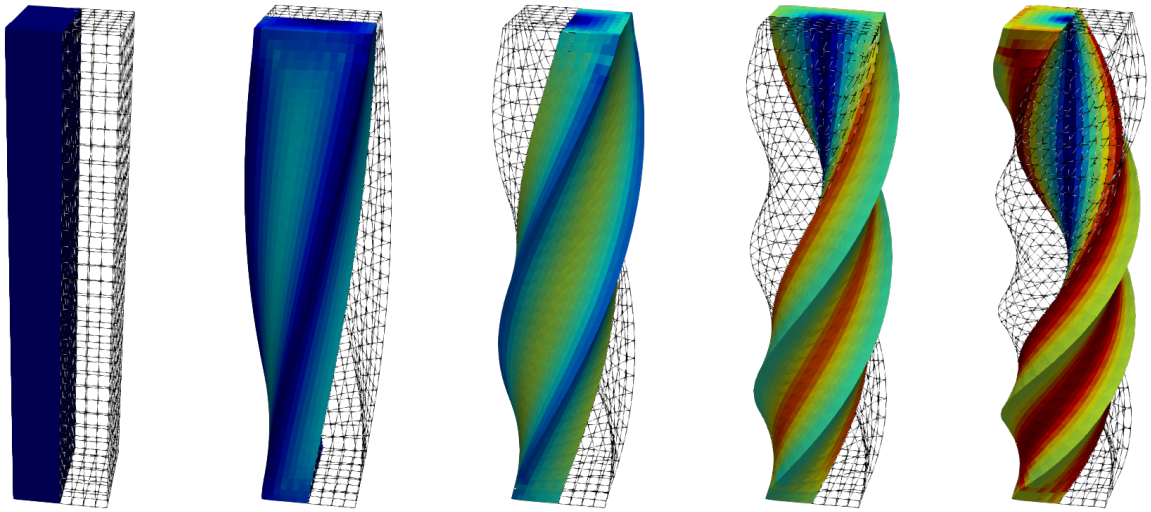


Figure 1. Torsion of a square-section bar: Equivalent plastic strain p (values between 0 (blue) to 0.49 (red)) for HHO with polynomial degree $k = 1$ at the quadrature points for different rotation angles Θ. From left to right: $\Theta = 0°$, $\Theta = 90°$, $\Theta = 180°$, $\Theta = 270°$, and $\Theta = 360°$

Our team contributes actively to the development of hybrid high-order (HHO) methods for nonlinear solid mechanics. Within the PhD of Nicolas Pignet in collaboration with EDF we have addressed several nonlinearities, including plasticity, large deformations, contact, and (Tresca) friction [15], [14], [49]. The advantage with respect to conforming finite elements is the robustness with respect to volumetric locking. The advantage with respect to mixed approaches is computational efficiency avoiding saddle-point formulations and additional unknowns. The advantage with respect to discontinuous Galerkin methods is avoiding the integration of the nonlinear behavior law at face quadrature nodes and the use of symmetric tangent matrices within Newton's method. The torsion of a square-section elastoplastic bar is presented in Figure 1 . The color filling reports the equivalent plastic strain. The solution is obtained with the HHO method using the polynomial degree $k = 1$ for the face and the cell unknowns.

7.2. A hybrid high-order method for flow simulations in discrete fracture networks

Participants: Florent Hédin, Géraldine Pichot, Alexandre Ern.

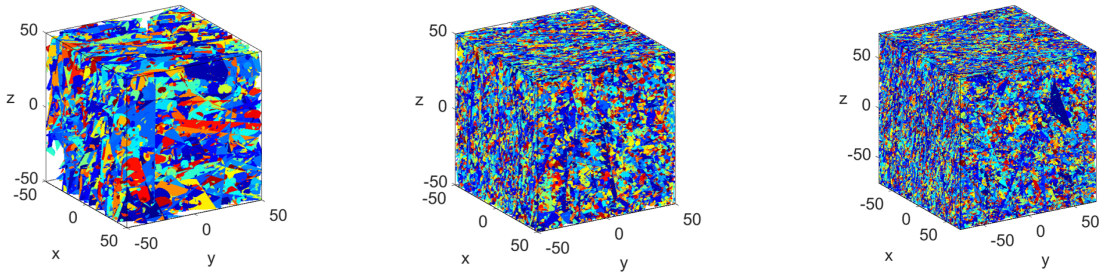


Figure 2. Examples of DFN: (left) B1: 19,007 fractures; (center) B2: 152,399 fractures ; (right) B3: 508,338 fractures.

In [36], we are interested in solving flow in large trimensional Discrete Fracture Networks (DFN) (cf Figure 2) with the hybrid high-order (HHO) method. The objectives of this paper are: (1) to demonstrate the benefit of using a high-order method for computing macroscopic quantities, like the equivalent permeability of fracture rocks; (2) to present the computational efficiency of our C++ software, `NEF++`, which implements the solving of flow in fractures based on the HHO method.

7.3. Analytic expressions of the solutions of advection-diffusion problems in 1D with discontinuous coefficients

Participants: Antoine Lejay, Lionel Lenôtre, Géraldine Pichot.

In [30], we provide a method to compute analytic expressions of the resolvent kernel of differential operators of the diffusion type with discontinuous coefficients in one dimension. Then we apply it when the coefficients are piecewise constant. We also perform the Laplace inversion of the resolvent kernel to obtain expressions of the transition density functions or fundamental solutions. We show how these explicit formula are useful to simulate advection-diffusion problems using particle tracking techniques.

7.4. An exponential timestepping algorithm for diffusion with discontinuous coefficients

Participants: Antoine Lejay, Lionel Lenôtre, Géraldine Pichot.

In [29], we present 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.

7.5. Polynomial-degree-robust multilevel algebraic error estimator & solver

Participants: Ani Miraci, Jan Papež, Martin Vohralík.

In [58], we devise a novel multilevel a posteriori estimator of the algebraic error. It delivers a fully computable, guaranteed lower bound on the error between an unknown exact solution of a system of linear algebraic equations and its approximation by an algebraic solver. The bound is also proved to be efficient, i.e., it also gives an upper bound on the algebraic error. Remarkably, the quality of these bounds is independent of the approximation polynomial degree. The derived estimates give immediately rise to a multilevel iterative algebraic solver whose contraction factor is independent of the polynomial degree of the approximation. We actually prove an equivalence between efficiency of the estimator and contraction of the solver. The estimator/solver are based on a global coarsest-level solve of lowest-order ($p = 1$), followed by local patchwise p-degree problems solved on the other levels. It corresponds to a V-cycle geometric multigrid solver with zero pre- and one post-smoothing step via block-Jacobi. A salient feature is the choice of the optimal step size for the descent direction.

7.6. Local- and global-best equivalence, simple projector, and optimal hp approximation in $\mathbf{H}(\mathrm{div})$

Participants: Alexandre Ern, Thirupathi Gudi, Iain Smears, Martin Vohralík.

$$\min_{\substack{\boldsymbol{v}_{\mathcal{T}}\in\boldsymbol{RTN}_p(\mathcal{T})\cap\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)\\ \nabla\cdot\boldsymbol{v}_{\mathcal{T}}=\Pi_{\mathcal{T}}^p(\nabla\cdot\boldsymbol{v})}} \|\boldsymbol{v}-\boldsymbol{v}_{\mathcal{T}}\|_\Omega^2$$
$$\approx \sum_{K\in\mathcal{T}} \min_{\boldsymbol{v}_K\in\boldsymbol{RTN}_p(K)} \|\boldsymbol{v}-\boldsymbol{v}_K\|_K^2$$

Figure 3. Equivalence between global-best and local-best approximation for any $\mathbf{H}(\mathrm{div})$ function $\mathbf{v}$ with zero normal flux over part of the boundary and piecewise polynomial divergence

In [53], we prove that a global-best approximation in $\mathbf{H}(\mathrm{div})$, with constraints on normal component continuity and divergence, is equivalent to the sum of independent local-best approximations, without any constraints, as illustrated in Figure 3 . This may seem surprising on a first sight since the right term in Figure 3 is seemingly much smaller (since the minimization set is unconstrained and thus much bigger). This result leads to optimal a priori hp-error estimates for mixed and least-squares finite element methods, which were missing in the literature until 2019. Additionally, the construction we devise gives rise to a simple stable local commuting projector in $\mathbf{H}(\mathrm{div})$, which delivers approximation error equivalent to the local-best approximation and applies under the minimal necessary Sobolev $\mathbf{H}(\mathrm{div})$ regularity, which is another result that has been sought for a very long time.

STEEP Project-Team

6. New Results

6.1. Analysis of socio-ecological dimensions of human activities – A case study of Beaufort cheese production in the Maurienne Valley

The PhD thesis of Michela Bevione aims at analysing socio-ecological dimensions of human activities creating wealth by coupling quantitative-biophysical approaches and qualitative and socio-economic methodologies to assess territorial metabolism. By focusing on the interactions between flows and actors, the methodology we propose aims at providing a methodological framework for the understanding of a territory and its capability.

As a case study for this thesis, we chose to focus on the production of the AOC-labelled cheese Beaufort in the Maurienne Valley (Savoie department, Auvergne-Rhône-Alpes region, France). Indeed, agriculture play a structuring role for the economic and social dynamics of the valley, and the landscape construction induced by farming activities contributes to create favourable conditions to the development of the touristic sector. Beaufort represents the flagship product of the agricultural sector in the valley and most of farms are dedicated to milk production for the Beaufort industry.

In [7] we represent the circulation of material flows through flow maps, showing the movement of material and monetary resources and products, their direction, source and destination. We focus on the circulation of flows related to the Beaufort industry within the Maurienne Valley and between the valley and other territories. Through Sankey diagrams (a specific kind of flow maps, where the width of the arrows is proportional to the flow quantity) we present the dominant contributions to the overall material flows circulation. This kind of representation is appropriate to characterise the circulation of material flows, the allocation of environmental pressures throughout the Beaufort industry, as well as the monetary dimension and the added value associated to Beaufort production. Mapping the geographical origin of input resources and the destination of output products and incomes allows to evaluate actors' capacity to create wealth through the activation and mobilisation of local resources and/or their dependence on foreign inputs.

Furthermore, results include schematic representations of the relations between local, extra-territorial actors and the circulation of material, environmental and monetary flows. The influence of immaterial resources (informational flows and traditional savoir-faire) and local infrastructures on the circulation of flows, and vice versa, is illustrated. Finally, positive and negative retroactions induced by output products on input resources for Beaufort production are drawn, as well as the interactions with other sub-systems creating wealth in the valley.

6.2. Sensitivity analysis of World3

World3 is a computer tool created to simulate the interactions between the world population, industrial growth and food production within the limits of the planet. It aims to highlight the problems posed by indefinite material growth in a finite world. The first version of this tool was proposed in 1972 by MIT researchers for the first report to the Club of Rome [19]. This report was both highly successful and polemic. The main detractors of the model criticized it for being too approximate in the choice of the parameters and for beeing too simplistic. We started to work on revisiting some aspects of the scientific validation of this model, in a context where growth is widely debated in the scientific and civil community, through a new and more sophisticated sensitivity analysis of the model, compared to what is available in the literature [11].

6.3. Efficient computation of solution space and conflicts detection for linear systems

Our work on Material Flow Analysis (see e.g. the *AF Filières* project), involves the analysis of systems of linear inequalities, $l \leq Ax \leq u$. There are three different but complementary goals for the analysis: (i) given some known variables x_i, efficiently compute the solution space of unknown variables, (ii) if the set of constraints is infeasible, efficiently identify the conflicts, (iii) efficiently classify variables to determine whether they are redundant, just measured, determinable or non-determinable. A baseline implementation for these tasks was available in the team but proved to be too inefficient for larger problem sizes. Through the internship of Alexandre Borthomieu we worked on various improvements, on the algorithmic and implementation side (e.g. choice of programming language), that eventually led to a reduction of execution by three orders of magnitude, compared to the previous implementation.

6.4. Mapping ecosystem services bundles in a heterogeneous mountain region

2019 was the final year of production of the ESNET project (which officially ended in 2017). This and the following section describe our two most complex pieces of work in that project.

Recent institutional and policy frameworks prescribe the incorporation of ecosystem services (ES) into land use management and planning, favouring co-production of ES assessments by stakeholders, land planners and scientists. Incorporating ES into land management and planning requires models to map and analyze ES. Also, because ES do not vary independently, many operational issues ultimately relate to the mitigation of ES trade-offs, so that multiple ES and their interactions need to be considered. Using a highly accurate LULC (Land Use Land Cover) database for the Grenoble urban region (French Alps), we mapped twelve ES using a range of models of varied complexity [5]. A specific, fine-grained (less than 1 ha) LULC database at regional scale (4450 km^2) added great spatial precision in individual ES models, in spite of limits of the typological resolution for forests and semi-natural areas. We analysed ES bundles within three different socio-ecosystems and associated landscape types (periurban, rural and forest areas). Such type-specific bundles highlighted distinctive ES trade-offs and synergies for each landscape. Advanced approaches combining remote sensing, targeted field data collection and expert knowledge from scientists and stakeholders are expected to provide the significant progress that is now required to support the reduction of trade-offs and enhance synergies between management objectives.

6.5. Co-constructing future land-use scenarios for the Grenoble region, France

Physically and socially heterogeneous mountain landscapes support high biodiversity and multiple ecosystem services. But rapid landscape transformation from fast urbanisation and agricultural intensification around cities to abandonment and depopulation in higher and more remote districts, raises urgent environmental and planning issues. For anticipating their future in a highly uncertain socio-economic context, we engaged stakeholders of a dynamic urban region of the French Alps in an exemplary interactive participatory scenario planning (PSP) for co-creating salient, credible and legitimate scenarios. Stakeholders helped researchers adapt, downscale and spatialize four normative visions from the regional government, co-producing four storylines of trend versus break-away futures. Stakeholder input, combined with planning documents and analyses of recent dynamics, enabled parameterisation of high-resolution models of urban expansion, agriculture and forest dynamics. With similar storylines in spite of stakeholders insisting on different governance arrangements, both trend scenarios met current local and European planning objectives of containing urban expansion and limiting loss and fragmentation of agricultural land. Both break-away scenarios induced considerable conversion from agriculture to forest, but with highly distinctive patterns. Under a commonly investigated, deregulated liberal economic context, encroachment was random and patchy across valleys and mountains. A novel reinforced nature protection scenario affecting primarily mountain and hilly areas fostered deliberate consolidation of forested areas and connectivity. This transdisciplinary approach demonstrated the potential of combining downscaled normative scenarios with local, spatially-precise dynamics informed by stakeholders for local appropriation of topdown visions, and for supporting land planning and subsequent assessment of ecosystem service trade-offs. This work is described in [4].

TONUS Project-Team

7. New Results

7.1. Relaxation method for Guiding-Center equation

Participants: E. Franck, R. Helie, L. Navoret, P. Helluy.
In previous years, implicit kinetic relaxation methods have been developed to treat conservation laws without CFL and without a non-trivial matrix to reverse [6]-[4]. We have started to apply this method with a spectral discretization for transport equations such as the guiding-center equation (a non constant advection equation coupled with elliptic problem used in plasma physics). The scheme obtained has a very high order of convergence for an instability test case and is very simple to implement. We have also investigated the different kinetic relaxation representations. However, they suffer from inaccuracy at the boundaries. We have proposed a new approach in 1D [8] to analyse this behaviour and a new way to apply boundary conditions to ensure they are compatible both with the approximated system and its kinetic approximation. Extending this approach to higher dimensions is one of the objectives of the thesis of Romane Helie.

7.2. Relaxation method for transport in Tokamak

Participants: M. Boileau, P. Helluy, B. Bramas (Inria Camus).
To apply the relaxation method in a Tokamak context, we have developed a code called Chukrut (in Schnaps) that can handle kinetic relaxation models in Tokamak geometry [15]-[17]-[13]. In the poloidal direction the code uses an unstructured Discontinuous Galerkin solver which solves the transport equation (the main ingredient of the kinetic relaxation method) by using the scheduling graph linked to the upwind scheme. In the toroidal direction we use an exact solver on uniform grids (which will be replaced by a semi-Lagrangian solver). The algorithm is parallelised in the poloidal plane by a task-based OpenMP implementation and in the toroidal direction by MPI parallelism.

7.3. Relaxation method for Euler/MHD in low-Mach regime

Participants: E. Franck, L. Navoret, F. Bouchut (Marne la Vallée university).
Previously, we have proposed implicit relaxation methods for fluid models that allow us to reverse a simple system. However, previous methods [5] were not very effective in the multi-scale regimes of interest. We therefore proposed a semi-implicit scheme based on a dynamic splitting and a relaxation of fast waves only. The scheme was first applied to the Euler equations in low Mach regime. The scheme is stable and accurate regardless of the Mach number. We have successfully applied the method for the equilibria of the Shallow Water equations. Since last summer we have begun the extension for the 1D MHD with and without dispersive effects. The first results show that we obtain a similar method compared to the Euler case with acceptable stability conditions as for the Euler equations.

7.4. Reduced model for the Scrape-Off Layer

Participants: L. Navoret, M. Mehrenberger, P. Ghendrih (CEA Cadarache)
In this work, we consider a one-dimensional model for describing the two-species plasma dynamics in the scrape-off layer. This region is defined as the transition between the core of the plasma and the edge and is located around the first non-closed magnetic field line. The electron and ion distribution functions satisfy a Vlasov-Poisson system with source and absorption terms and a non-homogeneous equilibrium is expected to develop. A high-order semi-Lagrangian scheme has been implemented to correctly capture such a dynamics.

7.5. Recurrence phenomenon for finite element grid based Vlasov solver

Participants: L. Navoret, M. Mehrenberger, N. Pham

We have improved our previous (last year) result concerning the recurrence phenomenon by providing a complete proof of the asymptotic behaviour of the correlation function. Indeed, we prove that, in the fine grid limit, the correlation function of the density exactly concentrates at multiple times of the recurrence time. This thus fully confirms the fact the amplitude of the recurrence phenomenon is actually linked to the spectral accuracy of the velocity quadrature when computing the charge density at least for the linear transport equation.

7.6. Machine learning techniques for reduced model and stabilization

Participants: E. Franck, L. Navoret, V. Vigon (IRMA Strasbourg).
Just recently, we have begun to work on applications of machine learning techniques for the plasma simulation. This preliminary work is in the context of "Action exploratoire MALESI" and will really begin in 2020. The first point is the construction of a new closure for the fluid models using kinetic simulation as data. We have constructed 1D solvers for the Vlasov-Poisson equation with collisional operator and Compressible Navier-Stokes Poisson models. Comparing the models we observe that the classical Navier-Stokes closure is not sufficient when the Knudsen number is larger than 0.3-0.4. Currently we generate data using the Vlasov-Poisson code to train a neural-network for the closure. The second point is about the stabilization of the numerical method using CNN. We began a study to construct a Convulational Neural Network (CNN) to detect the Gibbs oscillations in the fluid simulations.

7.7. Asymptotic Preserving scheme for Vlasov-Maxwell to MHD

Participants: E. Franck, A. Crestetto (Nantes university), M. Badsi (Nantes university).
The MHD equation can be obtained by taking the limit of different small parameters of the bi-species Vlasov-Maxwell equations. Obtaining an "asymptotic preserving" scheme for the Vlasov equation (cost independent of the small parameters) is an important goal. Indeed, this type of scheme would allows us to construct coupling methods between MHD and Vlasov equations or to make simulations in various regimes to construct closures with data (see the previous point). During this year we have written a scheme able to treat the "quasi-neutral" and "mass-less" limits between the two-species Euler-Maxwell equations and the MHD model. The scheme is partially validated. We will finish the validation and add the collisional limit between Vlasov-Maxwell and Euler-Maxwell equations.

7.8. Optimal control for population dynamics

Participants: Y. Privat, L. Almeida (Sorbonne University), M. Duprez (Dauphine University) and N. Vauchelet (Paris 13 University).
Particular attention is being paid to the transmission of dengue fever, an arbovirus transmitted to humans by mosquitoes [3]-[2]. There is no vaccine to immunize a population. It has been observed that when a mosquito population was infected with the Wolbachia virus, they stopped transmitting the disease. In addition, the virus is transmitted from mother to child and is characterized by cytoplasmic incompatibility (no possible crosses between infected males and healthy females). On the other hand, infected mosquitoes have a reduced lifespan and fertility. Mathematically, this situation can be modelled (in a simplified way) using a controlled reaction-diffusion system. The control term represents the strategy of releasing (time-space) mosquitoes infected by Wolbachia. The practical questions that arise and that we wish to address are:

- how to carry out these releases to ensure the invasion?
- how to optimize the domain and form of releases?

Preliminary work has made it possible to determine a plausible temporal control strategy.

7.9. Observability for wave equation and high frequency behavior

Participants: Y. Privat, E. Humbert (Tours University) et E. Trélat (Sorbonne University).

We have determined the asymptotic in time of the observability constant in closed manifolds. In particular, we have proved that this limit can be represented as the minimum of two quantities: one purely spectral and another called the geometric quantity representing the limit of the average time spent by geodesics within the observation domain [19]-[19].

7.10. Developement of a Python library for tomography diagnostics

Participants: L. Mendoza

In the tofu code, a big component of both the direct and inverse solvers is the integration module. During the 2019 project it was developed and accelerated. Special attention was brought to memory optimization. Core functions for the inversions routines were developed and parallelized using OpenMP. The number of users and developers of the library has significantly increased in the last year (collaborators in CEA cadarache, ITER, CEA saclay, IPP Garching, etc.) so one of the main objectives was to better the continuous integration and documentation of the code: more unitary and simulation tests have been implemented, an online Web site with the documentation has been added, the library can be used on more platforms (windows, mac os x, and linux), and more fusion devices are now available (West, ITER, JET, etc.).

7.11. Finite volume methods for complex hyperbolic systems

Participants: P. Helluy, L. Quibel (EDF)

This year we have developed a Lattice Boltzmann scheme able to treat really complex and tabuled EOS (Equation Of State) for compressible multiphase flows (two and three phases). This new scheme have been implemented in the PyOpenCL Patapon. Additionally, in order to perform realistic simulations of such situations, we have also proposed a code based on a model that can handle both the thermodynamical disequilibrium between liquid and vapor and complex equations. This code is based on a relaxation scheme which is the best compromise between accuracy and stability.

7.12. The study of domain walls in micromagnetism

Participants: C. Courtès, R. Côte (IRMA)

A ferromagnetic material consists of a succession of isolated subdomains (known as the magnetic domains) in which the magnetic moments are aligned and point in the same direction. The interface separating two magnetic domains is called the domain wall and corresponds to a localized area where the direction of the magnetization suddenly changes. Mathematically, those domain walls correspond to the minimizers of the well-known micromagnatics energy. The magnetic behavior of ferromagnetic materials is due to the arrangement of the magnetic domains and to the dynamics of their domains walls. That dynamic is governed by the nonlinear Landau-Lifshitz-Gilbert equation. We study numerically and theoretically the stability and the interaction of two domain walls. Depending on the initial topological configuration, two domain walls may collide to give rise to a persistent profile or annihilate both, which results in aligning all magnetization vectors of the nanowire in the same direction.

7.13. Maxwell solvers

Participants: P. Helluy, M. Houillon

In collaboration with the AxesSim company, we continue the development of our CLAC software devoted to electromagnetic simulations in biological environment. We have implemented a new wire model. We have also run computations on the new CNRS supercomputer: Jean Zay. We now routinely launch simulations on 64 V100 GPUs in parallel for performing parameter studies of various antennas near to the human body (we can for instance vary the humidity level of the skin).

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. Mathematical study of ecological models

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Clotilde Djuikem, Israël Tankam Chedjou.

Semi-discrete models. Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [41]. We used such models and analyzed their properties in several practical situations, some of them requiring such a modeling to describe external perturbations of natural systems such as harvest, and others to take seasonality into account. We developed such models in the context of the analysis of the effect of stochasticity and Allee effects on the introduction of populations [14], seasonality in the dynamics of coffee leaf rust [59] and of banana and plantain burrowing nematodes [67], as well as for the protection of plant resistance against root-knot nematodes [66].

Models in plant epidemiology. We developed and analysed dynamical models describing plant-parasite interactions, in order to better understand, predict and control the evolution of damages in crops. We considered several pathosystems, further described in Section 7.2.3 , describing and controlling the impact on plants of fungi [59], [39], viruses [36], nematodes [67], [66], and pests [60].

7.1.1.2. Estimation and control

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Yves Fotso Fotso, Samuel Nilusmas, Israël Tankam Chedjou.

Parameter identification in complex systems. In complex biological systems, identifying model parameters is a challenge that raises identifiability issues. To fit a within-host immunological model to a large data set of individual viremia profiles, we developed an Approximate Bayesian Computation (ABC)-like method that yielded several parameter sets compatible with the data and reflecting the variability among individuals [25]. This work benefited from the resources and support of NEF computation cluster.

Optimal control and optimisation. We developed several approaches to control the evolution of crop pests. To reduce crop losses due to plant-parasitic nematodes, we optimised (i) rotation strategies between resistant and susceptible cultivars of horticultural crops [76], or (ii) fallow periods between plantain cropping seasons [67]. These optimisation problems were solved on a finite time horizon. They benefited from the resources and support of NEF computation cluster.

We also solved an optimal control problem to limit the damages due to coffee berry borers [60]. It consisted in designing the most cost-efficient application of a biopesticide over time. Using Pontryagin's maximum principle, we determined the existence and structure of the solution. The problem was solved numerically using BOCOP (https://www.bocop.org/).

7.1.1.3. Analysis of multistability and periodic behavior with hybrid models

Participants: Madalena Chaves, Eleni Firippi.

Probabilistic dynamics tool for hybrid models In a collaboration with D. Figueiredo and M.A. Martins from the University of Aveiro, Portugal (project PHC Pessoa), a tool was developed for simulating weighted reactive models [55]. These are essentially discrete models with dynamics described by state transition graphs: each transition has a given weight and the graph has the capacity to alter its accessibility relations.

M. Chaves and M.A. Martins jointly edited a book with selected papers from the Symposium on Molecular Logic and Computational Synthetic Biology [70], gathering work on different formalisms and applications of hybrid models.

Coupling and synchronization of piecewise linear systems This work studies the coupling of N identical positive feedback loops described by piecewise linear differential equations. Under diffusive coupling, and for different conditions on the coupling parameters, the N systems may synchronize or, alternatively, generate a set ofnew steady states that form a specific pattern [49]. An unexpected result is the existence of a special relationship between the number of components N and the maximal concentration-to-activity threshold ratio $(V_1/(\gamma_1\theta_1))$. This relationship implies that, for very specific parameter sets, the N compartments cannot be guaranteed to synchronize.

7.1.1.4. Dynamics of complex feedback architectures

Participants: Madalena Chaves, Jean-Luc Gouzé.

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 [19]. 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 [20].

7.1.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Olivier Bernard, Valentina Baldazzi, Lucie Chambon, Carlos Martinez Von Dossow, Agustin Yabo, Alex Dos Reis de Souza, Walid Djema, Sofya Maslovskaya.

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 [80]. A reduction pipeline combining several reduction strategies has been 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) [64]. A paper is currently under revision for Mathematical Biosciences.

Analysis of an integrated cell division-endoreduplication and expansion model. The development of a new organ depends on cell-cycle progression and cell expansion, but the interaction and coordination between these processes is still unclear [27]. An integrated model of fruit development has been developed and used to test different interaction schemes, by comparing simulation results to the observed cell size distribution in tomato fruit [15].

Modeling cell growth and resource allocation. In the framework of the Maximic project (collaboration with IBIS team) and as a follow up of our previous work [82], we investigated the impact of energy metabolism on cell's strategy for resource allocation. Preliminary results show that the inclusion of energy costs leads to the emergence of a trade-off between growth rate and yield, as experimentally observed in many bacterial cells .

The allocation of cellular resources can strongly influence not only the rate of cell growth but also the resulting cell size [78]. To better investigate the connection between proteome allocation and cell volume, the original model by Giordano et al. [82] has been connected to a biophysical model of cell growth, explicitly describing cell volume increase as a function of cell's internal pressure and mechanical properties. The resulting model will be used to investigate the mechanisms (control of osmotic pressure or wall mechanics) behind cell size control under different environmental constrains [84].

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 and MCTAO project-teams. We developed different versions of the problem, and consider a new problem where the aim is to optimize the production of a product [68], [40], [50], (ANR project Maximic, PhD thesis of A. Yabo). We also study variations of the model, for example in the chemostat [57]. The precise mathematical analysis of the optimal behavior (turnpike property) is under investigation.

A synthetic community of bacteria. In the framework of IPL Cosy, we study the coexistence of two strains of bacteria E. Coli in a bioreactor. The strains have been modified synthetically to achieve some goals. The aim is to obtain a better productivity in the consortium than in a single strain, by control technics. The description of models is in revision for Plos Comp. Biol.

In collaboration with team VALSE (Lille), we also studied several problems of estimation and robust stabilization related to IPL Cosy, for two bacterial species in a bioreactor [53], [54].

Control of a model of synthesis of a virulence factor. In collaboration with J.-A. Sepulchre (UCA), we modeled 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 [37]. We also studied the problem of periodic inputs for maximization of some yield [97].

Hybrid control of genetic networks. We designed 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 [18]. Several other control strategies are studied [47], [48], [46]. This is part of the PhD thesis of L. Chambon.

7.1.3. Biochemical and signaling models

Participants: Madalena Chaves, Eleni Firippi, Sofia Almeida, Marielle Péré, Luis Gomes Pereira, Jérémie Roux.

7.1.3.1. Analysis and coupling of biological oscillators

Modeling, analysis and coupling of the mammalian cell cycle and clock A transcriptional model of the mammalian circadian clock was developed in [13] and its parameters calibrated against experimental data from F. Delaunay's lab. A cell cycle model was also previously developed by us [77]. The interactions between the two oscillators are investigated under uni- or bi-directional coupling schemes [44]. 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). This work is in collaboration with F. Delaunay (ANR ICycle) and part of the PhD thesis of Sofia Almeida.

Period-control in a coupled system of two genetic oscillators In the context of ANR project ICycle, we consider two reduced models that mimic the dynamics of the cell cycle and clock oscillators and study the effect of each oscillator on the coupled system, from a synthetic biology perspective [56]. The first observation is that oscillator A is more likely to be the controller of the coupled system period when the dynamis of oscillator B becomes stable due to the coupling strength. Another interesting observation is that the coupled system exhibits oscillatory dynamics over an increased region of the parameter space. This work is part of the PhD thesis of Eleni Firippi (ANR ICycle).

7.1.3.2. Modeling the apoptotic signaling pathway

A detailed model of the death receptor layer In a collaboration with J. Roux and within project Imodrez, he 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. In a first approach, we constructed a detailed model to represent the death receptor-ligand binding and subsequent signaling cascade [11]. This model was used to study the effect of intrinsic and extrinsic noise sources, and suggested the need to expand a set of reactions on the model, to account for the observed cell heterogeneity (this was part of the PhD thesis of Luis Pereira).

A basic model to explore the effect of a positive feedback loop Analysis of the detailed apoptosis receptor model uncovered a set of reactions for which the introduction of a positive feedback loop from caspase 8 was able to significantly increase the range of variability in the model in response to extrinsic noise. To better understand this mechanism and the role of positive loop in cell response variability, we are constructing a reduced model representing only the basic components: death ligand and receptor, caspase 8 and two intermediate complexes. This is part of the work of the PhD student Marielle Péré.

7.2. Fields of application

7.2.1. Bioenergy

7.2.1.1. Modelling microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Walid Djema, Ignacio Lopez, David Demory, Ouassim Bara, Jean-Philippe Steyer.

Experimental developments

Running experiments in controlled dynamical environments. The experimental platform made of continuous photobioreactors driven by a set of automaton controlled by the ODIN software is a powerful and unique tool which gave rise to a quantity of very original experiments. Such platform improved knowledge of several biological processes such as lipid accumulation or cell cycle under light fluctuation, etc [69].

This experimental platform was used to control the long term stress applied to a population of microalgae [72]. This Darwinian selection procedure generated several new strains more resistant to oxidative stresses after several months in the so called selectiostats [58].

Experiments were run to understand the interactions in a simplified ecosystem between microalgae and cyanobacteria. The initial idea was to use a nitrogen fixing cyanobacteria providing nitrogen to the microalgae. It turns out that negative interactions appear in this ecosystem, first because of the mutual shadowing of these organisms, and second because of the production of allelopathic substances inhibiting the competitive organisms [79].

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.

Experimental work was also carried out in collaboration with the Inalve startup with microalgal biofilm to determine the impact of light and dark sequences on cell growth and photoacclimation [26], [63]. The architecture of the biofilms was also observed for different species with confocal microscopic techniques [23].

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 developed by I. Lopez to represent the dynamics of polar lipids, especially when faced to higher oxygen concentration. In particular, this model represents the microalgae growth under different conditions of temperature, light and oxygen.

Modeling photosynthetic biofilms. Several models have been developed to represent the growth of microalgae within a biofilm. A first structured physiological model, extending the one proposed in [95] 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 . Another approach accounts for the dynamics of the light harvesting systems when cells are submitted to rapid successions of light and dark phases [28], [71]. 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 [45]. The model was used to identify the worldwide potential of microalgal biofilms under different climates [26].

Modeling microalgae production processes.

A synthesis has been written on the different aspects for developing models of microalgae in the field of wastewater treatment [38]. The paper is completed by a position paper proposing guidelines for the development of models in biotechnology [31]. A model representing the dynamics of microalgae when growing in suboptimal conditions of light, nitrogen and phosphorus was developed. It consists in an extension of the Droop model accounting for the two quota of nitrogen and phosphorus [65]. This was the topic of the

internship of Luis Plaza Alvarez. The model also represents the pigment acclimation to various light intensities. We have studied in [75] the response of a Droop model forced by periodic light or temperature signals. We transformed the model into a planar periodic system generating a monotone dynamical system. Combined with results on periodic Kolmogorov equations, the global dynamics of the system can be described.

Modeling thermal adaptation in microalgae.

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 [22]. 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 [22].

Modeling viral infection in microalgae. In collaboration with A.-C. Baudoux (Biological Station of Roscoff) 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 [22]. The model was then extrapolated to the whole ocean to better understand how the warming will impact the mortality due to viruses.

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 [21].

Control of microalgal biofilms.

Determining the optimal operating conditions for a rotating algal biofilm process [63] is a difficult question. A 1D model was developed, and the gradient associated to the productivity at the process scale was computed. Then the conditions maximizing productivity were derived, playing on the conveyer belt velocity and geometry [71].

Interactions between species. We have proposed an optimal control strategy to select in minimal time the microalgal strain with the lowest pigment content [51]. The control takes benefit from photoinhibition to compute light stresses penalizing the strains with a higher pigment content and finally selecting microalgae with lower chlorophyll content. Another optimal control problem was considered for selecting a strain of interest within two species competing for the same substrate, when dynamics is represented by a Droop model [52], [73], [74]. In both cases, the optimal control derived from the Pontryagin maximum principle also exhibit a turnpike behaviour. This is a collaboration with team MCTAO.

Strategies to improve the temperature response have also been studied. We modelled the adaptive dynamics for a population submitted to a variable temperature [58]. This was used at the LOV to design experiments with periodic temperature stresses aiming at enhancing polyunsaturated long chain fatty acids content of *Tisochrysis lutea* [72].

7.2.1.3. Modelling mitochondrial inheritance patterns

Most eukaryotes inherit their mitochondria from only one of their parents. When there are different sexes, it is almost always the maternal mitochondria that are transmitted. Indeed, maternal uniparental inheritance has been reported for the brown alga Ectocarpus but we show in this study [33] that different strains of Ectocarpus can exhibit different patterns of inheritance: Ectocarpus siliculosus strains showed maternal uniparental inheritance, as expected, but crosses using different Ectocarpus species 7 strains exhibited either

paternal uniparental inheritance or an unusual pattern of transmission where progeny inherited either maternal or paternal mitochondria, but not both. A possible correlation between the pattern of mitochondrial inheritance and male gamete parthenogenesis was investigated. Moreover, in contrast to observations in the green lineage, we did not detect any change in the pattern of mitochondrial inheritance in mutant strains affected in life cycle progression. Finally, an analysis of field-isolated strains provided evidence of mitochondrial genome recombination in both Ectocarpus species.

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 [79].

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 [96].

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 [91]. Control strategies playing with the dilution rate, shadowing or modifying depth were then proposed [90].

Finally, a study of the possible sensors which would enhance the monitoring of these process was proposed [30], [29]

7.2.2.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer, Marjorie Alejandra Morales Arancibia.

Environmental impact assessment. To follow up the pioneering life cycle assessment (LCA) work of [87], we identified the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable [93].
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 was carried out. Depending on the period of the year, changing the species can both improve productivity and reduce environmental footprint [34].

We have also studied the environmental impact of protein production from microalgae in an algal biofilm process and compared it to other sources (fisheries, soy,...). This study confirms the interest of microalgae for reducing the environmental impact.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE (Laboratory of Environmental Biotechnology, Narbonne).

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, 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 [89], and extends more generally to pulse perturbations in population dynamics.

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 in Python MESA, on the basis of a previous work in 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 [88]. The modelling platform is interactive and can be accessed online at http://popintro.sophia.inra.fr/.

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 [14].

Characteristics of space and the behavior and population dynamics of biological control agents. 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 [94] and Marjorie Haond's PhD theses [85]. The last article associated with Thibaut Morel Journel's Thesis appeared this year [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 [86] as represented by its local carrying capacity can positively influence the spreading speeed of an expanding population. This work has been published as a preprint recommended by *Peer Community in Ecology* and is on the verge to be submitted to a regular scientific journal. This work has been performed in collaboration with Elodie Vercken (ISA) and Lionel Roques (BioSP, Avignon).

In a different context, we studied how predatory mite population development can be enhanced by the provision of artificial habitats. One paper focused on the influence of different artificial materials on the oviposition and survival of predatory mites appeared this year [16]. This topic was also at the core of the Master 2 intership of Lucas Etienne [81] during which he studied the combined influences of artificial habitats and additional food on the development of a predatory mite and on the control of a phytophagous mite. An article reporting on this study is currently under preparation.

Modelling 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. A control was introduced, based on a biopesticide (entomopathogenic fungus such as *Beauveria bassiana*) that is sprayed and persist on the berries. An optimal control problem was solved (see Section 7.1.1.2). 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. Depending on the initial pest infestation, the optimal solution structure varied [60], [62]. This research pertains to Yves Fotso Fotso's PhD thesis, who visited BIOCORE during 5 months in 2019 through the EPITAG associate team.

7.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Clotilde Djuikem.

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.

We have developed a (spatio-)temporal epidemiological model of the phoma stem canker of oilseed rape, to test and assess the durability of deployment strategies of various cultivars. Based on this model, we aim at developing a user-friendly, upgradeable and efficient simulation tool designed for researchers as well as non academic partners from technical institutes and agriculture cooperatives. We hence applied and obtained the SiDRes AMDT, which will start in 2020.

A stochastic model was developed to help determine the efficiency of pyramiding qualitative resistance and quantitative resistance narrowing population bottlenecks exerted on plant 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 [36]. This studiy 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).

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 [61]. This was done in collaboration with Pauline Clin and 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.

Plant-parasitic 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.

Based on an interaction model between plantain roots and *Radopholus similis*, we solved an optimisation problem (see Section 7.1.1.2). It aimed at determining 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. Fallow periods reduce the nematode population in the soil, as these pests need roots to feed on and reproduce. For a relatively long time horizon, deploying one season less than the maximum possible number of cropping seasons resulted in a better multi-seasonal profit. The optimal solution consisted in applying long fallows at the beginning, to drastically reduce the nematode population. The profit was lower for more regular fallows, but the final soil infestation was also lower [67]. This research pertains to Israël Tankam Chedjou's PhD thesis, who visited BIOCORE during 5 months in 2019 through the EPITAG associate team.

We also studied resistance-based root-knot nematode control. As virulent nematodes exhibit a reduced fitness on susceptible crops, alternating resistant and susceptible plants could help increase the efficiency and durability of such control methods. Optimal crop rotations (see Section 7.1.1.2 were characterised by low ratios of resistant plants and were robust to parameter uncertainty. Rotations provided significant gains over resistant-only strategies, especially for intermediate fitness costs and severe epidemic contexts. Switching from the current general deployment of resistant crops to custom rotation strategies could not only maintain or increase crop yield, but also preserve the few and valuable R-genes available. This research pertains to Samuel Nilusmas' PhD thesis. This work has been publisehd as a pre-print [76] and is currently under review. It has also been presented at several national and international conferences this year [66], [42].

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 [39]. This work 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 [17].

CARMEN Project-Team

7. New Results

7.1. Modelling, direct simulation and prediction of cardiac phenomena

Using high-performance simulations on a detailed model of the human atria [58] we investigated several aspects of atrial fibrillation (AF). We showed that AF initiation by rapid pacing is sensitive to very small changes in parameter values [70] [32], [38] , and investigated effects of antiarrhythmic drugs and interventions [59] [34] and pathologies [35]. An example movie is available online at https://www.potse.nl/papers/potse/potse-fimh19.html.

High-performance simulations of human ventricular activity have contributed to the testing of new electrocardiographic mapping methods ("inverse models") [73], [36].

We are also developing new methods to help with the treatment of these patients by rapidly guiding an ablation catheter to the origin (strictly speaking: the exit site) of an arrhythmia. These methods are also being tested with simulated data. [33].

We contributed to work by Prof. Michel Haissaguerre and his team in which it is argued that many patients who are now believed to suffer from abnormalities in the genes for specific cardiac ion channels are in reality affected by structural diseases of the heart muscle [62] [26], [41]. These influential publications represent an important change in thinking about these patients and their treatment.

Simple mitochondrial model based on thermodynamic fluxes (PhD work of B. Tarraf): Mitochondria are involved in the regulation of calcium which plays a crucial role in the propagation of cardiac action potentials. However, they are not taken into account in the ionic models that are used to perform simulations at the tissue level. In the framework of the ANR MITOCARD project, we wrote a simple model of mitochondrial calcium regulation based on an extensive review of models of the litterature, which are not suited for further calibration due to their excessive complexity [39]. Now that the equations are written down, we are performing a parameter analysis on the whole model before including other key biological mechanisms.

In a collaboration with Jeremy Darde (IMT Toulouse), we have developed a numerical method on a cartesian grid to solve the direct problem of Electrical Impedance Tomography (EIT) in complex geometries, with first-order convergence. The objective is to solve then the inverse problem of EIT to identify heterogeneities of conductivities on the torso volume.

7.2. Inverse problems: parameter estimation, data assimilation and ECGi

- Data assimilation: In A. Gérard PhD thesis, it is showed that accounting for the anisotropy in the atria is crucial to reconstruct correctly activation maps compatible with a mono-domain model from sparse punctuals activation times. To this purpose, we have developed a new data assimilation method for the mono-domain model, using a Luenberger filter and a Kalman-type filter (ROUKF), based on the dissimilarity measure introduced in A. Collin PhD thesis.
- Parameter estimation: We have been working on the following theoretical question: What are the condition under which the parameters, in the mathematical codomain and bidomain models, are identifiables. Then we proposed an algorithm capable of estimating different ion-channels conductance parameters. mettre ref?
- ECGI: Several approaches have been investigated to improve the resolution of ECGi.
  - development of a new algorithm to choose the regularization coefficient for the resolution of ECGi with the Method of Fondamental Solutions (MFS).
  - a study using a parametrized model of action potentials, showing that accounting for the endocardium can improve the resolution of ECGi.

- in joint work with Laura Bear (IHU Liryc), development of a new method for improving the resolution of ECGI by combining several solutions obtained with various numerical methods (FEM and MFS). The method is based on the selection of the smallest residuals on the torso surface.
- Development of new methods for the ECGI problem, based on machine learning methods. The idea is to learn activation maps from body surface signals.
- collaborative work on the set up of an experimental platform for the experimental non-invasive validation of the reconstruction of cardiac signals

7.3. Numerical schemes

- Very-high order Finite Volume methods: We have showed the very good behavior of a specifically devised domain decomposition technique: the communications are minimized without impacting the accuracy or the order of convergence of the scheme. The total amount of communications dos not increase significantly between the second and the 6th order. The 6th-order Finite-Volume scheme is thus the most performing scheme.
- Numerical analysis of a cartesian method for elliptic problems with immersed interfaces: We have studied the convergence of a cartesian method for elliptic problems with immersed interfaces previously published [54]. The convergence is proved for the original second-order method in one-dimension and for a first-order version in two dimensions. The proof uses a discrete maximum principle to obtain estimates of the coefficients of the inverse matrix.

7.4. Miscellanous

- Refactoring of the CEPS software. Our CEPS software underwent an important refactoring phase so that future students spend less time getting hands on the code. Also, several features developped by previous PhD students were merged into the version that we intend to distribute : high order numerical schemes suited for ionic models, volume fraction due to tissue heterogeneities. We are currently in the process of licensing with Inria and University of Bordeaux.
- Deep brain stimulation procedure: We have been working on three different learning methods for the prediction of the optimal pacing sites in the Deep brain stimulation procedure. We also compared the found position to the position of the stimulation sites to the anatomical geometries in the Ewert brain atlas. Results show the robustness of the methods in founding the stimulation regions.
- Generation of boundary conditions for the Boussinesq system: In a collaboration with David Lannes (IMB, Bordeaux) we have developed a new method for the numerical implementation of generating boundary conditions for a one dimensional Boussinesq system. The method is based on a reformulation of the equations and a resolution of the dispersive boundary layer that is created at the boundary.
- Simulation of solid suspensions in incompressible fluids: In a collaboration with B. Lambert and M. Bergmann (IMB, Bordeaux) we have extended a previously published local lubrication model to non-Brownian suspensions of ellipsoidal solid particles in incompressible flows. This lubrication model used virtual spheres to evaluate local lubrication corrections instead of the global corrections found in the classical lubrication theory.

COMMEDIA Project-Team

7. New Results

7.1. Cardiovascular hemodynamics

Participant: Miguel Ángel Fernández Varela.

Mitral regurgitation is one of the most prevalent valvular heart disease. Proper evaluation of its severity is necessary to choose appropriate treatment. The PISA method, based on Color Doppler echocardiography, is widely used in the clinical setting to estimate various relevant quantities related to the severity of the disease. In [19], the use of a pipeline to quickly generate image-based numerical simulation of intracardiac hemodynamics is investigated. The pipeline capabilities are evaluated on a database of twelve volunteers. Full pre-processing is achieved completely automatically in 55 minutes, on average, with small registration errors compared to the image spatial resolution. This pipeline is then used to study the intracardiac hemodynamics in the presence of diseased mitral valve. A strong variability among the simulated cases, mainly due to the valve geometry and regurgitation specifics, is found. The results from those numerical simulations is used to assess the potential limitations of the PISA method with respect to different MR types. While the PISA method provides reasonable estimates in the case of a simple circular regurgitation, it is shown that unsatisfying estimates are obtained in the case of non-circular leakage. Moreover, it is shown that the choice of high aliasing velocities can lead to difficulties in quantifying MR.

7.2. Respiratory flows

Participant: Céline Grandmont.

In [21], we propose a coupled fluid-kinetic model taking into account the radius growth of aerosol particles due to humidity in the respiratory system. We aim to numerically investigate the impact of hygroscopic effects on the particle behaviour. The air flow is described by the incompressible Navier-Stokes equations, and the aerosol by a Vlasov-type equation involving the airhumidity and temperature, both quantities satisfying a convection-diffusion equation with a sourceterm. Conservations properties are checked and an explicit time-marching scheme is proposed. Two-dimensional numerical simulations in a branched structure show the influence of the particle size variations on the aerosol dynamics.

7.3. Fluid flow reconstruction from medical imaging

Participants: Muriel Boulakia, Miguel Ángel Fernández Varela, Felipe Galarce Marin, Damiano Lombardi, Olga Mula, Colette Voisembert.

In [22], we are interested in designing and analyzing a finite element data assimilation method for laminar steady flow described by the linearized incompressible Navier-Stokes equation. We propose a weakly consistent stabilized finite element method which reconstructs the whole fluid flow from velocity measurements in a subset of the computational domain. Using the stability of the continuous problem in the form of a three balls inequality, we derive quantitative local error estimates for the velocity. Numerical simulations illustrate these convergences properties and we finally apply our method to the flow reconstruction in a blood vessel.

In [29] a state estimation problem is investigated, that consists in reconstructing the blood flow from ultrasound Doppler images. The method proposed is based on a reduced-order technique. Semi-realistic 3D configurations are tested.

7.4. Safety pharmacology

Participants: Damiano Lombardi, Fabien Raphel.

In [31] a greedy method is used to classify molecules action on cardiac myocites. The method is applied to a realistic dataset: the experiments were performed at Ncardia (Netherlands). The experimental dataset is complemented by a synthetic dataset, obtained by simulating experimental meaningful scenarios. The results obtained are encouraging.

7.5. Mathematical analysis of PDEs

Participants: Muriel Boulakia, Jean-Jerome Casanova, Céline Grandmont.

In [23], we consider a reaction-diffusion equation where the reaction term is given by a cubic function and we are interested in the numerical reconstruction of the time-independent part of the source term from measurements of the solution. For this identification problem, we present an iterative algorithm based on Carleman estimates which consists of minimizing at each iteration strongly convex cost functionals. Despite the nonlinear nature of the problem, we prove that our algorithm globally converges and the convergence speed evaluated in weighted norm is linear. In the last part of the paper, we illustrate the effectiveness of our algorithm with several numerical reconstructions in dimension one or two.

In [25] a coupled system of pdes modelling the interaction between a two–dimensional incompressible viscous fluid and a one–dimensional elastic beam located on the upper part of the fluid domain boundary is considered. A good functional framework to define weak solutions in case of contact between the elastic beam and the bottom of the fluid cavity is designed.It is then prove that such solutions exist globally in time regardless a possible contact by approximating the beam equation by a damped beam and letting this additional viscosity vanishes.

7.6. Numerical methods for multi-physics problems

Participants: Miguel Ángel Fernández Varela, Fannie Gerosa.

In [24] we introduce a mixed dimensional Stokes-Darcy coupling where a d-dimensional Stokes' flow is coupled to a Darcy model on the $d-1$ dimensional boundary of the domain. The porous layer introduces tangential creeping flow along the boundary and allows for the modelling of boundary flow due to surface roughness. This leads to a new model of flow in fracture networks with reservoirs in an impenetrable bulk matrix. Exploiting this modelling capability, we then formulate a fluid-structure interaction method with contact, where the porous layer allows for mechanically consistent contact and release. Physical seepage in the contact zone due to rough surfaces is modelled by the porous layer. Some numerical examples are reported, both on the Stokes-Darcy coupling alone and on the fluid-structure interaction with contact in the porous boundary layer.

Unfitted mesh finite element approximations of immersed incompressible fluid-structure interaction problems which efficiently avoid strong coupling without compromising stability and accuracy are rare in the literature. Moreover, most of the existing approaches introduce additional unknowns or are limited by penalty terms which yield ill conditioning issues. In [28], we introduce a new unfitted mesh semi-implicit coupling scheme which avoids these issues. To this purpose, we provide a consistent generalization of the projection based semi-implicit coupling paradigm of [Int. J. Num. Meth. Engrg.,69(4):794-821, 2007] to the unfitted mesh Nitsche-XFEM framework.

7.7. Statistical learning and mathematical modeling interactions

Participants: Damiano Lombardi, Fabien Raphel.

In [30] a greedy dimension reduction method is proposed to deal with classification problems. The method proposed can be seen as a goal oriented dimension reduction method. Elements of a Stiefel manifold (whose dimension is not fixed a priori) are computed in such a way that the classification score is maximised. Several examples are proposed to illustrate the method features and to highlight its differences with classical reduction methods used in classification.

7.8. Tensor approximation and HPC

Participant: Damiano Lombardi.

In [26] a hierarchical adaptive piece-wise tensor decomposition is proposed to approximate high-dimensional functions. Neither the subtensor partitioning nor the rank of the approximation in each of the partitions are fixed a priori. Instead, they are computed to fulfill a prescribed accuracy. Two main contributions are proposed. A greedy error distribution scheme, that allows to adaptively construct the approximation in each of the partitions and a hierarchical tree algorithm that optimise the subtensor partitioning to minimise the storage. Several example on challenging functions are proposed.

7.9. Miscellaneous

Participants: Damiano Lombardi, Olga Mula.

In [20] an approximated formulation of the multi-marginal optimal transport problem (Kantorovich formulation) is proposed. In the formulation, called MCOT, the constraints on the marginal densities are replaced by moments of the densities. This formulation allows to deal simply with a wide spectrum of high-dimensional multi-marginal problems, with non-standard (martingale) constraints.

In [27] a reduced-order modeling framework is proposed, in which a set of model instances is part of a metric space. The introduction of the exponential and logarithmic maps (Riemannian geometry) makes it possible to reduce in an effective way solutions that are classically challenging for standard model reduction methods. Some examples on 1D hyperbolic equations and Wasserstein distance are proposed.

DRACULA Project-Team

5. New Results

5.1. Mathematical models describing the interaction between cancer and immune cells in the lymph node

To study the interplay between tumor progression and the immune response, we develop in [5] two new models describing the interaction between cancer and immune cells in the lymph node. The first model consists of partial differential equations (PDEs) describing the populations of the different types of cells. The second one is a hybrid discrete-continuous model integrating the mechanical and biochemical mechanisms that define the tumor-immune interplay in the lymph node. We use the continuous model to determine the conditions of the regimes of tumor-immune interaction in the lymph node. While we use the hybrid model to elucidate the mechanisms that contribute to the development of each regime at the cellular and tissue levels. We study the dynamics of tumor growth in the absence of immune cells. Then, we consider the immune response and we quantify the effects of immunosuppression and local EGF concentration on the fate of the tumor.

5.2. WASABI: a dynamic iterative framework for gene regulatory network inference

Background Inference of gene regulatory networks from gene expression data has been a long-standing and notoriously difficult task in systems biology. Recently, single-cell transcriptomic data have been massively used for gene regulatory network inference, with both successes and limitations. In the work [8], we propose an iterative algorithm called WASABI, dedicated to inferring a causal dynamical network from time-stamped single-cell data, which tackles some of the limitations associated with current approaches. We first introduce the concept of waves, which posits that the information provided by an external stimulus will affect genes one-by-one through a cascade, like waves spreading through a network. This concept allows us to infer the network one gene at a time, after genes have been ordered regarding their time of regulation. We then demonstrate the ability of WASABI to correctly infer small networks, which have been simulated in silico using a mechanistic model consisting of coupled piecewise-deterministic Markov processes for the proper description of gene expression at the single-cell level. We finally apply WASABI on in vitro generated data on an avian model of erythroid differentiation. The structure of the resulting gene regulatory network sheds a new light on the molecular mechanisms controlling this process. In particular, we find no evidence for hub genes and a much more distributed network structure than expected. Interestingly, we find that a majority of genes are under the direct control of the differentiation-inducing stimulus. Conclusions Together, these results demonstrate WASABI versatility and ability to tackle some general gene regulatory networks inference issues. It is our hope that WASABI will prove useful in helping biologists to fully exploit the power of time-stamped single-cell data.

5.3. A multiscale model of platelet-fibrin thrombus growth in the flow

Thrombosis is a life-threatening clinical condition characterized by the obstruction of blood flow in a vessel due to the formation of a large thrombus. The pathogenesis of thrombosis is complex because the type of formed clots depends on the location and function of the corresponding blood vessel. To explore this phenomenon, we develop in [9] a novel multiscale model of platelet-fibrin thrombus growth in the flow. In this model, the regulatory network of the coagulation cascade is described by partial differential equations. Blood flow is introduced using the Navier–Stokes equations and the clot is treated as a porous medium. Platelets are represented as discrete spheres that migrate with the flow. Each platelet can attach to the thrombus, aggregate, become activated, express proteins on its surface, detach, and/or become non-adhesive. The interaction of platelets with blood flow is captured using the Immersed Boundary Method (IBM). We use the model to

investigate the role of flow conditions in shaping the dynamics of venous and arterial thrombi. We describe the formation of red and white thrombi under venous and arterial flow respectively and highlight the main characteristics of each type. We identify the different regimes of normal and pathological thrombus formation depending on flow conditions.

5.4. Mathematical modeling of platelet production

- In [10], we analyze the existence of oscillating solutions and the asymptotic convergence for a nonlinear delay differential equation arising from the modeling of platelet production. We consider four different cell compartments corresponding to different cell maturity levels: stem cells, megakaryocytic progenitors, megakaryocytes, and platelets compartments, and the quantity of circulating thrombopoietin (TPO), a platelet regulation cytokine.
- In [11], we analyze the stability of a differential equation with two delays originating from a model for a population divided into two subpopulations, immature and mature, and we apply this analysis to a model for platelet production. The dynamics of mature individuals is described by the following nonlinear differential equation with two delays: $x'(t) = -\lambda x(t) + g(x(t-\tau_1)) - g(x(t-\tau_1-\tau_2))e^{-\lambda\tau_2}$. The method of D-decomposition is used to compute the stability regions for a given equilibrium. The center manifold theory is used to investigate the steady-state bifurcation and the Hopf bifurcation. Similarly, analysis of the center manifold associated with a double bifurcation is used to identify a set of parameters such that the solution is a torus in the pseudo- phase space. Finally, the results of the local stability analysis are used to study the impact of an increase of the death rate γ or of a decrease of the survival time τ_2 of platelets on the onset of oscillations. We show that the stability is lost through a small decrease of survival time (from 8.4 to 7 days), or through an important increase of the death rate (from 0.05 to 0.625 days^{-1}).
- In [12], we analyze the stability of a system of differential equations with a threshold-defined delay arising from a model for platelet production. We consider a maturity-structured population of megakaryocyte progenitors and an age-structured population of platelets, where the cytokine thrombopoietin (TPO) increases the maturation rate of progenitors. Using the quasi-steady-state approximation for TPO dynamics and the method of characteristics, partial differential equations are reduced to a system of two differential equations with a state-dependent delay accounting for the variable maturation rate. We start by introducing the model and proving the positivity and boundedness of the solutions. Then we use a change of variables to obtain an equivalent system of two differential equations with a constant delay, from which we prove existence and uniqueness of the solution. As linearization around the unique positive steady state yields a transcendental characteristic equation of third degree, we introduce the main result, a new framework for stability analysis on models with fixed delays. This framework is then used to describe the stability of the megakaryopoiesis with respect to its parameters. Finally, with parameters being obtained and estimated from data, we give an example in which oscillations appear when the death rate of progenitors is increased 10-fold.

5.5. Nonlinear analysis of a model for yeast cell communication

In [13], we study the non-linear stability of a coupled system of two non-linear transport-diffusion equations set in two opposite half-lines. This system describes some aspects of yeast pairwise cellular communication, through the concentration of some protein in the cell bulk and at the cell boundary. We show that it is of bistable type, provided that the intensity of active molecular transport is large enough. We prove the non-linear stability of the most concentrated steady state, for large initial data, by entropy and comparison techniques. For small initial data we prove the self-similar decay of the molecular concentration towards zero. Informally speaking, the rise of a dialog between yeast cells requires enough active molecular transport in this model. Besides, if the cells do not invest enough in the communication with their partner, they do not respond to each other; but a sufficient initial input from each cell in the dialog leads to the establishment of a stable activated state in both cells.

5.6. Alzheimer's disease and prion: An in vitro mathematical model

Alzheimer's disease (AD) is a fatal incurable disease leading to progressive neuron destruction. AD is caused in part by the accumulation in the brain of $A\beta$ monomers aggregating into oligomers and fibrils. Oligomers are amongst the most toxic structures as they can interact with neurons via membrane receptors, including PrP^c proteins. This interaction leads to the misconformation of PrP^c into pathogenic oligomeric prions, PrP^{ol}. In [14], we develop a model describing in vitro $A\beta$ polymerization process. We include interactions between oligomers and PrP^c, causing the misconformation of PrP^c into PrP^{ol}. The model consists of nine equations, including size structured transport equations, ordinary differential equations and delayed differential equations. We analyse the well-posedness of the model and prove the existence and uniqueness of solutions of our model using Schauder fixed point theorem and Cauchy-Lipschitz theorem. Numerical simulations are also provided to give an illustration of the profiles that can be obtained with this model.

5.7. Calibration, Selection and Identifiability Analysis of a Mathematical Model of the in vitro Erythropoiesis in Normal and Perturbed Contexts

The in vivo erythropoiesis, which is the generation of mature red blood cells in the bone marrow of whole organisms, has been described by a variety of mathematical models in the past decades. However, the in vitro erythropoiesis, which produces red blood cells in cultures, has received much less attention from the modelling community. In the paper [15], we propose the first mathematical model of in vitro erythropoiesis. We start by formulating different models and select the best one at fitting experimental data of in vitro erythropoietic differentiation obtained from chicken erythroid progenitor cells. It is based on a set of linear ODE, describing 3 hypothetical populations of cells at different stages of differentiation. We then compute confidence intervals for all of its parameters estimates, and conclude that our model is fully identifiable. Finally, we use this model to compute the effect of a chemical drug called Rapamycin, which affects all states of differentiation in the culture, and relate these effects to specific parameter variations. We provide the first model for the kinetics of in vitro cellular differentiation which is proven to be identifiable. It will serve as a basis for a model which will better account for the variability which is inherent to the experimental protocol used for the model calibration.

5.8. Model-based assessment of the role of uneven partitioning of molecular content on heterogeneity and regulation of differentiation in CD8 T-cell immune responses

Activation of naive CD8 T-cells can lead to the generation of multiple effector and memory subsets. Multiple parameters associated with activation conditions are involved in generating this diversity that is associated with heterogeneous molecular contents of activated cells. Although naive cell polarisation upon antigenic stimulation and the resulting asymmetric division are known to be a major source of heterogeneity and cell fate regulation, the consequences of stochastic uneven partitioning of molecular content upon subsequent divisions remain unclear yet. In [16], we aim at studying the impact of uneven partitioning on molecular-content heterogeneity and then on the immune response dynamics at the cellular level. To do so, we introduce a multiscale mathematical model of the CD8 T-cell immune response in the lymph node. In the model, cells are described as agents evolving and interacting in a 2D environment while a set of differential equations, embedded in each cell, models the regulation of intra and extracellular proteins involved in cell differentiation. Based on the analysis of in silico data at the single cell level, we 1 show that immune response dynamics can be explained by the molecular-content heterogeneity generated by uneven partitioning at cell division. In particular, uneven partitioning acts as a regulator of cell differentiation and induces the emergence of two coexisting sub-populations of cells exhibiting antagonistic fates. We show that the degree of unevenness of molecular partitioning, along all cell divisions, affects the outcome of the immune response and can promote the generation of memory cells.

5.9. Spatial lymphocyte dynamics in lymph nodes predicts the cytotoxic T-Cell frequency needed for HIV infection control

The surveillance of host body tissues by immune cells is central for mediating their defense function. In vivo imaging technologies have been used to quantitatively characterize target cell scanning and migration of lymphocytes within lymph nodes (LNs). The translation of these quantitative insights into a predictive understanding of immune system functioning in response to various perturbations critically depends on computational tools linking the individual immune cell properties with the emergent behavior of the immune system. By choosing the Newtonian second law for the governing equations, we developed in [17] a broadly applicable mathematical model linking individual and coordinated T-cell behaviors. The spatial cell dynamics is described by a superposition of autonomous locomotion, intercellular interaction, and viscous damping processes. The model is calibrated using in vivo data on T-cell motility metrics in LNs such as the translational speeds, turning angle speeds, andmeandering indices. Themodel is applied to predict the impact of T-cell motility on protection against HIV infection, i.e., to estimate the threshold frequency of HIV-specific cytotoxic T cells (CTLs) that is required to detect productively infected cells before the release of viral particles starts. With this, it provides guidance for HIV vaccine studies allowing for the migration of cells in fibrotic LNs.

5.10. Drugs modulating stochastic gene expression affect the erythroid differentiation process

To better understand the mechanisms behind cells decision-making to differentiate, we assessed in [18] the influence of stochastic gene expression (SGE) modulation on the erythroid dif- ferentiation process. It has been suggested that stochastic gene expression has a role in cell fate decision-making which is revealed by single-cell analyses but studies dedicated to demonstrate the consistency of this link are still lacking. Recent observations showed that SGE significantly increased during differentiation and a few showed that an increase of the level of SGE is accompanied by an increase in the differentiation process. However, a consistent relation in both increasing and decreasing directions has never been shown in the same cellular system. Such demonstration would require to be able to experimentally manipulate simultaneously the level of SGE and cell differentiation in order to observe if cell behavior matches with the current theory. We identified three drugs that modulate SGE in primary erythroid progenitor cells. Both Artemisinin and Indomethacin decreased SGE and reduced the amount of differentiated cells. On the contrary, a third component called MB-3 simultaneously increased the level of SGE and the amount of differentiated cells. We then used a dynamical modelling approach which confirmed that differentiation rates were indeed affected by the drug treatment. Using single-cell analysis and modeling tools, we provide experimental evidence that, in a physiologically relevant cellular system, SGE is linked to differentiation.

5.11. Stochastic gene expression with a multistate promoter: breaking down exact distributions

We consider in [19] a stochastic model of gene expression in which transcription depends on a multistate promoter, including the famous two-state model and refractory promoters as special cases, and focus on deriving the exact stationary distribution. Building upon several successful approaches, we present a more unified viewpoint that enables us to simplify and generalize existing results. In particular, the original jump process is deeply related to a multivariate piecewise-deterministic Markov process that may also be of interest beyond the biological field. In a very particular case of promoter configuration, this underlying process is shown to have a simple Dirichlet stationary distribution. In the general case, the corresponding marginal distributions extend the well-known class of Beta products, involving complex parameters that directly relate to spectral properties of the promoter transition matrix. Finally, we illustrate these results with biologically plausible examples.

5.12. Cell generation dynamics underlying naive T-cell homeostasis in adult humans

Thymic involution and proliferation of naive T-cells both contribute to shaping the naive T-cell repertoire as humans age, but a clear understanding of the roles of each throughout a human life span has been difficult to determine. By measuring nuclear bomb test-derived 14 C in genomic DNA, we determined in [22] the turnover rates of CD4 + and CD8 + naive T-cell populations and defined their dynamics in healthy individuals ranging from 20 to 65 years of age. We demonstrate that naive T-cell generation decreases with age because of a combination of declining peripheral division and thymic production during adulthood. Concomitant decline in T-cell loss compensates for decreased generation rates. We investigated putative mechanisms underlying age-related changes in homeostatic regulation of CD4+ naive T-cell turnover, using mass cytometry to profile candidate signaling pathways involved in T-cell activation and proliferation relative to CD31 expression, a marker of thymic proximity for the CD4+ naive T-cell population. We show that basal nuclear factor kB (NF-kB) phosphory-lation positively correlated with CD31 expression and thus is decreased in peripherally expanded naive T-cell clones. Functionally, we found that NF-kB signaling was essential for naive T-cell proliferation to the homeostatic growth factor interleukin (IL)-7, and reduced NF-kB phosphorylation in CD4 + CD31 - naive T cells is linked to reduced homeostatic proliferation potential. Our results reveal an age-related decline in naive T-cell turnover as a putative regulator of naive T-cell diversity and identify a molecular pathway that restricts proliferation of peripherally expanded naive T-cell clones that accumulate with age.

5.13. Erythroid differentiation displays a peak of energy consumption concomitant with glycolytic metabolism rearrangements

Our previous single-cell based gene expression analysis pointed out significant variations of LDHA level during erythroid differentiation. Deeper investigations highlighted that a metabolic switch occurred along differentiation of erythroid cells. More precisely we showed in [26] that self-renewing progenitors relied mostly upon lactate-productive glycolysis, and required LDHA activity, whereas differentiating cells, mainly involved mitochondrial oxidative phosphorylation (OXPHOS). These metabolic rearrangements were coming along with a particular temporary event, occurring within the first 24h of erythroid differentiation. The activity of glycolytic metabolism and OXPHOS rose jointly with oxgene consumption dedicated to ATP production at 12-24h of the differentiation process before lactate-productive glycolysis sharply fall down and energy needs decline. Finally, we demonstrated that the metabolic switch mediated through LDHA drop and OXPHOS upkeep might be necessary for erythroid differentiation. We also discuss the possibility that metabolism, gene expression and epigenetics could act together in a circular manner as a driving force for differentiation.

M3DISIM Project-Team

6. New Results

6.1. Mathematical and Mechanical Modeling

6.1.1. Stochastic modeling of chemical-mechanical coupling in striated muscles

Participants: Matthieu Caruel, Philippe Moireau, Dominique Chapelle [correspondant].

In [18] 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.

6.1.2. Upscaling of elastic network models

Participant: Patrick Le Tallec.

This work is done in collaboration with Julie Diani from École Polytechnique. The purpose of the approach is to develop general upscaling strategy for deriving macroscopic constitutive laws for rubberlike materials from the knowledge of the network distribution and a mechanical description of the individual chains and of their free energy. It is based on a variational approach in which the microscopic configuration is described by the position of the crosslinks and is obtained not by an affine assumption but by minimizing the corresponding free energy on stochastic large Representative Volume Elements with adequate boundary conditions. This general framework is then approximated by using a microsphere (directional) description of the network and by performing a local minimisation of the network free energy on this simplified configuration space under a maximal advance path kinematic constraint. This approximation framework takes into account anisotropic damage and is extended to handle situations with tube like constraints and stress induced cristallisation. For more detail see [23].

6.1.3. Stochastic construction of surrogate multiphase materials

Participant: Patrick Le Tallec.

Random microstructures of heterogeneous materials play a crucial role in the material macroscopic behavior and in predictions of its effective properties. A common approach to modeling random multiphase materials is to develop so-called surrogate models approximating statistical features of the material. However, the surrogate models used in fatigue analysis usually employ simple microstructure, consisting of ideal geometries such as ellipsoidal inclusions, which generally does not capture complex geometries. In our work , we introduce a simple but flexible surrogate microstructure model for two-phase materials through a level-cut of a Gaussian random field with covariance of Matern class. In addition to the traditional morphology descriptors such as porosity, size and aspect ratio of inclusions, our approach provides control of the regularity of the inclusions interface and sphericity. These parameters are estimated from a small number of real material images using

Bayesian inversion. An efficient process of evaluating the samples, based on the Fast Fourier Transform, makes possible the use of Monte-Carlo methods to estimate statistical properties for the quantities of interest in a given material class. This work in progress is done in collaboration with Andrei Constantinescu (École Polytechnique), Ustim Khristenko and Barbara Wohlmuth (Technical University Munich) and Tinsley Oden (University of Texas at Austin). This work has been submitted for publication in an international journal.

6.1.4. Apprehending the effects of mechanical deformations in cardiac electrophysiology – A homogenization approach

Participants: Annabelle Collin [MONC], Sébastien Imperiale, Philippe Moireau, Jean-Frédéric Gerbeau [Inria Siège], Dominique Chapelle [correspondant].

In this work [22], we follow a formal homogenization approach to investigate the effects of mechanical deformations in electrophysiology models relying on a bidomain description of ionic motion at the microscopic level. To that purpose, we extend these microscopic equations to take into account the mechanical deformations, and proceed by recasting the problem in the frame- work of classical two-scale homogenization in periodic media, and identifying the equations satisfied by the first coefficients in the formal expansions. The homogenized equations reveal some interesting effects related to the microstructure – and associated with a specific cell problem to be solved to obtain the macroscopic conductivity tensors – in which mechanical deformations play a non-trivial role, i.e. do not simply lead to a standard bidomain problem posed in the deformed configuration. We then present detailed numerical illustrations of the homogenized model with coupled cardiac electrical-mechanical simulations – all the way to ECG simulations – albeit without taking into account the abundantly-investigated effect of mechanical deformations in ionic models, in order to focus here on other effects. And in fact our numerical results indicate that these other effects are numerically of a comparable order, and therefore cannot be disregarded.

6.1.5. Patient-specific pulmonary mechanics - Modelling and estimation. Application to pulmonary fibrosis.

Participants: Cecile Patte [correspondant], Martin Genet, Dominique Chapelle.

Interstitial pulmonary diseases, like Idiopathic Pulmonary Fibrosis (IPF), affect the alveolar structure of lung tissue, which impacts lung mechanical properties and pulmonary functions. In this work [43], we aim to better understand the pulmonary mechanics in order to improve IPF diagnosis. We developped a poromechanical model for the lung at the organ scale and at the breathing scale. This model is then used to estimate regional mechanical parameters based on clinical data. In the future, this process can be used as an augmented diagnosis tool for clinicians. This work has been presented at the CSMA conference.

6.1.6. Energy preserving cardiac circulation models: formulation, reduction, coupling, inversion, and discretization

Participants: Jessica Manganotti, Philippe Moireau, Sébastien Imperiale [correspondant], Miguel Fernandez [Inria Paris, COMMEDIA].

The modeling of the heart cannot be satisfying if not coupled to the body circulation, and at least to the arterial circulation, which is its direct output “boundary condition”. But more importantly in the clinical context, it is still difficult – ant very invasive – to access the ventricular pressure, which is absolutely necessary for specifying the heart activity. By comparison, more and more devices allow to register non-invasively a distal pressure, for instance at the wrist or the finger, which could be used to estimate the ventricular pressure by inversion of a well adapted arterial circulation model. Such relation is of major interest for clinicians, for example anesthetists, since it could allow real-time monitoring and prediction of the effects of injected drugs during a clinical intervention. Models of the arterial circulation is a well-known subject where dimension reduction has been widely studied for more than half a century. However, the question remains of formulating energy-consistent formulations that can be consistently maintained during the reduction, when coupled to a heart model, and also when discretized. Yet the question is crucial for a better understanding of the physical phenomena of blood flow ejection from the heart as well as the propagation in the arterial network. Moreover,

as these models are non-linear, energy-preserving approaches are one of the few tools at our disposal to mathematically justify modeling, discretization or inversion approaches. Finally, inverting this unsteady model for estimation purposes of medical data also benefits from energy-preserving formulation as the inverse approach should also satisfy some stability properties. The subject here is twofold and part of the thesis of J. Manganotti. First we plan to develop accurate models, coupling strategies and robust numerical methods of the arterial network propagation coupled to the heart. Second, we want to develop observer-based strategies that will allow to easily feed these models with measurements in order to perform state estimation of hidden variables or identify key biophysical parameters.

6.1.7. Hierarchical modeling of force generation in cardiac muscle

Participants: Matthieu Caruel, François Kimmig [correspondant].

Performing physiologically relevant simulations of the beating heart in clinical context requires to develop detailed models of the microscale force generation process. These models however may be difficult to implement in practice due to their high computational costs and complex calibration. We propose a hierarchy of three interconnected cardiac muscle contraction models – from the more refined to the more simplified – that are rigorously and systematically related with each other, offering a way to select, for a specific application, the model that yields the best trade-off between physiological fidelity, computational cost and calibration complexity. Our starting model takes into account the stochastic dynamics of the molecular motors force producing conformational changes– and in particular the power stroke – and captures all the timescales of appearing in classical experimental isotonic responses of a heart papillary muscle submitted to rapid load changes. Adiabatic elimination of fast relaxing variables of the stochastic model yields a formulation based on partial differential equations (PDEs) that falls into the family of the Huxley'57 model, while embedding some properties of the process occurring at the fastest timescales. The third family of models is deduced from the PDE model by making minimal assumptions on the parameters, which leads to a computationally light formulation based on ordinary differential equations only. The three models families are compared to the same set of experimental data to systematically assess what physiological indicators can be reproduced or not and how these indicators constrain the model parameters. Finally, we discuss the applicability of these models for heart simulation. This work has been submitted for publication in an international journal.

6.1.8. A relaxed growth modeling framework for controlling growth-induced residual stresses

Participant: Martin Genet.

Background Constitutive models of the mechanical response of soft tissues have been established and are widely accepted, but models of soft tissues remodeling are more controversial. Specifically for growth, one important question arises pertaining to residual stresses: existing growth models inevitably introduce residual stresses, but it is not entirely clear if this is physiological or merely an artifact of the modeling framework. As a consequence, in simulating growth, some authors have chosen to keep growth-induced residual stresses, and others have chosen to remove them. **Methods** In this work, we introduce a novel "relaxed growth" framework allowing for a fine control of the amount of residual stresses generated during tissue growth. It is a direct extension of the classical framework of the multiplicative decomposition of the transformation gradient, to which an additional sub-transformation is introduced in order to let the original unloaded configuration evolve, hence relieving some residual stresses. We provide multiple illustrations of the framework mechanical response, on time-driven constrained growth as well as the strain-driven growth problem of the artery under internal pressure, including the opening angle experiment. **Findings** The novel relaxed growth modeling framework introduced in this paper allows for a better control of growth-induced residual stresses compared to standard growth models based on the multiplicative decomposition of the transformation gradient. **Interpretation** Growth-induced residual stresses should be better handled in soft tissues biomechanical models, especially in patient-specific models of diseased organs that are aimed at augmented diagnosis and treatment optimization. See [27] for more detail.

6.1.9. Multiscale population dynamics in reproductive biology: singular perturbation reduction in deterministic and stochastic models

Participants: Frédérique Clément [correspondant], Romain Yvinec.

During the supervision of a CEMRACS2018 project performed by Céline Bonnet (CMAP) and Keltoum Chahour (LERMA and JLAD), we have described (with Marie Postel, Sorbonne Université and Romain Yvinec, INRA) different modeling approaches for ovarian follicle population dynamics, based on either ordinary (ODE), partial (PDE) or stochastic (SDE) differential equations, and accounting for interactions between follicles [50]. We have put a special focus on representing the population-level feedback exerted by growing ovarian follicles onto the activation of quiescent follicles. We have taken advantage of the timescale difference existing between the growth and activation processes to apply model reduction techniques in the framework of singular perturbations. We have first studied the linear versions of the models to derive theoretical results on the convergence to the limit models. In the nonlinear cases, we have provided detailed numerical evidence of convergence to the limit behavior. We have reproduced the main semi-quantitative features characterizing the ovarian follicle pool, namely a bimodal distribution of the whole population, and a slope break in the decay of the quiescent pool with aging.

6.1.10. Stochastic nonlinear model for somatic cell population dynamics during ovarian follicle activation

Participants: Frédérique Clément [correspondant], Frédérique Robin, Romain Yvinec.

In mammals, female germ cells are sheltered within somatic structures called ovarian follicles, which remain in a quiescent state until they get activated, all along reproductive life. We have investigated the sequence of somatic cell events occurring just after follicle activation [54]. We have introduced a nonlinear stochastic model accounting for the joint dynamics of two cell types, either precursor or proliferative cells. The initial precursor cell population transitions progressively to a proliferative cell population, by both spontaneous and self-amplified processes. In the meantime, the proliferative cell population may start either a linear or exponential growing phase. A key issue is to determine whether cell proliferation is concomitant or posterior to cell transition, and to assess both the time needed for all precursor cells to complete transition and the corresponding increase in the cell number with respect to the initial cell number. Using the probabilistic theory of first passage times, we have designed a numerical scheme based on a rigorous Finite State Projection and coupling techniques to assess the mean extinction time and the cell number at extinction time. We have also obtained analytical formulas for an approximating branching process. We have calibrated the model parameters using an exact likelihood approach using both experimental and in-silico datasets. We have carried out a comprehensive comparison between the initial model and a series of submodels, which help to select the critical cell events taking place during activation. We have finally interpreted these results from a biological viewpoint.

6.1.11. A multiscale mathematical model of cell dynamics during neurogenesis in the mouse cerebral cortex

Participant: Frédérique Clément.

This work is a collaboration with Marie Postel and Sylvie Schneider-Maunoury (Sorbonne Université), 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. In this work [33] we aim at understanding 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. A special focus is put on the population of intermediate progenitors (IPs), a transit amplifying progenitor type critically involved in the size of the final neuron pool. A 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, reveals a shortening of the neurogenic period associated with an increased influx of newborn IPs from apical progenitors at mid-neurogenesis. Additional information is 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.

6.2. Numerical Methods

6.2.1. Numerical analysis for an energy-stable total discretization of a poromechanics model with inf-sup stability

Participants: Dominique Chapelle [correspondant], Philippe Moireau.

In this joint work with Bruno Burtschell [16], 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.

6.2.2. Conservative and entropy controlled remap for multi-material ALE simulations

Participant: Patrick Le Tallec.

For many multi-material problems such as fluid-structure interaction, impact or implosion problems, materials are in very large strains due to their nature or to the applied forces. In our situations of interest, we also have a strong coupling between energy and momentum conservation laws, due to intense transfers between internal and kinetic energies and to strong advection effects. Such situations are classically governed by the Euler's equations, written in Lagrangian form, and using a multi-material, single velocity framework, but their numerical solution demands a strict control of energy conservation and entropy production, which is hard to achieve in situations where dynamic remeshing is mandatory. In this framework, our approach deals with the analysis of the impact of a second-order staggered remap using an intersection-based approach on conservation properties and on the entropy control. We show that an accurate remap with exact mesh intersections and exact integrations affects both the momentum and the kinetic energy because of node mass re-localizations and node velocity remap. We propose therefore a staggered remapping strategy in order to take into account these discrepancies at a low computational cost. While preserving the strict conservation of total energy, our strategy allows to recover a proper entropy control at the expense of strict momentum conservation and monotonicity losses. This work [32] is done in collaboration with Alexandra Claisse (CEA DAM) and Alexis Marboeuf (École Polytechnique and CEA DAM).

6.2.3. Multipatch isogeometric analysis for complex structures

Participant: Patrick Le Tallec.

This work – done in collaboration with Nicolas Adam (École Polytechnique and PSA) and Malek Zarroug (PSA) – introduces, analyzes and validates isogeometric mortar methods for the solution of thick shells problems which are set on a multipatch geometry. It concerns industrial parts of complex geometries for which the effects of transverse shear cannot be neglected. For this purpose, Reissner-Mindlin model was retained and rotational degrees of freedom (DOF) of the normal are taken into account. A particular attention is devoted to the introduction of a proper formulation of the coupling conditions at patches interfaces, with a particular interest on augmented lagrangian formulations, to the choice and validation of mortar spaces, and to the derivation of adequate integration rules. The relevance of the proposed approach is assessed numerically on various significative examples of industrial relevance. This work has been submitted for publication in an international journal.

6.2.4. Mathematical and numerical study of transient wave scattering by obstacles with the Arlequin Method

Participant: Sébastien Imperiale.

In this work [14] we extend the Arlequin method to overlapping domain decomposition technique for transient wave equation scattering by obstacles. The main contribution of this work is to construct and analyze from the continuous level up to the fully discrete level some variants of the Arlequin method. The constructed discretizations allow to solve wave propagation problems while using non-conforming and overlapping meshes for the background propagating medium and the surrounding of the obstacle respectively. Hence we obtain a flexible and stable method in terms of the space discretization – an inf-sup condition is proven – while the stability of the time discretization is ensured by energy identities.

6.2.5. Construction and analysis of fourth-order, energy consistent, family of explicit time discretizations for dissipative linear wave equations

Participants: Juliette Chabassier [MAGIQUE-3D], Julien Diaz [MAGIQUE-3D], Sébastien Imperiale [correspondant].

This work and the corresponding article [19], deal with the construction of a family of fourth order, energy consistent, explicit time discretizations for dissipative linear wave equations. The schemes are 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 explicit approximations of its inverse. The stability of the schemes are studied using an energy analysis and a convergence analysis is carried out. Numerical results in 1D illustrate the space/time convergence properties of the schemes and their efficiency is compared to more classical time discretizations.

6.2.6. Energy decay and stability of a perfectly matched layer For the wave equation

Participants: Sébastien Imperiale [correspondant], Maryna Kachanovska [POEMS].

We follow a previous work where PML formulations was proposed for the wave equation in its standard second-order form. In the present work [15], energy decay and L^2 stability bounds in two and three space dimensions are rigorously proved both for continuous and discrete formulations with constant damping coefficients. Numerical results validate the theory.

6.2.7. A high-order spectral element fast Fourier transform for the poisson equation

Participants: Federica Caforio, Sébastien Imperiale [correspondant].

The aim of this work [17] is to propose a novel, fast solver for the Poisson problem discretised with High-Order Spectral Element Methods (HO-SEM) in canonical geometries (rectangle in 2D, rectangular parallelepiped in 3D). 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 the frequency space. The proposed solver is endowed with several properties. First, it preserves the efficiency of the standard FFT algorithm; then, the matrix storage is drastically reduced (in particular, it is independent of the space dimension); a pseudo-explicit Singular Value Decomposition (SVD) is used for the inversion of the symbols; finally, it can be extended to non-periodic boundary conditions. Furthermore, 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.

6.2.8. Thermodynamic properties of muscle contraction models and associated discrete-time principles

Participants: François Kimmig, Dominique Chapelle [correspondant], 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. See [30] for more detail.

6.2.9. Mechanical and imaging models-based image registration

Participants: Radomir Chabiniok, Martin Genet [correspondant].

Image registration plays an increasingly important role in many fields such as biomedical or mechanical engineering. Generally speaking, it consists in deforming a (moving) source image to match a (fixed) template image. Many approaches have been proposed over the years; if new model-free machine learning-based approaches are now beginning to provide robust and accurate results, extracting motion from images is still most commonly based on combining some statistical analysis of the images intensity and some model of the underlying deformation as initial guess or regularizer. These approaches may be efficient even for complex type of motion; however, any artifact in the source image (e.g., partial voluming, local decrease of signal-to-noise ratio or even local signal void), drastically deteriorates the registration. This work introduces a novel approach of extracting motion from biomedical image series, based on a model of the imaging modality. It is, to a large extent, independent of the type of model and image data – the pre-requisite is to incorporate biomechanical constraints into the motion of the object (organ) of interest and being able to generate data corresponding to the real image, i.e., having an imaging model at hand. We will illustrate the method with examples of synthetically generated 2D tagged magnetic resonance images. This work was presented at the VipIMAGE 2019 conference. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD. See [44] for more detail. This work was done in collaboration with Katerina Skardova (Czech Technical University in Prague) and Matthias Rambausek (École Polytechnique).

6.2.10. Validation of finite element image registration-based cardiac strain estimation from magnetic resonance images

Participants: Martin Genet [correspondant], Philippe Moireau.

Accurate assessment of regional and global function of the heart is an important readout for the diagnosis and routine evaluation of cardiac patients. Indeed, recent clinical and experimental studies suggest that compared to global metrics, regional measures of function could allow for more accurate diagnosis and early intervention for many cardiac diseases. Although global strain measures derived from tagged magnetic resonance (MR) imaging have been shown to be reproducible for the majority of image registration techniques, the measurement of regional heterogeneity of strain is less robust. Moreover, radial strain is underestimated with the current techniques even globally. Finite element (FE)-based techniques offer a mechanistic approach for the regularization of the ill-posed registration problem. This work presents the validation of a recently proposed FE-based image registration method with mechanical regularization named equilibrated warping. For this purpose, synthetic 3D-tagged MR images are generated from a reference biomechanical model of the left ventricle (LV). The performance of the registration algorithm is consequently tested on the images with different signal-to-noise ratios (SNRs), revealing the robustness of the method. See [35] for more detail.

6.3. Inverse Problems

6.3.1. Analysis of an observer strategy for initial state reconstruction of wave-like systems in unbounded domains

Participants: Sébastien Imperiale, Philippe Moireau [correspondant].

In [29] 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 stabilisation properties obtained from the use of the available measurements, hence allowing to still be able to reconstruct the unknown initial condition.

6.3.2. *Analysis and numerical simulation of an inverse problem for a structured cell population dynamics model*

Participants: Frédérique Clément, Frédérique Robin [correspondant].

We have studied (with Béatrice Laroche, INRA) a multiscale inverse problem associated with a multi-type model for age structured cell populations [20] (see also [21] for another application). In the single type case, the model is a McKendrick-VonFoerster like equation with a mitosis-dependent death rate and potential migration at birth. In the multi-type case, the migration term results in a unidirectional motion from one type to the next, so that the boundary condition at age 0 contains an additional extrinsic contribution from the previous type. We consider the inverse problem of retrieving microscopic information (the division rates and migration proportions) from the knowledge of macroscopic information (total number of cells per layer), given the initial condition. We have first shown the well-posedness of the inverse problem in the single type case using a Fredholm integral equation derived from the characteristic curves, and we have used a constructive approach to obtain the lattice division rate, considering either a synchronized or non-synchronized initial condition. We have taken advantage of the unidirectional motion to decompose the whole model into nested submodels corresponding to self-renewal equations with an additional extrinsic contribution. We have again derived a Fredholm integral equation for each submodel and deduced the well-posedness of the multi-type inverse problem. In each situation, we illustrate numerically our theoretical results.

6.3.3. *Inverse problem based on data assimilation approaches for protein aggregation*

Participants: Philippe Moireau [correspondant], Cécile Della Valle [MAMBA], Marie Doumic [MAMBA].

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation. In a depolarization configuration, we here extend some previous results obtained during the PhD Thesis of A. Armiento. Now, the depolarization rate is time-dependent or in the presence of an additional vanishing viscosity term. We continue to develop our framework mixing inverse problems methodologies and optimal control approaches typically encountered in data assimilation, allowing to justify mathematically the methods but also to adopt efficient numerical strategies. Publications of this work will be soon submitted.

6.3.4. *Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the RxCADRE field-scale experiment*

Participants: Annabelle Collin [MONC], Philippe Moireau [correspondant].

Data-driven wildfire spread modeling is emerging as a cornerstone for forecasting real-time fire behavior using thermal-infrared imaging data. One key challenge in data assimilation lies in the design of an adequate measure to represent the discrepancies between observed and simulated firelines (or “fronts”). A first approach consists in adopting a Lagrangian description of the flame front and in computing a Euclidean distance between simulated and observed fronts by pairing each observed marker with its closest neighbor along the simulated front. However, this front marker registration approach is difficult to generalize to complex front topology that can occur when fire propagation conditions are highly heterogeneous due to topography, biomass fuel and micrometeorology. To overcome this issue, we investigate in this paper an object-oriented approach derived from the Chan–Vese contour fitting functional used in image processing. The burning area is treated as a

moving object that can undergo shape deformations and topological changes. We combine this non-Euclidean measure with a state estimation approach (a Luenberger observer) to perform simulations of the time-evolving fire front location driven by discrete observations of the fireline. We apply this object-oriented data assimilation method to the three-hectare RxCADRE S5 field-scale experiment. This collaboration with CERFACS (M. Rochoux) and University of Maryland (C. Zhang and A. Trouvé) led to a publication [34] in the Proceedings of the Combustion Institute.

6.3.5. *Model assessment through data assimilation of realistic data in cardiac electrophysiology*

Participants: Antoine Gerard [CARMEN], Annabelle Collin [MONC], Gautier Bureau, Philippe Moireau [correspondant], Yves Coudière [CARMEN].

We consider a model-based estimation procedure – namely a data assimilation algorithm – of the atrial depolarization state of a sub- ject using data corresponding to electro-anatomical maps. Our objective is to evaluate the sensitivity of such a model-based reconstruction with respect to model choices. The followed data assimilation approach is capable of using electrical activation times to adapt a monodomain model simulation, thanks to an ingenious model-data fitting term inspired from image processing. The resulting simulation smoothes and completes the activation maps when they are spatially incomplete. Moreover, conductivity parameters can also be inferred. The model sensitivity assessment is performed based on synthetic data generated with a validated realistic atria model and then inverted using simpler modeling ingredients. In particular, the impact of the muscle fibers definition and corresponding anisotropic conductivity parameters is studied. Finally, an application of the method to real data is presented, showing promising results. This collaborative work has been published, see [37].

6.4. Experimental Assessments

6.4.1. *Combination of traction assays and multiphoton imaging to quantify skin biomechanics*

Participant: Jean-Marc Allain.

An important issue in tissue biomechanics is to decipher the relationship between the mechanical behavior at macroscopic scale and the organization of the collagen fiber network at microscopic scale. We have formalized a definitive protocol [46] to combine traction assays with multiphoton microscopy in ex vivo murine skin. This multiscale approach provides simultaneously the stress/stretch response of a skin biopsy and the collagen reorganization in the dermis by use of second harmonic generation (SHG) signals and appropriate image processing.

6.4.2. *Monitoring dynamic collagen reorganization during skin stretching with fast polarization-resolved second harmonic generation imaging*

Participant: Jean-Marc Allain.

The mechanical properties of biological tissues are strongly correlated to the specific distribution of their collagen fibers. Monitoring the dynamic reorganization of the collagen network during mechanical stretching is however a technical challenge, because it requires mapping orientation of collagen fibers in a thick and deforming sample. In this work [24], a fast polarization-resolved second harmonic generation microscope is implemented to map collagen orientation during mechanical assays. This system is based on line-to-line switching of polarization using an electro-optical modulator and works in epi-detection geometry. After proper calibration, it successfully highlights the collagen dynamic alignment along the traction direction in ex vivo murine skin dermis. This microstructure reorganization is quantified by the entropy of the collagen orientation distribution as a function of the stretch ratio. It exhibits a linear behavior, whose slope is measured with a good accuracy. This approach can be generalized to probe a variety of dynamic processes in thick tissues.

6.4.3. *Multiscale characterisation of skin mechanics through in-situ imaging*

Participant: Jean-Marc Allain.

The complex mechanical properties of skin have been studied intensively over the past decades. They are intrinsically linked to the structure of the skin at several length scales, from the macroscopic layers (epidermis, dermis and hypodermis) down to the microstructural organization at the molecular level. Understanding the link between this microscopic organization and the mechanical properties is of significant interest in the cosmetic and medical fields. Nevertheless, it only recently became possible to directly visualize the skin's microstructure during mechanical assays, carried out on the whole tissue or on isolated layers. These recent observations have provided novel information on the role of structural components of the skin in its mechanical properties, mainly the collagen fibers in the dermis, while the contribution of others, such as elastin fibers, remains elusive. We performed in [45] a systematic review of the current methods used to observe skin's microstructure during a mechanical assay, along with their strengths and limitations, as well as a review of the unique information they provide on the link between structure and function of the skin.

6.4.4. Root Hair Sizer: an algorithm for high throughput recovery of different root hair and root developmental parameters

Participant: Jean-Marc Allain.

The root is an important organ for water and nutrient uptake, and soil anchorage. It is equipped with root hairs (RHs) which are elongated structures increasing the exchange surface with the soil. RHs are also studied as a model for plant cellular development, as they represent a single cell with specific and highly regulated polarized elongation. For these reasons, it is useful to be able to accurately quantify RH length employing standardized procedures. Methods commonly employed rely on manual steps and are therefore time consuming and prone to errors, restricting analysis to a short segment of the root tip. Few partially automated methods have been reported to increase measurement efficiency. However, none of the reported methods allow an accurate and standardized definition of the position along the root for RH length measurement, making data comparison difficult. In this work [28] we are developing an image analysis algorithm that semi-automatically detects RHs and measures their length along the whole differentiation zone of roots. This method, implemented as a simple automated script in ImageJ/Fiji software that we termed Root Hair Sizer, slides a rectangular window along a binarized and straightened image of root tips to estimate the maximal RH length in a given measuring interval. This measure is not affected by heavily bent RHs and any bald spots. RH length data along the root are then modelled with a sigmoidal curve, generating several biologically significant parameters such as RH length, positioning of the root differentiation zone and, under certain conditions, RH growth rate. Image analysis with Root Hair Sizer and subsequent sigmoidal modelling of RH length data provide a simple and efficient way to characterize RH growth in different conditions, equally suitable to small and large scale phenotyping experiments.

6.4.5. Calcium and plasma membrane force-gated ion channels behind development

Participant: Jean-Marc Allain.

During development, tissues are submitted to high variation of compression and tension forces. The roles of the cell wall, the cytoskeleton, the turgor pressure and the cell geometry during this process have received due attention. In contrast, apart from its role in the establishment of turgor pressure, the involvement of the plasma membrane as a transducer of mechanical forces during development has been under studied. Force-gated (FG) or Mechanosensitive (MS) ion channels embedded in the bilayer represent 'per se' archetypal mechanosensor able to directly and instantaneously transduce membrane forces into electrical and calcium signals. We reviewed in [26] how their fine-tuning, combined with their ability to detect micro-curvature and local membrane tension, allows FG channels to transduce mechanical cues into developmental signals.

6.5. Clinical Applications

6.5.1. Cardiac displacement tracking with data assimilation combining a biomechanical model and an automatic contour detection

Participants: Radomir Chabiniok, Gautier Bureau, Dominique Chapelle, Philippe Moireau [correspondant].

Data assimilation in computational models represents an essential step in building patient-specific simulations. This work aims at circumventing one major bottleneck in the practical use of data assimilation strategies in cardiac applications, namely, the difficulty of formulating and effectively computing adequate data-fitting term for cardiac imaging such as cine MRI. We here provide a proof-of-concept study of data assimilation based on automatic contour detection. The tissue motion simulated by the data assimilation framework is then assessed with displacements extracted from tagged MRI in six subjects, and the results illustrate the performance of the proposed method, including for circumferential displacements, which are not well extracted from cine MRI alone. This work was presented at the Functional Imaging and Modeling of Heart Conference (FIMH2019, Bordeaux, France) and published in [36].

6.5.2. *Minimally-invasive estimation of patient-specific end-systolic elastance using a biomechanical heart model*

Participants: Arthur Le Gall, Fabrice Vallée, Dominique Chapelle, Radomir Chabiniok [correspondant].

The end-systolic elastance (E_{es}) – the slope of the end-systolic pressure-volume relationship (ESPVR) at the end of ejection phase – has become a reliable indicator of myocardial functional state. The estimation of E_{es} by the original multiple-beat method is invasive, which limits its routine usage. By contrast, non-invasive single-beat estimation methods, based on the assumption of the linearity of ESPVR and the uniqueness of the normalised time-varying elastance curve $E^N(t)$ across subjects and physiology states, have been applied in a number of clinical studies. It is however known that these two assumptions have a limited validity, as ESPVR can be approximated by a linear function only locally, and $E^N(t)$ obtained from a multi-subject experiment includes a confidence interval around the mean function. Using datasets of 3 patients undergoing general anaesthesia (each containing aortic flow and pressure measurements at baseline and after introducing a vasopressor noradrenaline), we first study the sensitivity of two single-beat methods — by Sensaki et al. and by Chen et al. — to the uncertainty of $E^N(t)$. Then, we propose a minimally-invasive method based on a patient-specific biophysical modelling to estimate the whole time-varying elastance curve $E^{model}(t)$. We compare E_{es}^{model} with the two single-beat estimation methods, and the normalised varying elastance curve $E^{N,model}(t)$ with $E^N(t)$ from published physiological experiments. This work was presented at the Functional Imaging and Modeling of Heart conference (FIMH2019, Bordeaux, France) and published in [38].

6.5.3. *Model-based indices of early-stage cardiovascular failure and its therapeutic management in Fontan patients*

Participant: Radomir Chabiniok.

Investigating the causes of failure of Fontan circulation in individual patients remains challenging despite detailed combined invasive cardiac catheterisation and magnetic resonance (XMR) exams at rest and during stress. In this work, we use a biomechanical model of the heart and Fontan circulation with the components of systemic and pulmonary beds to augment the diagnostic assessment of the patients undergoing the XMR stress exam. We apply our model in 3 Fontan patients and one biventricular “control" case. In all subjects, we obtained important biophysical factors of cardiovascular physiology – contractility, contractile reserve and changes in systemic and pulmonary vascular resistance – which contribute to explaining the mechanism of failure in individual patients. Finally, we used the patient-specific model of one Fontan patient to investigate the impact of changes in pulmonary vascular resistance, aiming at in silico testing of pulmonary vasodilation treatments. This work (in collaboration with Bram Ruijsink and Kuberan Pushparajah from St Thomas Hospital, King’s College London) was presented at the Functional Imaging and Modeling of Heart conference (FIMH2019, Bordeaux, France) and published in [40]. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.4. *Dobutamine stress testing in patients with Fontan circulation augmented by biomechanical modeling*

Participants: Philippe Moireau, Dominique Chapelle, Radomir Chabiniok [correspondant].

Understanding (patho)physiological phenomena and mechanisms of failure in patients with Fontan circulation — a surgically established circulation for patients born with a functionally single ventricle — remains challenging due to the complex hemodynamics and high inter-patient variations in anatomy and function. In this work, we present a biomechanical model of the heart and circulation to augment the diagnostic evaluation of Fontan patients with early-stage heart failure. The proposed framework employs a reduced-order model of heart coupled with a simplified circulation including venous return, creating a closed-loop system. We deploy this framework to augment the information from data obtained during combined cardiac catheterization and magnetic resonance exams (XMR), performed at rest and during dobutamine stress in 9 children with Fontan circulation and 2 biventricular controls. We demonstrate that our modeling framework enables patient-specific investigation of myocardial stiffness, contractility at rest, contractile reserve during stress and changes in vascular resistance. Hereby, the model allows to identify key factors underlying the pathophysiological response to stress in these patients. In addition, the rapid personalization of the model to patient data and fast simulation of cardiac cycles makes our framework directly applicable in a clinical workflow. We conclude that the proposed modeling framework is a valuable addition to the current clinical diagnostic XMR exam that helps to explain patient-specific stress hemodynamics and can identify potential mechanisms of failure in patients with Fontan circulation. This work has been submitted for publication in an international journal. This work (in collaboration with Bram Ruijsink and Kuberan Pushparajah from St Thomas Hospital, King's College London and Tarique Hussain, UT Southwestern Medical Center Dallas) also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.5. Signed-distance function based non-rigid registration of image sequences with varying image intensity

Participant: Radomir Chabiniok.

In this work we deal with non-rigid registration of the image series acquired by the Modified Look-Locker Inversion Recovery (MOLLI) magnetic resonance imaging sequence, which is used for a pixel-wise estimation of T_1 relaxation time. The spatial registration of the images within the series is necessary to compensate the patient's imperfect breath-holding. The evolution of intensities and a large variation of the image contrast within the MOLLI image series, together with the myocardium of left ventricle (the object of interest) typically not being the most distinct object in the scene, makes the registration challenging. We propose a method for locally adjusted optical flow-based registration of multimodal images, which uses the segmentation of the object of interest and its representation by the signed-distance function. We describe all the components of the proposed OF^{dist} method and their implementation. The OF^{dist} method is then compared to the performance of a standard mutual information maximization-based registration method, applied either to the original image (MIM) or to the signed-distance function (MIM^{dist}). Several experiments with synthetic and real MOLLI images are carried out. On synthetic image with a single object, MIM performed the best, while OF^{dist} and MIM^{dist} provided better results on synthetic images with more than one object and on real images. When applied to signed-distance function of two objects of interest, MIM^{dist} provided a larger registration error (but more homogeneously distributed) compared to OF^{dist}. For the real MOLLI image sequence with left ventricle pre-segmented using level-set method, the proposed OF^{dist} registration performed the best, as is demonstrated visually and by measuring the increase of mutual information in the object of interest and its neighborhood. This collaborative work (Katerina Skardova, Czech Technical University, Institute of Clinical and Experimental Medicine in Prague) has been submitted for publication in an international journal. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.6. Estimation of left ventricular pressure-volume loop using hemodynamic monitoring augmented by a patient-specific biomechanical model. An observational study

Participants: Arthur Le Gall, Fabrice Vallée, Dominique Chapelle, Radomir Chabiniok [correspondant].

Background During general anaesthesia, direct analysis of the arterial pressure or aortic flow waveforms may be confusing in complex haemodynamic situations. Patient-specific biomechanical modelling allows to simulate Pressure-Volume (PV) loops and obtain functional indicators of the cardiovascular (CV) system, such as ventricular-arterial coupling (Vva), cardiac efficiency (CE) or myocardial contractility. It therefore augments

the information obtained by monitoring and could help in medical decision-making. **Methods** Patients undergoing GA for neuroradiological procedure were included in this prospective observational study. A biomechanical model of heart and vasculature specific to each patient was built using transthoracic echocardiography and aortic pressure and flow signals. If intraoperative hypotension (IOH) appeared, diluted noradrenaline (NOR) was administered and the model readjusted. **Results** The model was calibrated for 29 (64%) normotensive and for 16 (36%) hypotensive patients before and after NOR administration. The simulated mean aortic pressure (MAP) and stroke volume (SV) were equivalent to the measurements (Percentage Error: 6% for MAP and 18% for SV) in all 45 datasets at baseline. After NOR administration, the percentage of concordance with 10% exclusion zone between measurement and simulation was > 95% for both MAP and SV. The modelling results showed a decreased Vva (0.64 ± 0.37 vs 0.88 ± 0.43; p=0.039), and an increased CE (0.8 ± 0.1 vs 0.73 ± 0.11; p=0.042) in hypotensive as compared with normotensive patients. After NOR administration, Vva increased by 92 ± 101%, CE decreased by 13 ± 11% (p<0.001 for both) and contractility increased by 14 ± 11% (p=0.002). **Conclusions** The numerical models built for individual patients were applied to estimate patients' PV loops and functional indicators of CV system during haemodynamic alterations and during restoration by NOR. This study demonstrates the feasibility of patient-specific cardiovascular modelling using clinical data readily available during GA and paves the way for model-augmented haemodynamic monitoring at operating theatres and intensive care units. This work is about to be submitted for publication in an international journal. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.7. Investigation of phase contrast magnetic resonance imaging underestimation of turbulent flow through the aortic valve phantom: Experimental and computational study by using lattice Boltzmann method

Participant: Radomir Chabiniok.

Work in collaboration with Radek Fucik, Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague.

Objective The accuracy of phase-contrast magnetic resonance imaging (PC-MRI) measurement is investigated using a computational fluid dynamics (CFD) model with the objective to determine the magnitude of the flow underestimation due to turbulence behind a narrowed valve in a phantom experiment. **Materials and Methods** An acrylic stationary flow phantom is used with three insertable plates mimicking aortic valvular stenoses of varying degrees. Positive and negative horizontal fluxes are measured at equidistant slices using standard PC-MRI sequences by 1.5T and 3T systems. The CFD model is based on the 3D lattice Boltzmann method (LBM). The experimental and simulated data are compared using the Bland-Altman-derived limits of agreement. Based on the LBM results, the turbulence is quantified and confronted with the level of flow underestimation. **Results** Matching results of PC-MRI flow were obtained for valves up to moderate stenosis on both field strengths. The flow magnitude through a severely stenotic valve was underestimated due to signal void in the regions of turbulent flow behind the valve, consistently with the level of quantified turbulence intensity. **Discussion** Flow measured by PC-MRI is affected by noise and turbulence. LBM can simulate turbulent flow efficiently and accurately, it has therefore the potential to improve clinical interpretation of PC-MRI. This collaborative work (Czech Technical University, Institute of Clinical and Experimental Medicine in Prague and Inria) has been submitted for publication in an international journal. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.8. Left ventricular torsion obtained using equilibrated warping in patients with repaired Tetralogy of Fallot

Participants: Martin Genet, Radomir Chabiniok [correspondant].

Work in collaboration with Katerina Skardova, Department of Mathematics, Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague and Tarique Hussain UT Southwestern Medical Center Dallas.

Background Patients after surgical repair of Tetralogy of Fallot (rTOF) have right ventricular (RV) dysfunction and may subsequently suffer a decrease in left ventricular (LV) function. Previous studies evaluating the assessment of LV torsion have shown poor reproducibility using cardiovascular magnetic resonance imaging (CMR). The aim of our study is to evaluate a novel finite element method of image registration to assess LV torsion in patients with rTOF and explore the relationship between LV torsion and cardiac parameters routinely obtained with CMR. **Methods** The assessment of torsion is based on the finite element method for image registration, and the equilibrium gap principle for problem regularization, known as equilibrated warping developed by M. Genet (Inria Saclay). It has been shown to be able to predict global torsion in regular cine images as well in 3D tagged images, despite low contrast. Seventy-six cases of rTOF and ten controls were included. The group of control patients were assessed for reproducibility using equilibrated warping and standard tissue tracking software (cvi42, version 5.10.1, Calgary, Canada). RV end-systolic volume (RVESV), RV end-diastolic volume (RVEDV), RV ejection fraction (RVEF), LVESV, LVEDV, LVEF, LV peak systolic torsion and peak systolic torsion gradient (normalized by mesh length) were obtained for each patient with rTOF. Patients were dichotomized into two groups: those with normal torsion (systolic basal clockwise rotation and apical counterclockwise rotation, representative example is shown in Image 1) and those with loss of torsion, defined as a reversal of normal systolic basal clockwise rotation (representative example is shown in Image 2). **Results** Torsion by equilibrated warping was successfully obtained in 68 of 76 (89%) patients with rTOF and 9 of 10 (90%) normal controls. For equilibrated warping, the intra- and inter-observer coefficients of variation were 0.095 and 0.117, respectively; compared to 0.668 and 0.418 for tissue tracking by standard clinical software. The intra- and inter-observer intraclass correlation coefficients for equilibrated warping were 0.862 and 0.831, respectively; compared to 0.250 and 0.621 for tissue tracking. Loss of torsion was noted in 32 of the 68 (47%) patients with rTOF and there was a significant difference in peak systolic torsion gradient between patients with normal torsion and loss of torsion. There was no difference in LV or RV volumes or function between these groups. **Conclusion** The equilibrated warping method of image registration to assess LV torsion is feasible in patients with rTOF and shows good reliability. Loss of torsion is common in patients with rTOF. In our study, there was no significant association between loss of torsion and other ventricular parameters indicative of a worsening cardiac condition. Future studies committed to the long- term follow-up of this population are needed to assess the role of torsion in predicting ventricular dysfunction and death. This work was accepted for presentation at SCMR conference 2020 (Society for Cardiovascular Magnetic Resonance). It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.9. Volume administration protocol to assess ventricular mechanics during interventional cardiac magnetic resonance procedures

Participant: Radomir Chabiniok.

Work in collaboration with Joshua Greer and Tarique Hussain UT Southwestern Medical Center Dallas.

Background Failure in Fontan circulation occurs with supposed normal ventricular systolic and diastolic function, including normal ventricular end-diastolic pressures and ventricular ejection fraction. This highlights the difficulty in assessing systolic and diastolic ventricular function in patients with single ventricle physiology. Interventional cardiac magnetic resonance (CMR) provides an opportunity for simultaneous acquisition of pressure and volume measurements that may lend itself well to analysis of ventricular mechanics in this population. We aim to develop a protocol of volume administration to assess ventricular pressure and volume during the cardiac cycle to construct pressure-volume loops under different loading conditions and perform their biomechanical interpretation. **Methods** This is a single center prospective study conducted on single ventricle patients with Glenn or Fontan circulation referred for interventional CMR procedures. With a catheter advanced into the ventricle, a pressure tracing and a cine sequence accelerated by kt-BLAST is obtained. Two 2.5 mL/kg fluid boluses are then rapidly administered into the catheter sheath with repeated acquisition of the pressure tracing and cine imaging immediately following each. Cine images are post-processed after the procedure to obtain ventricular volumes. The data are combined to construct pressure-volume loops and plot the end-diastolic pressure-volume relationship (EDPVR). **Results** The protocol has been performed in six patients. Ventricular end-diastolic pressure readings increased by a median of 2.5 mmHg (range of 1-3 mmHg) after the first volume administration and a median of 1.5 mmHg (range of 1-8 mmHg) after the

second volume administration. Ventricular end-diastolic volumes increased by a median of 4.1 mL (range of 1.7-19.3 mL) after the first volume administration and a median of 1.5 mL (range of 0.4-24.2 mL) after the second volume administration. The data obtained during simultaneous volume and pressure measurements allowed for the construction of ventricular pressure-volume loops. Ventricular stroke work increased by a median of 0.0825 Joules (range of 0.010-0.167 Joules) after the first volume administration then decreased by a median of -0.062 Joules (range of -0.083 to 0.005 Joules) after the second volume administration. EDPVR curves were derived from the pressure-volume loops and differentiated patients with similar starting end-diastolic pressures. **Conclusions** We present a novel method for the acquisition of data to construct pressure-volume loops. Our protocol focuses on rapid volume administration and fast data acquisition with the goal of increasing preload but recording data prior to compensatory changes in afterload. In each patient, administration of 2.5 mL/kg fluid boluses achieved measurable increases in ventricular end-diastolic pressure and ventricular end-diastolic volume. The construction of pressure-volume loops with varying loading may facilitate in-depth assessment of ventricular mechanics in patients with single ventricle heart disease. The variation of preload may allow for the assessment of EDPVR, therefore ventricular stiffness, and to some extent also the contractile response in such a physiology-modifying situation. This work was submitted to the CHOP 2020 conference. It also represents a part of the objectives supported by the Inria-UTSW Associated Team TOFMOD.

6.5.10. Computational quantification of patient specific changes in ventricular dynamics associated with pulmonary hypertension

Participant: Martin Genet.

Pulmonary arterial hypertension (PAH) causes an increase in the mechanical loading imposed on the right ventricle (RV) that results in progressive changes to its mechanics and function. Here, we quantify the mechanical changes associated with PAH by assimilating clinical data consisting of reconstructed three-dimensional geometry, pressure, and volume waveforms, as well as regional strains measured in patients with PAH ($n = 12$) and controls ($n = 6$) within a computational modeling framework of the ventricles. Modeling parameters reflecting regional passive stiffness and load-independent contractility as indexed by the tissue active tension were optimized so that simulation results matched the measurements. The optimized parameters were compared with clinical metrics to find usable indicators associated with the underlying mechanical changes. Peak contractility of the RV free wall (RVFW) $\gamma_{RVFW,max}$ was found to be strongly correlated and had an inverse relationship with the RV and left ventricle (LV) end-diastolic volume ratio (i.e., RVEDV/LVEDV) (RVEDV/LVEDV)+0.44, $R^2 = 0.77$). Correlation with RV ejection fraction ($R^2 = 0.50$) and end-diastolic volume index ($R^2 = 0.40$) were comparatively weaker. Patients with $RVEDV/LVEDV > 1.5$ had 25% lower $\gamma_{RVFW,max}$ ($P < 0.05$) than that of the control. On average, RVFW passive stiffness progressively increased with the degree of remodeling as indexed by RVEDV/LVEDV. These results suggest a mechanical basis of using RVEDV/LVEDV as a clinical index for delineating disease severity and estimating RVFW contractility in patients with PAH. See [25] for more detail.

6.5.11. Validation of equilibrated warping-image registration with mechanical regularization-on 3D ultrasound images

Participant: Martin Genet.

Image registration plays a very important role in quantifying cardiac motion from medical images, which has significant implications in the diagnosis of cardiac diseases and the development of personalized cardiac computational models. Many approaches have been proposed to solve the image registration problem; however, due to the intrinsic ill-posedness of the image registration problem, all these registration techniques, regardless of their variabilities, require some sort of regularization. An efficient regularization approach was recently proposed based on the equilibrium gap principle, named equilibrated warping. Compared to previous work, it has been formulated at the continuous level within the finite strain hyperelasticity framework and solved using the finite element method. Regularizing the image registration problem using this principle is advantageous as it produces a realistic solution that is close to that of an hyperelastic body in equilibrium with arbitrary boundary tractions, but no body load.The equilibrated warping method has already been extensively

validated on both tagged and untagged magnetic resonance images. In this paper, we provide full validation of the method on 3D ultrasound images, based on the 2011 MICCAI Motion Tracking Challenge data. See [39] for more detail.

MAMBA Project-Team

7. New Results

7.1. Direct and inverse Problems in Structured-population equations

7.1.1. Modelling Polymerization Processes

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]; new experimental evidence reinforced these findings [19], [35]. The challenges are now to propose mathematical models capable of tracking such diversity while keeping sufficient simplicity to be tractable to analysis.

In collaboration with Klemens Fellner from the university of Graz, we propose a new model, variant of the Becker-Döring system but containing two monomeric species, capable of displaying sustained though damped oscillations as is experimentally observed. We also proposed a statistical test to validate or invalidate the presence of oscillations in experimental highly nonstationary signals [55].

7.1.2. Asymptotic behaviour of structured-population equations

Pierre Gabriel and Hugo Martin studied the mathematical properties of a model of cell division structured by two variables – the size and the size increment – in the case of a linear growth rate and a self-similar fragmentation kernel [16]. They first show that one can construct a solution to the related two dimensional eigenproblem associated to the eigenvalue 1 from a solution of a certain one dimensional fixed point problem. Then they prove the existence and uniqueness of this fixed point in the appropriate L^1 weighted space under general hypotheses on the division rate. Knowing such an eigenfunction proves useful as a first step in studying the long time asymptotic behaviour of the Cauchy problem.
Etienne Bernard, Marie Doumic and Pierre Gabriel proved in [9] that for the growth-fragmentation equation with fission into two equal parts and linear growth rate, under fairly general assumptions on the division rate, the solution converges towards an oscillatory function, explicitely given by the projection of the initial state on the space generated by the countable set of the dominant eigenvectors of the operator. Despite the lack of hypo-coercivity of the operator, the proof relies on a general relative entropy argument in a convenient weighted L^2 space, where well-posedness is obtained via semigroup analysis. They also propose a non-dissipative numerical scheme, able to capture the oscillations.

Pierre Gabriel and Hugo Martin then extended this asymptotic result in the framework of measure solutions [50]. To do so they adopt a duality approach, which is also well suited for proving the well-posedness when the division rate is unbounded. The main difficulty for characterizing the asymptotic behavior is to define the projection onto the subspace of periodic (rescaled) solutions. They achieve this by using the generalized relative entropy structure of the dual problem.

7.1.3. Estimating the division rate from indirect measurements of single cells

7.1.3.1. Marie Doumic and Adélaïde Olivier

Is it possible to estimate the dependence of a growing and dividing population on a given trait in the case where this trait is not directly accessible by experimental measurements, but making use of measurements of another variable? The article [46] adresses this general question for a very recent and popular model describing bacterial growth, the so-called incremental or adder model - the model studied by Hugo Martin and Pierre Gabriel in [16]. In this model, the division rate depends on the increment of size between birth and division, whereas the most accessible trait is the size itself. We prove that estimating the division rate from size measurements is possible, we state a reconstruction formula in a deterministic and then in a statistical setting, and solve numerically the problem on simulated and experimental data. Though this represents a severely ill-posed inverse problem, our numerical results prove to be satisfactory.

7.2. Stochastic Models of Biological Systems

7.2.1. Stochastic models for spike-timing dependent plasticity

7.2.1.1. Ph. Robert and G. Vignoud

Synaptic plasticity is a common mechanism used to model learning in stochastic neural networks, STDP is a great example of such mechanisms. We develop a simple framework composed by two neurons and one synaptic weight, seen as stochastic processes and study the existence and stability of such distributions, for a wide range of classical synaptic plasticity models. Using two simple examples of STDP, the calcium-based rule and the all-to-all pair-based rule, we apply stochastic averaging principles and obtain differential equations for the limit processes, based on the invariant distributions of the fast system when the slow variables are considered fixed. We study a general stochastic queue to approximate the calcium-based rule and are able to have an analytical solution for the invariant distribution of the fast synaptic processes. We also detail some simpler systems, either through some approximations or simulations to put into light the influences of different biologically-linked parameters on the dynamics of the synaptic weight.

7.2.2. Online Sequence Learning In The Striatum With Anti-Hebbian Spike-Timing-Dependent Plasticity

7.2.2.1. G. Vignoud. Collaboration with J. Touboul (Brandeis University)

Spike-Timing Dependent Plasticity (STDP) in the striatum is viewed as a substrate for procedural learning. Striatal projecting neurons (SPNs) express anti-Hebbian plasticity at corticostriatal synapses, (a presynaptic cortical spike followed by a postsynaptic striatal spike leads to the weakening of the connection, whereas the reverse pairing leads to potentiation). SPNs need to integrate many inputs to spike, and as such, their main role is to integrate context elements to choose between different sensorimotor associations. In this work, we develop a simple numerical model of the striatum, integrating cortical spiking inputs to study the role of anti-Hebbian STDP in pattern recognition and sequence learning. Cortical neurons are seen as binary input neurons and one striatal SPN is modeled as a leaky-integrate-and-fire neuron. Combined informations from the output, reward and timing between the different spikes modify the intensity of each connection, through two mechanisms: anti-Hebbian STDP and dopaminergic signaling, using three-factor learning rules. We have added a second output neuron with collateral inhibition which leads to an improvement of the global accuracy. In another project, we studied the dynamics of learning, by shutting off/on the dopaminergic plasticity, and compare it to DMS/DLS experimental and behavioral experiments. We show that anti-Hebbian STDP favors the learning of complete sequence of spikes, such as is needed in the striatum, whereas, even if Hebbian STDP helps to correlate the spiking of two connected neurons, it is not sufficient to integrate of long sequences of correlated inputs spikes.

7.2.3. D1/D2 detection from action-potential properties using machine learning approach in the dorsal striatum

7.2.3.1. G. Vignoud. Collaboration, with Team Venance (CIRB/Collège de France)

Striatal medium spiny neurons (MSNs) are segregated into two subpopulations, the D1 receptor-expressing MSNs (the direct striatonigral pathway) and the D2 receptorexpressing MSNs (the indirect striatopallidal pathway). The fundamental role of MSNs as output neurons of the striatum, and the necessary distinction between D1- and D2-expressing neurons accentuate the need to clearly distinguish both subpopulations in electrophysiological recordings in vitro and in vivo. Currently, fluorescent labelling of the dopaminergic receptors in mice enables a clear differentiation. However, multiplying in vivo the number of genetic markers (optogenetics, fluorescence) hinders possibilities for other genetic manipulations. Moreover, electrophysiological properties of fluorescents neurons can slightly differ from “native” cells and false-positive can be observed. The lack of a proper way to separate D1- and D2-MSNs based on electrophysiological properties led us to devise a detection algorithm based on action potential profile. We used more than 450 D1/D2 labelled MSNs from in vitro patch-clamp recordings (different experimentalists, different setups and protocols), to characterize and identify properties that facilitate the MSN discrimination. After analyzing passive and active MSN membrane

properties, we built an extensive dataset and fed it into classical machine learning classification methods. The training of the different algorithms (k-nearest neighbors, random forest, deep neural networks, ...) was performed with the scikit-learn Python library, and the optimized classifier was able to correctly discriminate neurons in the dorsolateral striatum at 76% (and up to 83% if we allow the classifier to reject some MSNs). This study developed an efficient classification algorithm for D1/D2-MSNs, facilitating cell discrimination without specific genetic fluorescent labelling, leaving some room for other genetic markers and optogenetic labeling.

7.2.4. The Stability of Non-Linear Hawkes Processes

7.2.4.1. Ph. Robert and G. Vignoud

We have investigated the asymptotic properties of self-interacting point processes introduced by Kerstan (1964) and Hawkes and Oakes (1974). These point processes have the property that the intensity at some point $t \in (-\infty, +\infty)$ is a functional of all points of the point process before t. Such a process is said to be stable if it has a version whose distribution is invariant by translation. By using techniques of coupling and Markovian methods, we have been able to obtain some existence and uniqueness results with weaker conditions than in the current literature.

7.2.5. Mathematical Models of Gene Expression

7.2.5.1. Ph. Robert

In Robert [30] we analyze the equilibrium properties of a large class of stochastic processes describing the fundamental biological process within bacterial cells, *the production process of proteins*. Stochastic models classically used in this context to describe the time evolution of the numbers of mRNAs and proteins are presented and discussed. An extension of these models, which includes elongation phases of mRNAs and proteins, is introduced. A convergence result to equilibrium for the process associated to the number of proteins and mRNAs is proved and a representation of this equilibrium as a functional of a Poisson process in an extended state space is obtained. Explicit expressions for the first two moments of the number of mRNAs and proteins at equilibrium are derived, generalizing some classical formulas. Approximations used in the biological literature for the equilibrium distribution of the number of proteins are discussed and investigated in the light of these results. Several convergence results for the distribution of the number of proteins at equilibrium are in particular obtained under different scaling assumptions.

7.2.6. Stochastic modelling of molecular motors

7.2.6.1. Marie Doumic, Dietmar Oelz, Alex Mogilner

It is often assumed in biophysical studies that when multiple identical molecular motors interact with two parallel microtubules, the microtubules will be crosslinked and locked together. The aim of the article [4] is to examine this assumption mathematically. We model the forces and movements generated by motors with a time-continuous Markov process and find that, counter-intuitively, a tug-of-war results from opposing actions of identical motors bound to different microtubules. The model shows that many motors bound to the same microtubule generate a great force applied to a smaller number of motors bound to another microtubule, which increases detachment rate for the motors in minority, stabilizing the directional sliding. However, stochastic effects cause occasional changes of the sliding direction, which has a profound effect on the character of the long-term microtubule motility, making it effectively diffusion-like. Here, we estimate the time between the rare events of switching direction and use them to estimate the effective diffusion coefficient for the microtubule pair. Our main result is that parallel microtubules interacting with multiple identical motors are not locked together, but rather slide bidirectionally. We find explicit formulae for the time between directional switching for various motor numbers.

7.3. Analysis and control of mosquitoe populations

7.3.1. 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 [152]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 [152].

7.3.2. 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 [8], [42].

7.3.3. 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 [33] 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 [153], [154] 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.

7.3.4. 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 [39]. 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].

7.4. Bacterial motion by Rerun 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 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 [142]. 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 [86]. 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.

Another activity concerns the relation between the tumbling rate and the internal state of bacteria. The study [58] aims at deriving models at the macroscopic scale from assumptions on the microscopic scales. In particular we are interested in comparisons between the stiffness of the response and the adaptation time. Depending on the asymptotics chosen both the standard Keller-Segel equation and the flux-limited Keller-Segel (FLKS) equation can appear. An interesting mathematical issue arises with a new type of equilibrium equation leading to solution with singularities.

7.5. 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 [82], 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 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 [131]. 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.

7.6. 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) [93], [94]. 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 [24].

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 [128], [127] and later continued in [91], [92], [126], has been recently summarised in [60] 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.

7.7. 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 [123]. 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 [155]. 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. Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer

Authors: *C. Nicolò; C. Périer*; M. Prague; C. Bellera; G. MacGrogan; *O.Saut; S. Benzekry*. Accepted for publication in the Journal of Clinical Oncology: Clinical Cancer Informatics.

Purpose: For patients with early-stage breast cancer, prediction of the risk of metastatic relapse is of crucial importance. Existing predictive models rely on agnostic survival analysis statistical tools (e.g. Cox regression). Here we define and evaluate the predictive ability of a mechanistic model for the time to metastatic relapse.

Methods: The data consisted of 642 patients with 21 clinicopathological variables. A mechanistic model was developed on the basis of two intrinsic mechanisms of metastatic progression: growth (parameter α) and dissemination (parameter μ). Population statistical distributions of the parameters were inferred using mixed-effects modeling. A random survival forest analysis was used to select a minimal set of 5 covariates with best predictive power. These were further considered to individually predict the model parameters, by using a backward selection approach. Predictive performances were compared to classical Cox regression and machine learning algorithms.

Results: The mechanistic model was able to accurately fit the data. Covariate analysis revealed statistically significant association of Ki67 expression with α (p=0.001) and EGFR with μ (p=0.009). Achieving a c-index of 0.65 (0.60-0.71), the model had similar predictive performance as the random survival forest (c-index 0.66-0.69) and Cox regression (c-index 0.62 - 0.67), as well as machine learning classification algorithms.

Conclusion: By providing informative estimates of the invisible metastatic burden at the time of diagnosis and forward simulations of metastatic growth, the proposed model could be used as a personalized prediction tool of help for routine management of breast cancer patients.

7.2. Numerical workflow for clinical electroporation ablation

Authors: *Olivier Gallinato*, *Baudouin Denis de Senneville*, Olivier Seror, *Clair Poignard*. Published in Physics in Medicine and Biology. https://hal.inria.fr/hal-02063020/file/paperIRE_workflow_R3.pdf.

The paper describes a numerical workflow, based on the “real-life” clinical workflow ofirreversible electroporation (IRE) performed for the treatment of deep-seated liver tumors. Thanksto a combination of numerical modeling, image registration algorithm and clinical data, our numerical workflow enables to provide the distribution of the electric field as effectively delivered by theclinical IRE procedure. As a proof of concept, we show on a specific clinical case of IRE ablation ofliver tumor that clinical data could be advantageously combined to numerical simulations in a nearfuture , in order to give to the interventional radiologists information on the effective IRE ablation.We also corroborate the simulated treated region with the post-treatment MRI performed 3 days after treatment.

7.3. A reduced Gompertz model for predicting tumor age using a population approach

Authors: *C. Vaghi*, A. Rodallec, R. Fanciullino, J. Ciccolini, J. Mochel, M. Mastri, *C. Poignard*, J. ML Ebos, *S. Benzekry*. Accepted for publication in PLoS Computational Biology. https://www.biorxiv.org/content/10.1101/670869v2

Tumor growth curves are classically modeled by means of ordinary differential equations. In analyzing the Gompertz model several studies have reported a striking correlation between the two parameters of the model, which could be used to reduce the dimensionality and improve predictive power.

We analyzed tumor growth kinetics within the statistical framework of nonlinear mixed-effects (population approach). This allowed the simultaneous modeling of tumor dynamics and inter-animal variability. Experimental data comprised three animal models of breast and lung cancers, with 833 measurements in 94 animals. Candidate models of tumor growth included the exponential, logistic and Gompertz. The exponential and – more notably – logistic models failed to describe the experimental data whereas the Gompertz model generated very good fits. The previously reported population-level correlation between the Gompertz parameters was further confirmed in our analysis ($R^2 > 0.92$ in all groups). Combining this structural correlation with rigorous population parameter estimation, we propose a reduced Gompertz function consisting of a single individual parameter (and one population parameter). Leveraging the population approach using Bayesian inference, we estimated times of tumor initiation using three late measurement timepoints. The reduced Gompertz model was found to exhibit the best results, with drastic improvements when using Bayesian inference as compared to likelihood maximization alone, for both accuracy and precision. Specifically, mean accuracy was 12.2% versus 78% and mean precision was 15.6 days versus 210 days, for the breast cancer cell line.

These results offer promising clinical perspectives for the personalized prediction of tumor age from limited data at diagnosis. In turn, such predictions could be helpful for assessing the extent of invisible metastasis at the time of diagnosis.

The code and the data used in our analysis are available at https://github.com/cristinavaghi/plumky.

7.4. T2-based MRI Delta-Radiomics Improve Response Prediction in Soft-Tissue Sarcomas Treated by Neoadjuvant Chemotherapy

Authors: *Amandine Crombé*, *Cynthia Perier*, Michèle Kind, *Baudouin Denis de Senneville*, Francois Le Loarer, Antoine Italiano, Xavier Buy, *Olivier Saut*. Published in Journal of Magnetic Reasonance Imaging https://hal.inria.fr/hal-01929807v2

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 remains 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: 65 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 2 cycles. Field strength/Sequence: Pre- and post-contrast enhanced T1-weighted imaging (T1-WI), turbo spin echo T2-WI at 1.5T.

Assessment: A threshold of <10% viable cells on surgical specimen 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 voxelsizes 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 cross-validation (training and validation) on 50 patients ('training cohort') and was validated on 15 other patients ('test cohort').

Results: 16 patients were good-HR. Neither RECIST status, nor semantic radiological variables were associated with response 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 3 features, which provided: AUROC=0.86, accuracy=88.1%, sensitivity=94.1%, 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 conclusions: A T2-based Delta-Radiomics approach can improve early response prediction in STS patients with a limited number of features.

7.5. Quantitative mathematical modeling of clinical brain metastasis dynamics in non-small cell lung cancer

Authors: *M. Bilous*, C. Serdjebi, A. Boyer, P. Tomasini, *C. Pouypoudat*, D. Barbolosi, F. Barlesi, F. Chomy, *S. Benzekry*. Published in Scientific Reports. https://hal.inria.fr/hal-01928442v2

Brain metastases (BMs) are associated with poor prognosis in non-small cell lung cancer (NSCLC), but are only visible when large enough. Therapeutic decisions such as whole brain radiation therapy would benefit from patient-specific predictions of radiologically undetectable BMs. Here, we propose a mathematical modeling approach and use it to analyze clinical data of BM from NSCLC. Primary tumor growth was best described by a gompertzian model for the pre-diagnosis history, followed by a tumor growth inhibition model during treatment. Growth parameters were estimated only from the size at diagnosis and histology, but predicted plausible individual estimates of the tumor age (2.1–5.3 years). Multiple metastatic models were further assessed from fitting either literature data of BM probability (n = 183 patients) or longitudinal measurements of visible BMs in two patients. Among the tested models, the one featuring dormancy was best able to describe the data. It predicted latency phases of 4.4–5.7 months and onset of BMs 14–19 months before diagnosis. This quantitative model paves the way for a computational tool of potential help during therapeutic management.

7.6. Optimizing 4D abdominal MRI: Image denoisingusing an iterative back-projection approach

Authors: *B Denis de Senneville*, C R Cardiet, A J Trotier, E J Ribot, L Lafitte, L Facq, S Miraux. Published in Physics in Medicine and Biology. https://hal.archives-ouvertes.fr/hal-02367839/file/2019_4D_reconstruction_revision2_final.pdf

4D-MRI is a promising tool for organ exploration, target delineationand treatmentplanning. Intra-scan motion artifacts may be greatly reduced by increasing the imagingframe rate. However, poor signal-to-noise ratios (SNR) are observed when increasingspatial and/or frame number per physiological cycle, in particular inthe abdomen.In the current work, the proposed 4D-MRI method favored spatial resolution, framenumber, isotropic voxels and large field-of-view (FOV) during MR-acquisition. Theconsequential SNR penalty in the reconstructed data is addressed retrospectively usingan iterative back-projection (IBP) algorithm. Practically, after computing individualspatial 3D deformations present in the images using a deformable image registration(DIR) algorithm, each 3D image is individually enhanced by fusing several successiveframes in its local temporal neighborood, these latter being likely tocover commonindependent informations. A tuning parameter allows one to freely readjust the balancebetween temporal resolution and precision of the 4D-MRI.The benefit of the method was quantitatively evaluated on the thorax of 6 miceunder free breathing using a clinically acceptable duration. Improved 4D cardiacimaging was also shown in the heart of 1 mice. Obtained results are comparedto theoretical expectations and discussed. The proposed implementation is easilyparallelizable and optimized 4D-MRI could thereby be obtained with a clinicallyacceptable duration.

7.7. Optimal Scheduling of Bevacizumab and Pemetrexed/cisplatin Dosing in Non-Small Cell Lung Cancer

Authors: Benjamin Schneider, Arnaud Boyer, Joseph Ciccolini, Fabrice Barlési, Kenneth Wang, *Sébastien Benzekry**, Jonathan Mochel*. * = co-senior authors. Published in CPT: Pharmacometrics and Systems Pharmacology. https://hal.inria.fr/hal-02109335

Bevacizumab-pemetrexed/cisplatin (BEV-PEM/CIS) is a first line therapeutic for advanced non-squamous non-small cell lung cancer (NSCLC). Bevacizumab potentiates PEM/CIS cytotoxicity by inducing transient tumor vasculature normalization. BE V- PEM/CIS has a narrow therapeutic window. Therefore, it is an attractive target for administration schedule optimization. The present study leverages our previous work on BEV-PEM/CIS pharmacodynamic modeling in NSCLC-bearing mice to estimate the optimal gap in the scheduling of sequential BEV-PEM/CIS. We predicted the optimal gap in BEV-PEM/CIS dosing to be 2.0 days

in mice and 1.2 days in humans. Our simulationssuggest that the efficacy loss in scheduling BEV-PEM/CIS at too great of a gap is much less than the efficacy loss in scheduling BEV-PEM/CIS at too short of a gap.

NUMED Project-Team (section vide)

REO Team

6. New Results

6.1. Numerical methods for fluid mechanics and application to blood flows

Participants: Irene Vignon-Clementel

If abdominal aortic aneurysms (AAA) are known to be associated with altered morphology and blood flow, intraluminal thrombus deposit and clinical symptoms, the growth mechanisms are yet to be fully understood. In this retrospective longitudinal study of 138 scans, morphological analysis and blood flow simulations for 32 patients with clinically diagnosed AAAs and several follow-up CT-scans, are performed and compared to 9 control subjects [21]. Local correlations between hemodynamic metrics and AAA growth are also explored. Finally, high-risk predictors trained with successively clinical, morphological, hemodynamic and all data, and their link to the AAA evolution are built from supervise learning.

In this paper [19], we perform a verification study of the Coupled-Momentum Method (CMM), a 3D fluid-structure interaction (FSI) model which uses a thin linear elastic membrane and linear kinematics to describe the mechanical behavior of the vessel wall. The verification of this model is done using Womersley's deformable wall analytical solution for pulsatile flow in a semi-infinite cylindrical vessel. This solution is, under certain premises, the analytical solution of the CMM and can thus be used for model verification. For the numerical solution, we employ an impedance boundary condition to define a reflection-free outflow boundary condition and thus mimic the physics of the analytical solution, which is defined on a semi-infinite domain. We first provide a rigorous derivation of Womersley's deformable wall theory via scale analysis. We then illustrate different characteristics of the analytical solution and verification tests comparing the CMM with Womersley's theory.

Superior cavopulmonary circulation can be achieved by either the Hemi-Fontan or Bidirectional Glenn connection. Debate remains as to which results in best hemodynamic results. In [22], adopting patient-specific multiscale computational modeling, we examined both the local dynamics and global physiology to determine if surgical choice can lead to different hemodynamic outcomes.

6.2. Liver biomedical research

Participants: Irene Vignon-Clementel, Nicolas Golse

Nicolas Golse, as part of his medical activity has published 7 articles in 2019 that are not reported here.

The hepatic volume gain following resection is essential for clinical recovery. Previous studies have focused on cellular regeneration. In [13], the study aims to explore the rate of hepatic regeneration of the porcine liver following major resection, highlighting estimates of the early microarchitectural changes that occur during the cellular regeneration. Nineteen large white pigs had 75% resection with serial measurements of the hepatic volume, density, blood flow, and architectural changes that are analyzed at different days to highlight differences pre-resection and in the days following resection.

SISTM Project-Team

7. New Results

7.1. Mechanistic learning

7.1.1. Ebola models

New models have been developed for the response to the Ebola vaccine. The first one has been fitted to Phase 1 trials and has given interesting predictions of the long term duration of the response that are confirmed with the new data coming from phase 2 trials. These results have been published in Journal of Virology. Then, a new model including the B cell memory response has been defined and its mathematical proprieties have been studied. A manuscript has been submitted to Journal of Theoretical Biology. The next step is to estimate model parameters using EBL2001 clinical trial data.

New publication: Pasin C et al. Dynamics of the Humoral Immune Response to a Prime-Boost Ebola Vaccine: Quantification and Sources of Variation. J Virol. 2019 Aug 28;93(18). pii: e00579-19. doi: 10.1128/JVI.00579-19. Print 2019 Sep 15.

7.1.2. Estimation method

A new approach is currently under development by Quentin Clairon to estimate model parameters using a regularization method based on the control theory. The estimation method used an approximation of the original ODE solution for each subject. The expected advantages of this approach are i) to mitigate the effect of model misspecification on estimation accuracy ii) to regularize the estimation problem in presence of poorly identifiable parameters, iii) to avoid estimation of initial conditions. The method is still under development but preliminary results have been presented at the Viral dynamics conference in October 2019.

7.2. High-dimensional and statistical learning

7.2.1. Automatic analysis of cell populations

New publication:

Hejblum BP, Alkhassim C, Gottardo R, Caron F, Thiébaut R, Sequential Dirichlet process mixture of skew t-distributions for model-based clustering of flow cytometry data, Annals of Applied Statistics, 13(1):638-660, 2019. DOI: 10.1214/18-AOAS1209.

7.2.2. High-dimensional compositional data analysis

Perrine Soret (PhD student in the axis "High-dimensional and statistical learning", supervised by M. Avalos) has applied our expertise in high dimensional data analysis to human microbiome field of research:

Soret P, Vandenborght LE, Francis F, Coron N, Enaud R, The Mucofong Investigation Group, Avalos M, Schaeverbeke T, Berger P, Fayon M, Thiébaut R and Delhaes L. Respiratory mycobiome and suggestion of inter-kingdom network during acute pulmonary exacerbation in cystic fibrosis. To appear in *Scientific Reports*.

7.2.3. 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. Hadrien Lorenzo (PhD student in the axis "High-dimensional and statistical learning", supervised by J. Saracco and R. Thiébaut) has developed a multi-block approach: the dd-sPLS method. dd-sPLS has been applied to high dimensional data analysis of different fields of research:

Lorenzo, H., Misbah, R., Odeber, J., Morange, P. E., Saracco, J., Trégouët, D. A., and Thiébaut, R. High-dimensional multi-block analysis of factors associated with thrombin generation potential. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 453-458). IEEE. https://hal.archives-ouvertes.fr/hal-02429302

Ellies-Oury, M. P., Lorenzo, H., Denoyelle, C., Saracco, J., and Picard, B. An Original Methodology for the Selection of Biomarkers of Tenderness in Five Different Muscles. Foods, 8(6), 206 (2019). https://hal.archives-ouvertes.fr/hal-02164157

7.3. Translational vaccinology

7.3.1. HIV vaccine development

We have finalized the data science analyses 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 149 LIGHT, a randomized phase II trial comparing a prime-boost therapeutic HIV vaccine strategy to placebo in HIV-infected patients undergoing antiretroviral treatment interruption. This included integrative statistical analyses using sPLS methods (as developed by the team) to relate markers from different high-dimensional immunogenicity or gene expression assays or virological assays to each other. In the ANRS VRI01 data set this allowed to disentangle the immune responses induced by the different vaccines used in the prime-boost strategies, showing specific effects of one of the vaccines (MVA HIV-B). We further identified a gene expression signature that correlates with later functional T-cell responses across the three different prime-boost association in which the MVA HIV-B vaccine was used. The corresponding manuscripts are currently in preparation.

Other HIV vaccine trials are currently being set-up by the French VRI (Vaccine Research Institute) and the European consortium EHVA, with strong contributions of SISTM team members to the trial designs.

7.3.2. Ebola vaccine development

The main results of the two randomized phase II Ebola vaccine trials conducted by the IMI-2 EBOVAC2 consortium (coordinated by Rodophe Thiébaut from the SISTM team) were finalized in 2019. and presented at international conferences. The results showed that the tested vaccine strategy (two-dose heterologous Ad26.ZEBOV and MVA-BN®-Filo Ebola vaccine regimen, developed by Janssen) was safe and immunogenic in both European and African volunteers (EBL2001 and EBL2002 trials). Deeper analyses of the induced immune responses are currently ongoing, and systems vaccinology analyses will soon start in the SISTM team.

The SISTM team is also a partner in the related IMI-2 EBOVAC1 and EBOVAC3 consortia (assesing the same vaccine regimen in other trial populations and/or trial phases), in which the main contributions of the team are related to mechanistic modeling of the immune responses (ongoing).

7.3.3. Vaccine development against other pathogens

Two other phase I vaccine trials (one testing a placental malaria vaccine, Primalvac trial; and one testing a nasal Pertussis vaccine, BPZE-1 trial), in which members of the SISTM team were strongly involved, have shown promising results. Results of the Primalvac trial have been accepted for publication in the Lancet Infectious Disease journal, and results of the BPZE-1 trial have been submitted for publication.

7.3.4. Methodological developments for vaccine trials

At the interface between the axis on "Mechanistic learning" and the axis "Translational vaccinology", modelling done within a PhD project (M. Alexandre, supervised by R. Thiébaut and M. Prague) has informed the definition of the primary endpoint and statistical analysis method to be used in two therapeutic HIV vaccine trials with antiretroviral treatment interruption (EHVA T02 and ANRS DALIA-2). The methodological choices and their rationale have been presented to the governance bodies of the research consortia, patient associations and have been submitted for ethics and regulatory approvals. This will be subject to a specific methodology publication.

Edouard Lhomme (PhD student in the axis "Translational Vaccinology, supervised by L. Richert) has developed a statistical method for functional T-cell assay data from vaccine trials (in particular the intracellular cytokine staining assay) that takes into account non-specific immune responses. We propose using a bivariate linear model for the analysis of the cellular immune responses to obtain accurate estimations of the vaccine effect. We benchmarked the performance of the model in terms of both bias and control of type-I and -II errors, and applied it to simulated data as well as real pre- and post-vaccination data from two recent HIV vaccine trials (ANRS VRI01 and ANRS 149 LIGHT in HIV-infected participants). This method has been published in the Journal of Immunological Methods and is now used in the SISTM team as the standard method for analyses of functional cellular data with non-stimulated control conditions, for instance in the currently ongoing analysis of cellular proliferation data from the EBOVAC2 EBL2001 trial. We have also established an online interface based on R Shiny to make this analysis method available for use by immunologists without specific training in statistical modelling (https://shiny-vici.apps.math.cnrs.fr/).

7.3.5. Prediction of the survival of patients based on RNA-seq data

In collaboration with the Inria MONC team, with the Inserm Angiogenesis and Tumor micro-environment team, and with clinicians from Milan and Bergen, the project GLIOMA-PRD aims to improve the prediction of the evolution of the lower grade glioma, a primary brain tumor, based on clinical, imaging and genomic data. In the SISTM team, we first had to determine the sufficient sample size for determining a predictive signature based on RNA-seq data for the survival of the patients. We concluded that 50 patients were a good enough sample size for this aim. We then explored the potential methods to analyze these data, with a particular focus on the methods grouping the genes by pathways, as the pilot data (The Cancer Genome Atlas Research Network, NEJM, 2015) showed a high correlation structure. We particularly compared two methods, Generalized Berk-Jones (GBJ), proposed by [54], and tcgsaseq, proposed by [1]. The first method could be applied to the survival context and thus be appropriate for our data.

Once the RNA-seq data were available, we could observe a high batch effect since the data were sequenced in two different lanes. One of these two batch included only patients that exhibited a particular tumor at the PET-scan, called "COLD", while the second batch included both patient labelled as "COLD", but also the other type of tumor, called "DIFFUSE". As this experiment structure might leads to confusing the difference due to the batch effect with the one due to the biological different, we had to explore methodologies that could remove this batch effect.

Once this batch effect removed, we will then analyze the RNA-seq data in order to identify the genes, or the group of genes, that could be predictive of the survival of the patients.

XPOP Project-Team

7. New Results

7.1. Modelling inheritance and variability of kinetic gene expression parameters in microbial cells

Modern experimental technologies enable monitoring of gene expression dynamics in individual cells and quantification of its variability in isogenic microbial populations. Among the sources of this variability is the randomness that affects inheritance of gene expression factors at cell division. Known parental relationships among individually observed cells provide invaluable information for the characterization of this extrinsic source of gene expression noise. Despite this fact, most existing methods to infer stochastic gene expression models from single-cell data dedicate little attention to the reconstruction of mother-daughter inheritance dynamics. Starting from a transcription and translation model of gene expression, we proposed a stochastic model for the evolution of gene expression dynamics in a population of dividing cells. Based on this model, we developed a method for the direct quantification of inheritance and variability of kinetic gene expression parameters from single-cell gene expression and lineage data. We demonstrated that our approach provides unbiased estimates of mother-daughter inheritance parameters, whereas indirect approaches using lineage information only in the post-processing of individual-cell parameters underestimate inheritance. Finally, we have shown on yeast osmotic shock response data that daughter cell parameters are largely determined by the mother, thus confirming the relevance of our method for the correct assessment of the onset of gene expression variability and the study of the transmission of regulatory factors [9].

7.2. Main effects and interactions in mixed and incomplete data frames

A mixed data frame (MDF) is a table collecting categorical, numerical and count observations. The use of MDF is widespread in statistics and the applications are numerous from abundance data in ecology to recommender systems. In many cases, an MDF exhibits simultaneously main effects, such as row, column or group effects and interactions, for which a low-rank model has often been suggested. Although the literature on low-rank approximations is very substantial, with few exceptions, existing methods do not allow to incorporate main effects and interactions while providing statistical guarantees. We proposed a new method that fills this gap [11], [3].

7.3. Quantification of gemcitabine intravenous drugs

This aim of this study was to assess the ability of Raman spectroscopy to quantify antineoplastic drugs directly in the finished product in plastic bags using a handled Raman spectrometer. Gemcitabine diluted in 0.9% sodium chloride was analyzed at various concentrations ranging from 1 to 20mg/mL directly through plastic bags using a handled 785nm Raman spectrometer.In accordance with EMA guidelines, quantitative models were developed to predict gemcitabine concentration in bag using partial least squares (PLS) regression. In order to evaluate the transposability of the developed Raman method and the routine method (flow injection analysis with UV detection), independent samples were analyzed using both techniques. The impact of the plastic bag was also evaluated by analysis samples through two different bags.The best model was obtained after standard normal variates preprocessing (SNV) for 15 latent variables. This model presented an excellent correlation between predicted and theoretical concentration values (R^2 of 0.9938 from the calibration set), a low limit of quantification (LLOQ) of 3.68mg/mL and acceptable repeatability and intermediate precision lower than the expected acceptance limit of 5% over the entire concentration range tested (except for the average concentration of 5.73mg/mL). For the 48 preparations higher than the LLOQ, the Bland-Altman approach showed the interchangeability of the wo methods with a difference bias of 2%. Moreover, no significant difference of predicted concentrations between the two containers tested ($p = 0.189$) was

observed.Despite some limitations for low concentrations, this study clearly shows promising results for real-time monitoring of gemcitabine infusion preparations without removing samples. The non-invasive nature of this method should ensure the correct dose before administration to patients and with heightened safety for operators [8].

7.4. Low-rank model with covariates for count data analysis

Count data are collected in many scientific and engineering tasks including image processing, single-cell RNA sequencing and ecological studies. Such data sets often contain missing values, for example because some ecological sites cannot be reached in a certain year. In addition, in many instances, side information is also available, for example covariates about ecological sites or species. Low-rank methods are popular to denoise and impute count data, and benefit from a substantial theoretical background. Extensions accounting for covariates have been proposed, but to the best of our knowledge their theoretical and empirical properties have not been thoroughly studied, and few softwares are available for practitioners. We propose a complete methodology called LORI (Low-Rank Interaction), including a Poisson model, an algorithm, and automatic selection of the regularization parameter, to analyze count tables with covariates. We also derive an upper bound on the estimation error. We provide a simulation study with synthetic data, revealing empirically that LORI improves on state of the art methods in terms of estimation and imputation of the missing values. We illustrate how the method can be interpreted through visual displays with the analysis of a well-know plant abundance data set, and show that the LORI outputs are consistent with known results. Finally we demonstrate the relevance of the methodology by analyzing a waterbirds abundance table from the French national agency for wildlife and hunting management (ONCFS). The method is available in the R package lori on the Comprehensive Archive Network (CRAN), [10].

7.5. Imputation and low-rank estimation with Missing Non At Random data

Missing values challenge data analysis because many supervised and unsupervised learning methods cannot be applied directly to incomplete data. Matrix completion based on low-rank assumptions are very powerful solution for dealing with missing values. However, existing methods do not consider the case of informative missing values which are widely encountered in practice. We propose matrix completion methods to recover Missing Not At Random (MNAR) data. Our first contribution is to suggest a model-based estimation strategy by modelling the missing mechanism distribution. An EM algorithm is then implemented, involving a Fast Iterative Soft-Thresholding Algorithm (FISTA). Our second contribution is to suggest a computationally efficient surrogate estimation by implicitly taking into account the joint distribution of the data and the missing mechanism: the data matrix is concatenated with the mask coding for the missing values ; a low-rank structure for exponential family is assumed on this new matrix, in order to encode links between variables and missing mechanisms. The methodology that has the great advantage of handling different missing value mechanisms is robust to model specification errors, [22].

7.6. A mathematical model to predict BNP levels in hemodialysis patients

Clinical interpretation of B-Type Natriuretic Peptide (BNP) levels in hemodialysis patients (HD) for fluid management remains elusive. We conducted a retrospective observational monocentric study. We built a mathematical model to predict BNP levels, using multiple linear regressions, [12].

7.7. Analysis of the global convergence of (fast) incremental EM methods

The EM algorithm is one of the most popular algorithm for inference in latent data models. The original formulation of the EM algorithm does not scale to large data set, because the whole data set is required at each iteration of the algorithm. To alleviate this problem, Neal and Hinton (1998) have proposed an incremental version of the EM (iEM) in which at each iteration the conditional expectation of the latent data (E-step) is updated only for a mini-batch of observations. Another approach has been proposed by Cappé and Moulines (2009) in which the E-step is replaced by a stochastic approximation step, closely related to stochastic gradient.

In this study, we analyzed incremental and stochastic version of the EM algorithm in a common unifying framework. We also introduced a new version incremental version, inspired by the SAGA algorithm by Defazio et al. (2014). We established non-asymptotic convergence bounds for global convergence, [15].

7.8. Efficient Metropolis-Hastings sampling 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 conduct such sampling, but such a method can converge slowly for high dimension problems, or when the joint structure of the distributions to sample is complex. We proposed a 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, in contrast with more sophisticated samplers such as the Metropolis Adjusted Langevin Algorithm or the No-U-Turn Sampler that involve costly tuning runs or intensive computation. Indeed, this distribution is automatically obtained thanks to a Laplace approximation of the original model. We have shown that such approximation is equivalent to linearizing the model in the case of continuous data, [14], [2].