RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2015

Section New Results

Edition: 2016-03-21

ABS Project-Team

6. New Results

6.1. Modeling Interfaces and Contacts

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

6.1.1. High Resolution Crystal Structures Leverage Protein Binding Affinity Predictions

Participants: Frédéric Cazals, Simon Marillet.

In collaboration with P. Boudinot, Unité de recherche en virologie et immunologie moléculaires, INRA Jouy-en-Josas.

Predicting protein binding affinities from structural data has remained elusive, a difficulty owing to the variety of protein binding modes. Using the structure-affinity-benchmark (SAB, 144 cases with bound/unbound crystal structures and experimental affinity measurements), prediction has been undertaken either by fitting a model using a handfull of pre-defined variables, or by training a complex model from a large pool of parameters (typically hundreds). The former route unnecessarily restricts the model space, while the latter is prone to overfitting.

We design models in a third tier [20], using twelve variables describing enthalpic and entropic variations upon binding, and a model selection procedure identifying the best sparse model built from a subset of these variables. Using these models, we report three main results. First, we present models yielding a marked improvement of affinity predictions. For the whole dataset, we present a model predicting K_d within one and two orders of magnitude for 48% and 79% of cases, respectively. These statistics jump to 62% and 89% respectively, for the subset of the SAB consisting of high resolution structures. Second, we show that these performances owe to a new parameter encoding interface morphology and packing properties of interface atoms. Third, we argue that interface flexibility and prediction hardness do not correlate, and that for flexible cases, a performance matching that of the whole SAB can be achieved. Overall, our work suggests that the affinity prediction problem could be partly solved using databases of high resolution complexes whose affinity is known.

6.1.2. Dissecting Interfaces of Antibody - Antigen Complexes: from Ligand Specific Features to Binding Affinity Predictions

Participants: Frédéric Cazals, Simon Marillet.

In collaboration with: P. Boudinot, Unité de recherche en virologie et immunologie moléculaires, INRA Jouy-en-Josas; M-P. Lefranc, Univ. of Montpellier 2.

B lymphocytes recognize the antigen through their membrane immunoglobulins (IG), that can also be secreted. The diversity of IG-Ag complexes challenges our understanding in terms of binding affinity and interaction specificity.

In this work [21], we dissect the interfaces of IG-Ag complexes from high resolution crystal structures. We show that global interface statistics distinguish ligand types and that interfacial side chains play a key role in the interaction. Our analysis of the relative positions of CDR identifies a remarkably conserved pattern involving seven seams between CDR, with specific variations depending on the ligand type. Finally, we show that structural features of the interface and of the partners yield binding affinity estimates of unprecedented accuracy (median absolute error of 1.02 kcal/mol).

Our findings will be of broad interest, as understanding Ag recognition at the atomic level will help guiding design of better IG targeting Ag for therapeutic or other uses.

6.2. Modeling Macro-molecular Assemblies

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

6.2.1. Unveiling Contacts within Macro-molecular assemblies by solving Minimum Weight Connectivity Inference Problems

Participants: Frédéric Cazals, Deepesh Agarwal.

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

Consider a set of oligomers listing the subunits involved in sub-complexes of a macro-molecular assembly, obtained e.g. using native mass spectrometry or affinity purification. Given these oligomers, connectivity inference (CI) consists of finding the most plausible contacts between these subunits, and minimum connectivity inference (MCI) is the variant consisting of finding a set of contacts of smallest cardinality. MCI problems avoid speculating on the total number of contacts, but yield a subset of all contacts and do not allow exploiting a priori information on the likelihood of individual contacts. In this context, we present two novel algorithms, `MILP-W` and `MILP-W`$_B$ [14]. The former solves the *minimum weight connectivity inference* (MWCI), an optimization problem whose criterion mixes the number of contacts and their likelihood. The latter uses the former in a bootstrap fashion, to improve the sensitivity and the specificity of solution sets.

Experiments on three systems (yeast exosome, yeast proteasome lid, human eIF3), for which reference contacts are known (crystal structure, cryo electron microscopy, cross-linking), show that our algorithms predict contacts with high specificity and sensitivity, yielding a very significant improvement over previous work, typically a twofold increase in sensitivity.

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

6.3. Modeling the Flexibility of Macro-molecules

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

6.3.1. Hybridizing Rapidly Growing Random Trees and Basin Hopping Yields an Improved Exploration of Energy Landscapes

Participants: Frédéric Cazals, Tom Dreyfus, Christine Roth.

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

The number of local minima of the potential energy landscape (PEL) of molecular systems generally grows exponentially with the number of degrees of freedom, so that a crucial property of PEL exploration algorithms is their ability to identify local minima which are low lying and diverse.

In this work [22], we present a new exploration algorithm, retaining the ability of basin hopping (`BH`) to identify local minima, and that of *transition based rapidly exploring random trees* (`T-RRT`) to foster the exploration of yet unexplored regions. This ability is obtained by interleaving calls to the extension procedures of `BH` and `T-RRT`, and we show tuning the balance between these two types of calls allows the algorithm to focus on low lying regions. Computational efficiency is obtained using state-of-the art data structures, in particular for searching approximate nearest neighbors in metric spaces.

We present results for the BLN69, a protein model whose conformational space has dimension 207 and whose PEL has been studied exhaustively. On this system, we show that the propensity of our algorithm to explore low lying regions of the landscape significantly outperforms those of `BH` and `T-RRT`.

6.4. Algorithmic Foundations

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

6.4.1. Beyond Two-sample-tests: Localizing Data Discrepancies in High-dimensional Spaces

Participants: Frédéric Cazals, Alix Lhéritier.

Comparing two sets of multivariate samples is a central problem in data analysis. From a statistical standpoint, the simplest way to perform such a comparison is to resort to a non-parametric two-sample test (TST), which checks whether the two sets can be seen as i.i.d. samples of an identical unknown distribution (the null hypothesis). If the null is rejected, one wishes to identify regions accounting for this difference. In this paper [17], we presents a two-stage method providing *feedback* on this difference, based upon a combination of statistical learning (regression) and computational topology methods.

Consider two populations, each given as a point cloud in $\mathbb{R}^d$. In the first step, we assign a label to each set and we compute, for each sample point, a discrepancy measure based on comparing an estimate of the conditional probability distribution of the label given a position versus the global unconditional label distribution. In the second step, we study the height function defined at each point by the aforementioned estimated discrepancy. Topological persistence is used to identify persistent local minima of this height function, their *basins* defining regions of points with high discrepancy and in spatial proximity.

Experiments are reported both on synthetic and real data (satellite images and handwritten digit images), ranging in dimension from $d = 2$ to $d = 784$, illustrating the ability of our method to localize discrepancies.

On a general perspective, the ability to provide feedback downstream TST may prove of ubiquitous interest in exploratory statistics and data science.

6.4.2. A Sequential Non-parametric Two-Sample Test

Participants: Frédéric Cazals, Alix Lhéritier.

Given samples from two distributions, a nonparametric two-sample test aims at determining whether the two distributions are equal or not, based on a test statistic. This statistic may be computed on the whole dataset, or may be computed on a subset of the dataset by a function trained on its complement. We propose a third tier [19], consisting of functions exploiting a sequential framework to learn the differences while incrementally processing the data. Sequential processing naturally allows optional stopping, which makes our test the first truly sequential nonparametric two-sample test.

We show that any sequential predictor can be turned into a sequential two-sample test for which a valid p-value can be computed, yielding controlled type I error. We also show that pointwise universal predictors yield consistent tests, which can be built with a nonparametric regressor based on k-nearest neighbors in particular. We also show that mixtures and switch distributions can be used to increase power, while keeping consistency.

AMIB Project-Team

6. New Results

6.1. RNA Design

In collaboration with J. Hales, J. Manuch and L. Stacho (Simon Fraser University/Univ. British Columbia, Canada), we have investigated the combinatorial RNA design problem, a minimal instance of the RNA design problem which aims at finding a sequence that admits a given target as its unique base pair maximizing structure. We obtained provide complete characterizations for the structures that can be designed using restricted alphabets. We provided a complete characterization of designable structures without unpaired bases. When unpaired bases are allowed, we provides partial characterizations for classes of designable/undesignable structures, and showed that the class of designable structures is closed under the stutter operation. Membership of a given structure to any of the classes can be tested in linear time and, for positive instances, a solution could be found in linear time. Finally, we considered a structure-approximating version of the problem that allows to extend helices and, assuming that the input structure avoids two motifs, we provided a linear-time algorithm that produces a designable structure with at most twice more base pairs than the input structure, as illustrated by Fig. 3 .

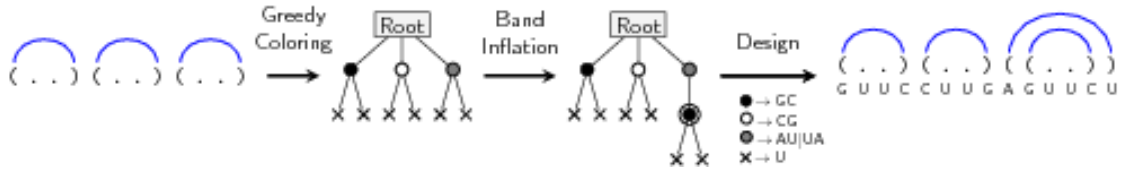


Figure 3. Principle of our structure-approximating version of RNA design: Starting from a potentially undesignable structure, a greedy coloring can be performed and corrected such that the final structure is provably designable in linear time.

Theses results were presented at the CPM 2015 conference in Italy [17], and open new avenues of research, both towards pratical, tractable versions of design, and constitute a first step towards long-awaited theoretical foundations for the problem.

6.2. Combinatorics of motifs and algorithms

We developed an $O(n)$-time and $O(n)$-space algorithm to compute minimal absent words. Their computation is used in sequence comparison [32] or to detect biologically significant events. For instance, in [52], it was shown that there exist three minimal words in *Ebola* virus genomes which are absent from human genome. The identification of such species-specific sequences may prove to be useful for the development of both diagnosis and therapeutics. In our new contribution [21] we provided an implementation that can be executed in parallel. Experimental resuts show that excluding the indexing data structure construction time, it achieves near-optimal speed-ups. The computation on the human genome is accelerated by a factor of 10 when using 16 processors, but it consummes a huge amout of RAM. Thus we are currently working on an external memory implementation, that will provide a trade-off between space and time consumption.

Combinatorial tools have been developed to predict the length of repetitions in a random sequence. This allows to distinguish biologically significant repetitions or tune some parameters in assembly or re-sequencing algorithms. For instance, unique mappability is strongly related to the length of the repetitions. A *trie profile* was defined in [45] to address this issue for binary alphabets, by the means of analytic combinatorics. General alphabets, where no closed formula exist, were adressed in [24]. An alternative, and simpler, approach is derived, thatexhibits a Large deviation Principle and makes use of Lagrange multipliers. Different domains and transition phases are exhibited. It is expected that htis approach extends to a Markov model and to approximate repetitions.

6.3. Structural variants

D. Iakovishina defended in 2015 a PhD thesis co-advised by M. Régnier and V. Boeva (Curie Institute). She proposed a new computational method to detect structural variants using whole genome sequencing data. It combines two techniques that are based either on the detection of paired-end mapping abnormalities or on the detection of the depth of coverage. SV-BAY relies on a probabilistic Bayesian approach and includes a modelization of possible sequencing errors, read mappability profile along the genome and changes in the GC-content. Keeping only somatic SVs is an additional option when matched normal control data are provided. SV-BAY compares favorably with existing tools on simulated and experimental data sets [12] Software SV-BAY is freely available https://github.com/InstitutCurie/SV-Bay.

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

Structural variations can also be observed and analyzed at larger time scales, and computational methods can be used to predict the structure of ancestral genomes. Within two collaborations with C. Chauve, A. Rajaraman (Simon Fraser University, Canada) and J. Zanetti (SFU, Canada & UniCAMP, Brazil), we revisited the problem of predicting a parsimonious set of adjacencies between ancestral genes, i.e. the most likely structure of an ancestral genome. More specifically, we modified the dynamic programming scheme underlying the DeCo algorithm [28] to compute indicators of robustness for predicting adjacencies. Our reimplementation, which relies on interesting meta-programming strategies, is available at https://github.com/yannponty/DeClone.

In a first study, we postulated a Boltzmann-Gibbs distribution over the set of evolutionary scenarii [9]. Our initial experiments relied on Boltzmann sampling to estimate the probabilities of ancestral adjacencies, but our extended version describes an exact polynomial-time computation of such probabilities, through an adaptation of the inside-outside algorithm. We interpreted such probabilities as supports for predicted adjacencies, and found that discarding adjacencies associated with low supports provided a good strategy for resolving synthenic conflicts.

However, the costs associated with the main operations (gaining/breaking adjacencies) in the underlying evolutionary models must be set beforehand in a somewhat arbitrary fashion. This has led us to investigate the influence of those costs on the characteristics of parsimonious predictions, i.e. the robustness of predictions with respect to perturbations of the scoring scheme [18]. To that purpose, we have performed an exact parametric analysis of the DeCo dynamic programming scheme (see Fig. 4 for details). This analysis revealed a quasi-independence, for a large subset of gene trees, of predicted adjacencies to the actual numerical values involved in the scoring scheme.

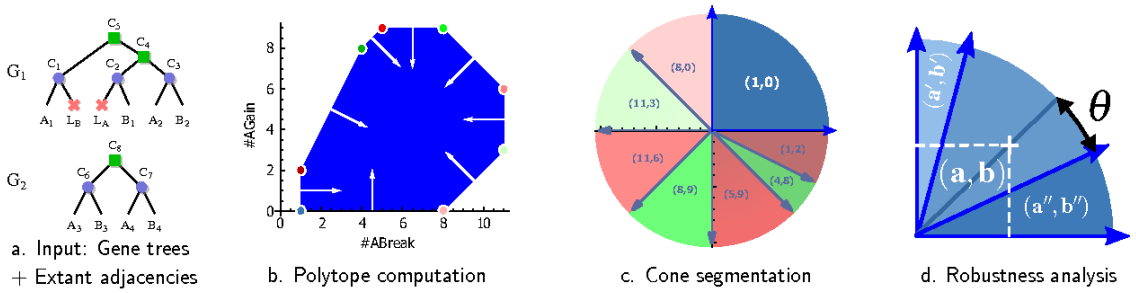


Figure 4. Main steps involved in the parametric prediction of ancestral adjacencies. Starting from two reconciled gene trees and a list of contemporary adjacencies (a.), the polytope of admissible Adjacency Gains/Breaks (+Presence/Absence of a given adjacency) is computed (b.) and projected onto a dual space which partitions the space of cost schemes into (infinite) regions leading to equivalent predictions (c.). The angular distance of the reference cost scheme $(1, 1)$ to a region representing an alternative prediction (d.) is used as a measure of robustness for the prediction.

BEAGLE Project-Team

7. New Results

7.1. Subspace Clustering Using Evolvable Genome Structure

We have developped an evolutionary algorithm to tackle the subspace clustering problem. Subspace clustering is recognized as more difficult than standard clustering since it requires to identify not only the clusters but also the various subspaces where the clusters hold. We propose to tackle this problem with a bio-inspired algorithm that includes many bio-like features like variable genome length and organization, functional and non-functional elements, and variation operators including chromosomal rearrangements. These features give the algorithm a large degree of freedom to achieve subspace clustering with satisfying results on a reference benchmark with respect to state of the art methods. One of the main advantages of the approach is that it needs only one subspace clustering ad-hoc parameter: the maximal number of clusters. This is a single and intuitive parameter that sets the maximal level of details of the clustering, while other algorithms require more complicated parameter space exploration. The other parameters of the algorithm are related to the evolution strategy (population size, mutation rate, ...) and for them we use a single setting that turns out to be effective on all the datasets of the benchmark.
This work has been presented at the main conference for genetic & evolutionary computation, GECCO [31], where it received the best paper award and during the EvoEvo Workshop of ECAL 2015 [35].

7.2. Epigenetic inheritance speeds up evolution of artificial organisms

DNA is not the sole medium by which parents transmit information to their offspring. Epigenetic inheritance, in particular, is based on the partial transmission of the cellular state of the parental cell to its descendants. Although the reality of epigenetic inheritance is now firmly established, whether it has an influence on the long term evolutionary process is still subject to debate. To address this question, we used the RAevol extention of the Aevol simulator developed in the team, and defined 4 scenarios with static or dynamic environments and with or without epigenetic inheritance. Simulations in dynamic environments show that protein inheritance indeed increases the rate of evolution on the long term. But they also show that it impedes evolution in its very first stages. This negative effect can be explained by instabilities generated by the interference between the two inheritance mediums. On the opposite, the long term gain can be explained by protein inheritance reducing the constraints on the genetic regulation network.
This work has been published in the article [33].

7.3. In silico evolution improves statistical models of genome dynamics

Using Aevol, we have proved that statistical frameworks published in the last twenty years for inferring evolutionary genome rearrangements are flawed in two ways. First, they mistranslated a null hypothesis on a uniform breakage model, and second, they assumed that genomic breakable regions are known *a priori*. We propose ways to correct these flaws by combining mathematical approaches, simulations, observations and validation on real genomic data. The results will be of interest for an audience from evolutionary biology, computational biology, bioinformatics and mathematics. We successively show that:

- a truly uniform hypothesis on rearrangement breakages leads to a model with an equilibrium intergene size distribution that fits the measured one on diverse genomes,
- estimations based on the flawed uniform breakage model completely fail on simulations with the truly uniform model,
- coherently with previous studies the flawed, and to a lesser extent, the truly uniform model are rejected on amniote genomes if breakable regions are identified with intergenic regions,
- co-estimating the number of breakable regions with the rearrangement distance gives coherent values on amniote genomes.

A paper reporting these results has been submitted by Priscila Biller, Carole Knibbe and Eric Tannier.

7.4. Temperature-induced variation in gene expression burst size in metazoan cells

Gene expression is an inherently stochastic process, owing to its dynamic molecular nature. Protein amount distributions, which can be acquired by cytometry using a reporter gene, can inform about the mechanisms of the underlying microscopic molecular system. By using different clones of chicken erythroid progenitor cells harboring different integration sites of a CMV-driven mCherry protein, we investigated the dynamical behavior of such distributions. We show that, on short term, clone distributions can be quickly regenerated from small population samples with a high accuracy. On longer term, on the contrary, we show variations manifested by correlated fluctuation in the Mean Fluorescence Intensity. In search for a possible cause of this correlation, we demonstrate that in response to small temperature variations cells are able to adjust their gene expression rate: a modest (2 °C) increase in external temperature induces a significant down regulation of mean expression values, with a reverse effect observed when the temperature is decreased. Using a two-state model of gene expression we further demonstrate that temperature acts by modifying the size of transcription bursts, while the burst frequency of the investigated promoter is less systematically affected. For the first time, we report that transcription burst size is a key parameter for gene expression that metazoan cells from homeotherm animals can modify in response to an external thermal stimulus.
This work has been published in the article [11].

7.5. Deciphering the signalling networks of synaptic plasticity

Synaptic plasticity, i.e. adaptive modifications of synaptic strength between two neurons depending on their activity, is a main substrate for learning and memory. Experimentally, synaptic plasticity is commonly assessed using prolonged electrical stimulations. Since learning can arise from few or even a single trial, synaptic strength is expected to adapt rapidly. However, whether synaptic plasticity occurs in response to limited event occurrences remains elusive. To address this question, we started a collaboration with Laurent Venance Lab (experimental neuroscience, College de France, Paris). Combining experimental and modelling approaches, we investigated whether a low number of stimulations can induce plasticity in a major synaptic learning rule, spike-timing-dependent plasticity (STDP). It is known that 100 stimulations induce bidirectional STDP, i.e. spike-timing-dependent potentiation (tLTP) and depression (tLTD) at most central synapses. In rodent striatum, we found that tLTD progressively disappears when the number of stimulations is decreased (below 50 pairings) whereas tLTP displays a biphasic profile: tLTP is observed for 75-100 stimulations, absent for 25-50 stimulations and re-emerges for 5-10 stimulations. This tLTP, induced by very few stimulations (5-10) depends on the endocannabinoid (eCB) system. The eCB system has recently emerged as a pivotal pathway for synaptic plasticity because of its widely characterized ability to depress synaptic transmission on short- and long-term scales. Our result therefore indicate that eCBs also mediate potentiation of the synapse. To understand how eCB signaling may support such bidirectionality, we combined electrophysiology experiments with mathematical modeling. Our model describes the temporal kinetics of the biochemical species involved in a first signaling pathway leading from NMDAR to calmodulin and CaMKII with that of a a second, distinct one that assembles mGluR and cytosolic calcium to eCB production and the resulting activation of CB1R. This demonstrated that STDP outcome is controlled by eCB levels and dynamics: prolonged and moderate levels of eCB lead to eCB-mediated long-term depression (eCB- tLTD) while short and large eCB transients produce eCB-mediated long-term potentiation (eCB-tLTP). Therefore, just like neurotransmitters glutamate or GABA, eCB forms a bidirectional system to encode learning and memory.
For reasons of publication strategy, our first co-publication on the subject presents our major experimental results [16]. A second article, featuring both experimental and modelling results, explains how the underlying signalling network can support the observed bidirectionality and is under submission.

7.6. Anomalous diffusion as an age-structured renewal process

Continuous-time random walks (CTRW) are one of the main mechanisms that are recurrently evoked to explain the emergence of subdiffusion in cells. CTRW were introduced fifty years ago as a generalisation of random walks, where the residence time (the time between two consecutive jumps) is a random variable. If the expectation of the residence time is defined, for instance when it is dirac-distributed or decays exponentially fast, one recovers "normal" Brownian motion. However, when the residence time expectation diverges, the CTRW describes a subdiffusive behavior. The classical approach to CTRW yields a non-Markovian (mean-field) transport equation, which is a serious obstacle when one wants to couple subdiffusion with (bio)chemical reaction. We took an alternative approach to CTRW that maintains the Markovian property of the transport equation at the price of a supplementary independent variable. We associate each random walker with an age a, that is the time elapsed since its last jump and describe the subdiffusive CTRW using an age-structured partial differential equations with age renewal upon each walker jump. In the spatially-homogeneous (zero-dimensional) case, we follow the evolution in time of the age distribution. An approach inspired by relative entropy techniques allows us to obtain quantitative explicit rates for the convergence of the age distribution to a self-similar profile, which corresponds to convergence to a stationary profile for the rescaled variables. An important difficulty arises from the fact that the equation in self-similar variables is not autonomous and we do not have a specific analytical solution. Therefore, in order to quantify the latter convergence, we estimate attraction to a time-dependent "pseudo-equilibrium", which in turn converges to the stationary profile.
The corresponding article is currently in press [38].

7.7. IGF-I signalling in neural stem cells during neurogenesis and aging

Downregulation of insulin-like growth factor (IGF) pathways prolongs lifespan in various species, including mammals. Still, the cellular mechanisms by which IGF signaling controls the aging trajectory of individual organs are largely unknown. Z. Chaker, in M. Holzenberg Lab (Centre de Recherche Saint-Antoine, Paris), asked whether suppression of IGF-I receptor (IGF-1R) in adult stem cells preserves long-term cell replacement, and whether this may prevent age-related functional decline in a regenerating tissue. Using neurogenesis as a paradigm, we showed that conditional knockout of IGF-1R specifically in adult neural stem cells maintained youthful characteristics of olfactory bulb neurogenesis within an aging brain. This in turn resulted in neuro-anatomical changes that improved olfactory function. To help interpret these results, we developed a mathematical model of stem cell differentiation using ordinary differential equations with time-dependent growth, division and death rates (to account for aging) and optimizing at each time step the amount of IGF-IR to maximize an experimentally-derived tissue efficiency criterion. The model predicts that decreased stimulation of growth in adults is indeed optimal for tissue aging. Thus, inhibiting growth and longevity gene IGF-1R in adult stem cells induced a gain-of-function phenotype during aging, marked by optimized management of cell renewal, and enhanced olfactory sensory function.
This work has been published in the article [14].

7.8. A novel model for leptin resistance

Leptin is a major hormone that regulates food intake and appetite in most mammals. Leptin increase in the blood tends to decrease the food intake and leptin is produced in proportion with fat depot. Leptin is therefore a simple probe that feed backs energy reserve to the brain and maintains a constant weight. It is a central hormone for this balance because KO mice without the leptin gene are quickly extremely obese. Also obese people (and animal) tend to have high concentration of leptin suggesting that after a certain point the brain ignores the leptin signal. We developed a mathematical model that explores this resistance developed by neural cells to leptin. This model predicts leptin resistance if food intake is artificially increased and predict a pathway to obesity by such mechanism. This work has been published by H. Soula (Beagle) in collaboration with F. Crauste (Dracula) with co-supervised PhD student Marine Jacquier[19].

7.9. Without eye contact, birds are Markovian!

Any social birds rely on acoustic messages to organize their daily activity (such as parenting and food foraging). In many occasions, birds are within earshot but not in visual contact and therefore should rely only on acoustic channel for this communication. In collaboration with the University of Saint-Etienne, we developed automatic extraction scripts that can detect birds vocalizations in a protocol of meeting with decreasing distance and with or without visual contact modality. Our worked showed that without visual contact birds are more synchronized and their vocal dynamics cannot be distinguished from a two state Markov chain. This markov property vanishes as soon as visual contact is restablished. This work has been published in the main ethology journal: Animal Behaviour[23].

BIGS Project-Team

7. New Results

7.1. Stochastic modeling

7.1.1. Tumor growth modeling

Participants: P. Vallois, S. Wantz-Mézières
External collaborator: J-S. Giet (IECL,Université de Lorraine)

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

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

where $(B_t)_{t\geq 0}$ is a standard brownian motion starting from zero. Then (i) We define a family of time continuous Markov chains which models the evolution of the rate of malignant cells and approximate (under some conditions) the diffusion process (V_t). (ii) We study in depth the solution to equation (1). This diffusion process lives in a domain delimited by two boundaries: 0 and $\kappa > 0$. In this stochastic setting, the role of κ is not so clear and we contribute to understand it. We describe the asymptotic behavior of the diffusion according to the values of the parameters. The tools we resort to are boundary classification criteria and Laplace transform of the hitting time to biological worthwhile level. We are able in particular to express the mean of the hitting time. We have an accepted paper in the journal Theory of Stochastic Processes [70].

7.1.2. A Multitype Branching Process Model of Heterogeneous Damages in vitro Cancer Cell Populations Treated by Radiotherapy

Participants: T. Bastogne, P. Vallois
External collaborator: S. Pinel (CRAN, Université de Lorraine)

Cancer is the result of inter-dependent multi-scale phenomena and this is mainly why the understanding of its spread is still an unsolved problem. In integrative biology, mathematical models play a central role; they help biologists and clinicians to answer complex questions through numerical simulations and statistical analyses. The main issue here is to better understand and describe the role of cell damage heterogeneity and associated mutant cell phenotypes in the therapeutic responses of cancer cell populations submitted to a radiotherapy sessions during *in vitro* experiments. The cell heterogeneity is often described as randomness in mathematical modeling and different representations, such as Markov chains, branching processes and even stochastic differential equations, have been recently used. Conversely to these previous studies, which only focused on the steady-state responses of cell populations, we are interested by modeling the transient behavior after treatment and to identify the role of mutation heterogeneity in the global dynamic response of the cell populations. We propose to describe the survival response of anin vitro cancer cell culture treated by radiotherapy as a superposition of independent dynamics. Each cell is represented by a finite collection of cell mutation states with possible transitions between them. The population dynamics is given by an age-dependent multi-type branching process. From this representation, we obtain equations satisfied by the average size of the global survival population as well as the one of subpopulations associated with 10 mutation phenotypes. This work was presented via a poster communication in a international congress [40].

7.1.3. Modeling of response to chemotherapy in gliomas

Participant: S. Wantz-Mézières
External collaborators: M. Ben Abdallah, Yann Gaudeau, J.-M. Moureaux (CRAN, Université de Lorraine) and M. Blonski, L. Taillandier (CHU Nancy)

In the framework of a collaboration with neurologists (Luc Taillandier, Marie Blonski, CHU Nancy) and automaticians (Jean-Marie Moureaux, Yann Gaudeau, CRAN), around the thesis supervision of M. Ben Abdallah, our aim is to work out personalized therapeutic strategy in the monitoring of diffuse low-grade glioma patients. Regular monitoring with MRI are used to estimate the tumour volume ; we proposed a method by manual segmentation and statistically assessed its reproducibility by a subjective test. In order to design a decision-aid tool for the response to chemotherapy, our approach is phenomelogical and we used simple regression tools to model and predict the cinetics of the tumour growth. We identified two different models. These results open up many perspectives, the main one being the modeling by multi-factor models, including biological and anatomopathological factors. This work is currently in progress.

7.1.4. Photodynamic therapy

Participant: C. Lacaux
External collaborators: T. Obara and M. Thomassin (CRAN, Université de Lorraine), L. Vinckenboch (Fribourg)

Our project focuses on an innovative application: the interstitial PDT for the treatment of high-grade brain tumors. This strategy requires the installation of optical fibers to deliver the light directly into the tumor tissue to be treated, while nanoparticles are used to carry the photosensitizer into the cancer cells. In order to optimize the intra-cerebral position of our optical fiber, two fundamental questions have to be answered: (1) What is the optimal shape and position of the light source in order to optimize the damage on malignant cells? (2) Is there a way to identify the physical parameters of the tissue which drive the light propagation?

Notice that we are obviously not the first ones to address these issues, and there is nowadays a consensus in favor of the algorithm proposed by L. Wang and S. L. Jacques for the simulation of light transport in biological tissues. However, our starting point is the observation that the usual methods slightly lack of formalism and miss formal representations that answer the questions of identifiability. In [16], in the framework of homogeneous biological tissues, we propose an alternative MC method to Wang's algorithm. Then we also propose a variance reduction method. Interestingly enough, our formulation also allows us to design quite easily a Markov chain Monte Carlo (MCMC) method based on Metropolis-Hastings algorithm and to handle the inverse problem (of crucial importance for practitioners), consisting in estimating the optical coefficients of the tissue according to a series of measurements. We have compared the proposed MC and MCMC method and Wang's algorithm: we see that our MC method is much more consistent. However, MCMC methods induce quick mutations, which paves the way to very promising algorithms in the inhomogenous case. To handle the inverse problem, we derive a probabilistic representation of the variation of the fluence with respect to the absorption and scattering coefficients. This leads us to the implementation of a Levenberg-Marquardt type algorithm that gives an approximate solution to the inverse problem. Our results open the way for new improvements of Monte-Carlo methods in the context of light propagation. They should rather be seen as a starting point for new methods, including in inhomogeneous tissue. This work has been presented in several french seminars (Lille, Avignon, Paris Descartes, Orléans).

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

Participant: C. Lacaux
External collaborator: G. Samorodnitsky (Cornell, USA)

In extreme value theory, one of the major topics is the study of the limiting behavior of the partial maxima of a stationary sequence. When this sequence is i.i.d., the unique limiting process is well-known and called the extremal process. Considering a long memory stable sequence, the limiting process is obtained as a simple power time change extremal process. Céline Lacaux and Gennady Samorodnistky have proved in [38] that this limiting process can also be interpreted as a restriction of a self-affine random sup measure. In addition, they have established that this random measure arises as a limit of the partial maxima of the same long memory stable sequence, but in a different space. Their results open the way to propose new self-similar processes with stationary max-increments. Céline Lacaux has presented this work in an invited session of the international conference *Extreme Value Analysis* at Ann Arbor (June 2015).

7.1.6. Modulus of continuity of some conditionally sub-Gaussian fields, application to stable random fields

Participant: C. Lacaux
External collaborator: H. Biermé (Poitiers)

Hermine Biermé and Céline Lacaux maintain their collaboration on the study of anisotropic random fields. They have extended their previous work in the framework of conditionally sub-Gaussian random series. For such anisotropic fields, they have obtained a modulus of continuity and a rate of uniform convergence. Their framework enables the study of study e.g., Gaussian fields, stable random fields and multi-stable random fields. As invited speaker, Céline Lacaux has presented this work in the international conference *Adventure in Self-similarity* at Cornell University (June 2015) [17]. Another of their works in progress deals with the simulation of anisotropic Gaussian random fields and the estimation of their parameters using quadratic variations.

7.1.7. DNA sequences analysis

Participants: P. Vallois
External collaborators: A. Lagnoux and S. Mercier (Toulouse)

Here we want to determine the sequences that are biologically interesting and compare the results using the single local score Hn and using the pair (Hn;Ln) where Ln is the length of the segment that realizes the best score. In that view, we work on the p-values associated to the observed samples.

7.1.8. Multicriteria Agregation for Health Economic Assessment

Participants: T. Bastogne, Y. Petot, P. Vallois

The framework of this work is the PhD thesis of Yann Petit. The first chapter of the thesis is a state of the art identifying the current challenges in medico-economic analyses. A review article should be submitted in spring 2016. We are currently working on the aggregation operators, based on fuzzy measures and the Choquet integral. Theoretical results have been obtained and a publication is planned to be submitted in the second half of 2016. Work continues by introducing probabilities. The next step will be to apply our theoretical results to real clinical cases.

7.1.9. Spatial and spatio-temporal modeling

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

In the framework of a collaboration with INRA Bordeaux about the esca-illness of vines, Anne Gégout-Petit with Shuxian Li developed different spatial models and spatio-temporal models for different purposes: (1) study the distribution and the dynamics of esca vines in order to tackle the aggregation and the potential spread of the illness (2) propose a spatio-temporal model in order to capture the dynamics of cases and measure the effects of environmental covariates. For this, we propose different hierarchic models with latent process associated with a bayesian inference. A part of the research has been submitted in a journal of biology [39]. Shuxian Li defended his PhD on December the 15th.

7.1.10. Stochastic modeling of fatigue crack propagation

Participants: R. Azaïs, A. Gégout-Petit
External collaborators: A.B. Abdessalem, M. Puiggali, M. Touzet (Bordeaux)

Fatigue crack propagation is a stochastic phenomenon due to the inherent uncertainties originating from material properties and environmental conditions. In a recent preprint [35], we propose to model and to predict the fatigue crack growth by a piecewise-deterministic Markov process associated with deterministic crack laws of the literature, namely the Paris-Erdogan equation defined by $\mathrm{d}a/\mathrm{d}N = C(\Delta K)^m$ and the Forman equation given by $\mathrm{d}a/\mathrm{d}N = C(\Delta K)^m/(K_c(1-R)-\Delta K)$, where a is the crack length, N denotes the number of cyclic mechanical loads, ΔK is the range of the stress intensity factor and C, m, K_c and R are different parameters. We introduce a regime-switching model to express the transition between Paris' regime and rapid propagation which occurs just before failure. We also investigate the prediction of the fatigue crack path and its variability based on measurements taken at the beginning of the propagation. This work has also been presented in an international conference [25].

7.2. Estimation and control for stochastic processes

7.2.1. Inference for dynamical systems driven by Gaussian noises

Participant: S. Tindel
External collaborators: K. Chouk, A. Deya, Y. Hu, L. Khoa, D. Nualart, E. Nualart, F. Xu. (US)

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

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

where θ is the unknown parameter, b is a smooth enough coefficient and B is a one-dimensional fractional Brownian motion. In contrast with this simple situation, our applications of interest (motivated by some anomalous diffusion phenomenon in proteins fluctuations) require the analysis of the following $\mathbb{R}^n$-valued equation:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\, du + \int_0^t \sigma(\theta; X_u)\, dB_t, \tag{2}$$

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

To this aim, here are the steps we have focused on in 2015:

- Some limit theorems for general functionals of Gaussian sequences [6], or for functionals of a Brownian motion [3], which give some insight on the asymptotic behavior of systems like (2).
- Extension of pathwise stochastic integration to processes indexed by the plane in [1], which helps to the definition of noisy systems such as partial differential equations.
- Definition of new systems driven by a (spatial) fractional Brownian motion, such as the stochastic PDE considered in [37].
- The local asymptotic normality obtained for the system (2), which implies a lower bound on general estimators of the coefficient θ. This is the contents of the preprint [41].

7.2.2. Optimal estimation of the jump rate of a piecewise-deterministic Markov process

Participants: R. Azaïs, A. Muller-Gueudin

A piecewise-deterministic Markov process is a stochastic process whose behavior is governed by an ordinary differential equation punctuated by random jumps occurring at random times. In a recent preprint [33], we focus on the nonparametric estimation problem of the jump rate for such a stochastic model observed within a long time interval under an ergodicity condition. More precisely, we introduce an uncountable class (indexed by the deterministic flow) of recursive kernel estimates of the jump rate and we establish their strong pointwise consistency as well as their asymptotic normality. In addition, we propose to choose among this class the estimator with the minimal variance, which is unfortunately unknown and thus remains to be estimated. We also discuss the choice of the bandwidth parameters by cross-validation methods. This paper has also been presented in two national workshops.

7.2.3. Estimation and optimal control for the TCP process

Participant : R. Azaïs
External collaborators: N. Krell (Rennes), B. de Saporta (Montpellier)

In [33], we assume that the transition kernel is continuous with respect to the Lebesgue measure. This condition may be not satisfied in some applications, as for instance for the well-known TCP process that appears in the modeling of the famous Transmission Control Protocol used for data transmission over the Internet. As a consequence, we propose to investigate estimation followed by optimal control for this ergodic process. The particular framework defined by this process allows us to define an optimal policy for the estimation of its jump rate. We obtain at present an efficient method for estimating the moments of the conditional distribution of the inter-congestion times in an optimal way. This work is currently in progress.

7.2.4. Estimation of integrals from a Markov design

Participant : R. Azaïs
External participants: B. Delyon, F. Portier

Monte-Carlo methods for estimating an integral assume that the distribution of the random design is known. Unfortunately, some applications generate a design whose density function f is unknown. In this case, a solution is to perform the classical Monte-Carlo estimate of the integral by replacing f by a leave-one-out kernel estimator, and one may expect the convergence

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\varphi(X_i)}{\widehat{f}^{(-i)}(X_i)} \to \int \varphi \mathrm{d}\lambda,$$

when the number n of independent data X_i goes to infinity. This difficult question has been investigated by François Portier and Bernard Delyon in a recent paper. We propose to extend this work to the more general case of a Markov design. This new model includes a large variety of applications, in particular in biology and climatology. Indeed, the data $(X_i, \varphi(X_i))$ are often obtained from a measuring instrument that is launched in its environment and thus follows a random walk in it. A paper on this work will be submitted soon.

7.2.5. Method of control for radiotherapy treatment using Decision Markov Processes

Participants : R. Azaïs, B. Scherrer, S. Tindel, S. Wantz-Mézières

In recent years, Bastogne, Keinj and Vallois designed a Markov model of the evolution of cells under a radiotherapy treatment. We are currently investigating the problem of optimizing the radiotherapy intensity sequence in order to kill as many cancerous cells as possible while preserving as many healthy cells, a problem that fits into the stochastic optimal control problem. Our preliminary efforts suggest that, since we are dealing with large populations of cells, the problem can be well approximated by a limit deterministic optimal control problem. We can solve this problem numerically with a Pontryagine approach, and symbolically (in the simplest cases) by identifying the critical points of some multivariate polynomials. The latter approach allows us to validate the fact that the former actually finds globally optimal solutions. This is a work in progress.

7.2.6. Numerical approximate schemes for large optimal control problems and zero-sum two player games

Participant: B. Scherrer
External collaborators: V. Gabillon, M. Ghavamzadeh, M. Geist, B. Lesner, J. Perolat, O. Pietquin, M. Tagorti

We have provided in [23] (ICML 2015) the first finite-sample analysis of the LSTD(λ) algorithm aimed at approximating the value of some fixed policy in a large MDP, through the approximation of the projected fixed point of the linear Bellman equation from samples. This analysis highlights the influence of the main parameter λ of the algorithm.

The long version of our previous work on the analysis of an approximate modified policy iteration for optimal control and its application to the Tetris domain is now published in JMLR [13]. The extension of this algorithm family for computing approximately-optimal non-stationary policies allows to improve the dependency with respect to the discount factor: we provide such improved bounds in [19], as well as examples that show that our analysis is tight (and cannot be further improved).

An original analysis of the variation of the approximate modified policy iteration for computing approximate Nash equilibria in the more general setting of two-player zero-sum games was published in ICML 2015 [22].

7.3. Algorithms and estimation for graph data

7.3.1. Modelling of networks of multiagent systems

Participant: A. Muller-Gueudin
External collaborators: A. Girard, S. Martin, I.C. Morarescu (CRAN, Nancy)

We relate here a starting of collaboration with researchers in Automatics in Nancy. We consider here networks, modeled as a graph with nodes and edges representing the agents and their interconnections, respectively. The objective is to study the evolution of the opinion of all the agents. The connectivity of the network, persistence of links and interactions reciprocity influence the convergence speed towards a consensus. The problem of consensus or synchronization is motivated by different applications as communication networks, power and transport grids, decentralized computing networks, and social or biological networks. We then consider networks of interconnected dynamical systems, called agents, that are partitioned into several clusters. Most of the agents can only update their state in a continuous way using only inner-cluster agent states. On top of this, few agents also have the peculiarity to rarely update their states in a discrete way by reseting it using states from agents outside their clusters. In social networks, the opinion of each individual evolves by taking into account the opinions of the members belonging to its community. Nevertheless, one or several individuals can change their opinions by interacting with individuals outside its community. These inter-cluster interactions can be seen as resets of the opinions. This leads us to a network dynamics that is expressed in term of reset systems. We suppose that the reset instants arrive stochastically following a Poisson renewal process. We have an accepted paper in the journal IEEE Transactions on Automatic Control [10].

7.3.2. Microbial interaction inference by network analysis

Participants: A. Gégout-Petit, A. Muller-Gueudin
External collaborators: A. Deveau (INRA Nancy), C. Raïssy (Inria Orpailleur)

The objective is to characterize microbial interactions in a particular environment: the truffles.

The truffle provides a habitat for complex bacterial communities. The role for bacteria in the development of truffles has been suggested but very little is known regarding the structure and the functional potential of the truffle's bacterial communities along truffle maturation. In a mathematical point of view, two micro-organisms are connected if they are not independent, conditionally to the other micro-organisms. Several models fit into this setting, especially the gaussian graphical models, the bayesians networks, and the graphical log-linear models. But the data, which can be zeros inflated, need developments and we have to proposed new models. Moreover, we are confronted to the problem that $n \ll p$, that is the sample size is much smaller that the number of variables ($n = 30, p = 200$). Last year, thanks to a financially supported project (PEPS), we have began a collaboration between statisticians and data-miners. The first approches have been notified in a report [31]. The statistical methodologies developed for this project could also be applied to human health (for instance identification of network between bacteria inside colon).

7.3.3. Lossy compression of unordered trees

Participant: R. Azaïs
External collaborators: J-B. Durand, C. Godin

A classical compression method for trees is to represent them by directed acyclic graphs. This approach exploits subtree repeats in the structure and is efficient only for trees with a high level of redundancy. The class of self-nested trees presents remarkable compression properties by this method because of the systematic repetition of subtrees. In particular, the compressed version of a self-nested tree T is a linear directed acyclic graph with only $1 + \text{height}(T)$ nodes. Unfortunately, it should be noted that trees without a high level of redundancy are often insufficiently compressed by this procedure. In a paper recently submitted for publication in an international conference [32], we introduce a lossy compression method that consists in computing in polynomial time for trees with bounded outdegree the reduction of a self-nested structure that closely approximates the initial data. We prove on a simulated dataset that the error rate of this lossy compression method is always better than the loss involved in a previous algorithm of the literature, while the compression rates are equivalent.

7.3.4. Inference for critical Galton-Watson trees from their Harris process

Participant: R. Azaïs
External collaborator: A. Genadot (Inria CQFD Bordeaux)

Galton-Watson trees are an elementary model for the genealogy of a branching population and thus play a central role in biology. Critical Galton-Watson trees are generated from a sibling distribution μ whose theoretical expectation $\sum k\mu(k)$ is equal to 1. Under this assumption, the well-known Harris process of a tree conditioned on having n nodes converges to a Brownian excursion characterized by the variance $\sigma^2 = \sum (k-1)^2 \mu(k)$ of μ. We propose to exploit this asymptotic approximation to define a new estimate of the unknown parameter of interest σ^2 based on a least-square method. In particular, this new technique allows us to take into account the behavior of the Harris path with respect to its asymptotic theoretical expectation. In certain cases, we obtain a better confidence interval than the classical approach. A paper on this work is in preparation.

7.4. Regression and machine learning

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

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

In this work with Myriam Maumy-Bertrand (IRMA, Strasbourg), we study the conditional cumulative distribution function and a nonparametric estimator associated to this function. The conditional cumulative distribution function has the advantages of completely characterizing the law of the random considered variable, allowing to obtain the regression function, the density function, the moments and the conditional quantile function. As a nonparametric estimator of this function, we focus on local polynomial techniques described in Fan and Gijbels [64]. In particular, we use the local linear estimation of the conditional cumulative distribution function.

The objective of this work is to establish uniform asymptotic certainty bands for the conditional cumulative distribution function. To this aim, we give exact rate of strong uniform consistency for the local linear estimator of this function. We show that limit laws of the logarithm are useful in the construction of uniform asymptotic certainty bands for the conditional distribution function. In particular, we use a single bootstrap to construct sharp uniform asymptotic bands of this estimator.

We illustrate our results with simulations and a study of fetal growth which is based on 694 fetuses (carefully selected by exclusion of multiple pregnancies, malformed, macerated or serious ill fetuses, or those with chromosomal abnormalities) autopsied in fetopathologic units of the "Service de foetopathologie et de placentologie" of the Maternité Régionale Universitaire (CHU Nancy, France) between 1996 and 2013.

We have presented our results in two international conferences with proceedings in Lille in June 2015 ("47èmes Journées de Statistique de la SFdS") [21] and London in December 2015 ("CM Statistics") [36].

7.4.2. Omnibus tests for regression models.

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

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

7.4.3. Data analysis techniques: a tool for cumulative exposure assessment

Participant: J-M. Monnez
External collaborators : W. Kihal, B. Lalloué, C. Padilla, D, S. Zmirou-Navier

Everyone is subject to environmental exposures from various sources, with negative health impacts (air, water and soil contamination, noise, etc.) or with positive effects (e.g. green space). Studies considering such complex environmental settings in a global manner are rare. We propose to use statistical factor and cluster analyses to create a composite exposure index with a data-driven approach, in view to assess the environmental burden experienced by populations. We illustrate this approach in a large French metropolitan area. The study was carried out in the Great Lyon area (France, 1.2 M inhabitants) at the census Block Group (BG) scale. We used as environmental indicators ambient air NO2 annual concentrations, noise levels and proximity to green spaces, to industrial plants, to polluted sites and to road traffic. They were synthesized using Multiple Factor Analysis (MFA), a data-driven technique without a priori modeling, followed by a Hierarchical Clustering to create BG classes. The first components of the MFA explained, respectively, 30, 14, 11 and 9% of the total variance. Clustering in five classes group: (1) a particular type of large BGs without population; (2) BGs of green residential areas, with less negative exposures than average; (3) BGs of residential areas near midtown; (4) BGs close to industries; and (5) midtown urban BGs, with higher negative exposures than average and less green spaces. Other numbers of classes were tested in order to assess a variety of clustering. We present an approach using statistical factor and cluster analyses techniques, which seem overlooked to assess cumulative

exposure in complex environmental settings. Although it cannot be applied directly for risk or health effect assessment, the resulting index can help to identify hot spots of cumulative exposure, to prioritize urban policies or to compare the environmental burden across study areas in an epidemiological framework [9].

7.4.4. Online Partial Principal Component Analysis of a Data Stream

Participant: J-M. Monnez
External collaborator: R. Bar (EDF, R & D)

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

Moreover, several incremental procedures of linear and logistic regression of a data stream were defined and tested and compared on existing batch data files and on simulated data streams.

7.4.5. Prognostic Value of Estimated Plasma Volume in Heart Failure.

Participant: J-M. Monnez
External collaborators: E. Albuisson, B. Pitt, P. Rossignol, F. Zannad (CHU, Nancy)

The purpose of this study was to assess the prognostic value of the estimation of plasma volume or of its variation beyond clinical examination in a post-hoc analysis of EPHESUS (Eplerenone Post-Acute Myocardial Infarction Heart Failure Efficacy and Survival Study).

Assessing congestion after discharge is challenging but of paramount importance to optimize patient management and to prevent hospital readmissions.

The present analysis was performed in a subset of 4,957 patients with available data (within a full dataset of 6,632 patients). The study endpoint was cardiovascular death or hospitalization for heart failure (HF) between months 1 and 3 after post-acute myocardial infarction HF. Estimated plasma volume variation (ΔePVS) between baseline and month 1 was estimated by the Strauss formula, which includes hemoglobin and hematocrit ratios. Other potential predictors, including congestion surrogates, hemodynamic and renal variables, and medical history variables, were tested.

An instantaneous estimation of plasma volume at month 1 was defined and also tested.

Multivariate analysis was performed with stepwise logistic regression. ΔePVS was selected in the model. The corresponding prognostic gain measured by integrated discrimination improvement was significant. Nevertheless, instantaneous estimation of plasma volume at month 1 was found to be a better predictor than ΔePVS. LDA with mixed variables was also performed and confirmed these results.

In HF complicating myocardial infarction, congestion as assessed by the Strauss formula and an instantaneous derived measurement of plasma volume provided a predictive value of early cardiovascular events beyond routine clinical assessment. Prospective trials to assess congestion management guided by this simple tool to monitor plasma volume are warranted [4].

7.4.6. Death or hospitalization scoring for heart failure patients

Participant: J-M. Monnez
External collaborator: E. Albuisson (CHU Nancy)

The purpose of this study was to define an event - death or hospitalization - score for heart failure patients based on the observation of biological, clinical and medical historical variables. Some of them were transformed or winsorized. Two methods of statistical learning were performed, logistic regression and linear discriminant analysis, with a stepwise selection of variables. Aggregation of classifiers by bagging was used. Finally a score taking values between 0 and 100 was established.

7.4.7. A simultaneous stepwise covariate selection and clustering algorithm to discriminate a response variable with numerous values

Participant: J-M. Monnez
External collaborator: O. Collignon (LIH, Luxembourg)

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

7.4.8. Statistical Analyses of Cell Impedance Signals in High-Throughput Cell Analysis

Participant: T. Bastogne, L. Batista
External Collaborator: El-Hadi Djermoune (Université de Lorraine, CRAN)

With the advent of high-throughput technologies, life scientists are starting to grapple with massive data sets, encountering challenges with handling, processing and moving information that were once the domain of astronomers and high-energy physicists [91]. We particularly focus the statistical analysis of large batch of time series with applications in the preclinical research in Cancerology. Our original contribution consists in developing new dynamical system identification methods suited to the processing of those type of data. System identification is a data-driven modeling approach more and more used in biology and biomedicine. In this application context, each assay is always repeated to estimate the response variability. The inference of the modeling conclusions to the whole population requires to account for the inter-individual variability within the modeling procedure. One solution consists in using mixed effects models but up to now no similar approach exists in the field of dynamical system identification. Therefore, our objective is to develop a new identification method integrating mixed effects within an ARX (Auto Regressive model with eXternal inputs) model structure. The parameter estimation step relies on the EM (Expectation-Maximisation) algorithm. First simulation results show the relevance of this solution compared with a classical procedure of system identification repeated for each subject. This work and derived was accepted in conference papers [34] [24] [18].

BONSAI Project-Team

7. New Results

7.1. Ancestral gene order reconstruction

In the field of **genomic rearrangement**, a topic of interest is to infer ancestral gene order from gene order known in extant species. The problem resumes to compute a set ancestral CARs (continuous ancestral regions) at a given node of a phylogeny. This work, initially published in a conference, was published this year in a journal [5].

7.2. Nonribosomal peptides

Norine is the unique and leading platform dedicated to computational biology analysis of nonribosomal peptides (NRPs). It is used by thousands of scientists all over the world to explore and better understand the diversity of the NRPs. To improve the data quality and quantity in Norine, we are now opening our resource to external contributors. To achieve this new challenge, we developed new tools (MyNorine, s2m) and communicate on our novelties.

- **Crowdsourcing.** To facilitate the submission of new nonribosomal peptides (NRPs) or modification of stored ones in Norine, we have developed a dedicated and user-friendly module named MyNorine [2]. It provides interactive forms to fill in the annotations with, for example, autocompletion and tools such as a monomeric structure editor. It has especially been designed for biologists and biochemists working on secondary metabolites to easily enrich the database with their own data.
- **Norine communication.** We advertise Norine by different promoting media. We organized an international workshop in Lille in October to teach biologists and biochemists how to annotate NRPs and their synthetases with bioinformatics tools such as Norine. It attracted 32 attendees from 8 countries. We participated, as invited contributors, to the special issue "Bioinformatics tools and approaches for synthetic biology" of the new journal Synthetic and Systems Biotechnology edited by KeAi Publishing, funded by Elsevier and Chinese Science Publishing & Media. Our article [6] describes the usefulness of Norine to discover novel nonribosomal peptides, with examples of biological results obtained thanks to Norine tools. More than 20 NRPs have already been submitted since September, proving the efficiency of our communication and usefulness and relevancy of Norine.
- **Monomeric structure.** The tool Smiles2Monomers (abbreviated s2m) infers efficiently and accurately the monomeric structure of a polymer from its chemical structure [1]. It is provided to the scientific community through the Norine website for on-line run or for download. Beside its utility to facilitate the annotation of new peptides, it allowed us to detect annotation errors in the Norine database.

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

High-throughput V(D)J repertoire analysis is an activity started in the group in 2012. As mentioned in previous reports, we produced a platform dedicated to analysing lymphocyte populations: Vidjil. Starting from DNA sequences, Vidjil is able to identify and quantify lymphocyte populations, visualise them and store metadata. Vidjil is now used routinely in Lille hospital and is also tested in other laboratories around the world.

With collaborators in Prague we used Vidjil in a retrospective study on patients suffering acute lymphoblastic leukemia [3]. The study identified a new measure of predicting relapse in patients, just a month after the diagnosis. This measure is simple as it relies on the diversity of the lymphocyte population.

7.4. Spaced seed coverage

In the field of spaced seed statistics these last two years, a new challenge is the selection of a set of spaced seeds that are at the same time sensitive, while providing a stable similarity measure for *alignment-free genomic sequence comparison*. One of the most stable estimators is the *coverage* provided by these seeds. We have proposed an efficient method to build the coverage automaton, in order to compute several statistics efficiently. This work was implemented in Iedera and published in the AISM journal [4].

7.5. Genome scaffolding with contaminated data

Scaffolding is a cornerstone in the assembly of genomes from next-generation sequencing data. It consists in ordering assembled sequences according to their putative order and orientation in the source genome. However, we are almost always in a setting where the genome is not known. Instead, order and orientation of sequences are inferred from partial information present in the sequencing data.

Unfortunately, sequencing data is noisy and often has contamination, i.e. a subset of the data which indicates a wrong genome order and/or orientation. We have investigated this effect and designed the first algorithm that explicitly models this contamination to better perform scaffolding.

This work appeared in the proceedings of the WABI 2015 conference [9] and has been accepted to the Bioinformatics journal, currently under revision. This work is in collaboration with K. Sahlin and L. Arvestad (KTH, Sweden).

7.6. Mining metatranscriptomic data

The team has recently developed the SortMeRNA software, which is a sequence analysis tool for filtering, mapping and OTU-picking NGS reads. The core algorithm is based on approximate seeds and allows for fast and sensitive analyses of nucleotide sequences. In [11], we demonstrate a computational technique for filtering ribosomal RNA from total RNA in metatranscriptomic data using it. Additionally, we propose a post-processing pipeline using the latest software tools to conduct further studies on the filtered data, including the reconstruction of mRNA transcripts for functional analyses and phylogenetic classification of a community using the ribosomal RNA. This work is a collaboration with Genoscope.

7.7. Structured RNAs

In many families of strutured RNAs, the signature of the family cannot be characterized by a single consensus structure, and is mainly described by a set of alternate secondary structures. For example, certain classes of RNAs adopt at least two distinct stable folding states to carry out their function. This is the case of riboswitches, that undergo structural changes upon binding with other molecules, and recently some other RNA regulators were proven to show evolutionary evidence for alternative structure. The necessity to take into account multiple structures also arises when modeling an RNA family with some structural variation across species, or when it comes to work with a set of predicted suboptimal foldings. In this perspective, we have introduced the concept of RNA multistructures, that is a formal grammar based framework specifically designed to model a set of alternate RNA secondary structures. Continuing our work of 2014, we provide several motivating examples and propose an efficient algorithm to search for RNA multistructures within a genomic sequence. This work was published in [7].

CAPSID Project-Team

7. New Results

7.1. Annotating 3D Protein Domains

Many protein chains in the Protein Data Bank (PDB) are cross-referenced with EC numbers and Pfam domains. However, these annotations do not explicitly indicate any relation between EC numbers and Pfam domains. In order to address this limitation, we developed EC-DomainMiner, a recommender-based approach for associating EC (Enzyme Commission) numbers with Pfam domains [19]. EC-DomainMiner is able to infer automatically 20,179 associations between EC numbers and Pfam domains from existing EC-chain/Pfam-chain associations from the SIFTS database as well as EC-sequence/Pfam-sequence associations from UniProt databases.

7.2. Large-Scale Analysis of 3D Protein Interactions

As part of a continuing collaboration with a former doctoral student in the Orpailleur team, Anisah Ghoorah (now at the University of Mauritius), we used her KBDOCK database of all known PPIs to perform a large-scale statistical analysis of the secondary structure composition of known protein-protein binding sites [14]. This showed that some combinations of secondary structure features are significantly favoured, whereas other combinations are considerably dis-favoured. These findings could provide knowledge-based rules for the prediction of unsolved protein-protein interactions.

7.3. Predicting Drug Side Effects

Together with Harmonic Pharma SAS (a LORIA / Inria spin-out company), we developed the “GESSE” method for proposing new uses for existing therapeutic drug molecules by associating the Gaussian shapes of known drug molecules with their clinically observed side-effects [15].

7.4. Modeling a GPCR Receptor Complex

In collaboration with the BIOS team (INRA Tours) and the AMIB team (Inria Saclay – Île de France) we used our Hex protein docking software to help model a multi-component G-protein coupled receptor (GPCR) complex [12]. The resulting 3D structure was shown to be consistent with the known experimental data for the protein components of this trans-membrane molecular signaling system.

7.5. Modeling the Apelin Receptor

The Apelin receptor (ApelinR) is a GPCR which is important in regulating cardiovascular homeostasis. As part of an on-going collaboration with the Centre for Interdisciplinary Research (CIRB) at Collège de France, we modeled the interaction between the Apelin peptide and ApelinR [13]. This study provides new mechanistic insights which could lead to the development of therapeutic agents for the treatment of heart failure.

7.6. Identifying New Anti-Fungal Agents

In this collaboration with several Brasilian laboratories (at University of Mato Grosso State, University of Maringá, Embrapa, and University of Brasilia), we identified several novel small-molecule drug leads against the pathogenic fungus *Paracoccidioides lutzii* [17] which is a serious health threat, especially in Brasilian hospitals.

DYLISS Project-Team

7. New Results

7.1. Data integration

Participants: Jacques Nicolas, Charles Bettembourg, Jérémie Bourdon, Jeanne Got, Marie Chevallier, Guillaume Collet, Olivier Dameron, Damien Eveillard, Julie Laniau, Anne Siegel.

Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies. Interaction graphs provide a suitable representation of cellular networks with information flows. Methods based on sign consistency have been shown to be valuable tools to (i) predict qualitative responses, (ii) test the consistency of network topologies and experimental data, and (iii) apply repair operations to the network model suggesting missing or wrong interactions. We present a framework to unify different notions of sign consistency and propose a refined method for data discretization that considers uncertainties in experimental profiles. We furthermore introduce a new constraint to filter undesired model behaviors induced by positive feedback loops. Finally, we generalize the way predictions can be made by the sign consistency approach. This corresponds to an extension of our *Bioquali* software. [*Anne Siegel*] [21]

Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach. Our software tool *shogen* was used to decipher functional roles within a consortium of five mining bacteria through the integration of genomic and metabolic knowledge at genome scale. We first reconstructed a global metabolic network. Next, using a parsimony assumption, we deciphered sets of genes, called Sets from Genome Segments (SGS), that (i) are close on their respective genomes, (ii) take an active part in metabolic pathways and (iii) whose associated metabolic reactions are also closely connected within metabolic networks. The use of SGS (shogen) pinpoints a functional compartmentalization among the investigated species and exhibits putative bacterial interactions necessary for promoting these pathways. [*Damien Eveillard, Anne Siegel*] [17]

Optimal Threshold Determination for Interpreting Semantic Similarity and Particularity: Application to the Comparison of Gene Sets and Metabolic Pathways Using GO and ChEBI. We developed a method for determining optimal semantic similarity and particularity thresholds in order to interprete the results of the comparison of ontology terms sets. We applied this method on the GO and ChEBI ontologies. Qualitative analysis using the thresholds on the PPAR multigene family yielded biologically-relevant patterns. [*Charles Bettembourg, Olivier Dameron*] [16]

AskOmics : Integration et interrogation de reseaux de regulation genomique et post-genomique. We present AskOmics, an integration and interrogation software using a RDF model and the SPARQL query language. The purpose of this work is to obtain quick answers to biological questions demanding currently hours of manual search in several spreadsheet results files. AskOmics allows biologists to integrate and interrogate their data by themselves without any knowledge about RDF and SPARQL required. [*Charles Bettembourg, Olivier Dameron*] [30]

7.2. Time-series and asymptotic dynamics

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

Identification of logical models for signaling pathways: towards a systems biology loop. Logical models of signaling pathways are a promising way of building effective in silico functional models of a cell. The automated learning of Boolean logic models describing signaling pathways can be achieved by training to phosphoproteomics data. This data is unavoidably subject to noise. As a result, the learning process leads to a family of feasible logical networks rather than a single model. This family is composed of logic models proposing different internal wirings for the system, implying that the logical predictions from this family may suffer a significant level of variability leading to uncertainty. In our work, combinatorial optimization methods based on recent logic programming paradigm allow to enumerate, and discriminate the family of logical models explaining data. Together, these approaches enable a robust understanding of the system response. The results are implemented in the *caspo* software [*Jacques Nicolas, Anne Siegel*] [22], [23]

Boolean Network Identification from Multiplex Time Series Data. The ASP-based learning algorithm developed in the team to train logical models of signaling networks focuses on the comparison of two time-points and assumes that the system has reached an early steady state. We have generalized such a learning procedure in order to discriminate Boolean networks according to their transient dynamics. To that goal, we exhibit a necessary condition that must be satisfied by a Boolean network dynamics to be consistent with a discretized time series trace. This approach was included in the ASP-based framework designed for the caspo software. We ended up with a global learning algorithm and compared it to learning approaches based on static data. [*Anne Siegel*] [31]

Representation of symbolic dynamical systems generated by a substitution. Iterated morphisms are combinatorial processes which are related to several classes of dynamical systems appearing in several fields of computer sciences and mathematics: numeration, ergodic theory, discrete geometry. They may be associated to fractal sets called "Rauzy fractals" whose topological properties are linked to the properties of the underlying dynamical system. We have introduced a generic algorithm framework to check such topological properties within a complete family of iterated morphism. This makes efficient the verification of conjectures on several families of substitutions related to multi-dimensional continued fraction algorithms. [*Anne Siegel*] [32], [25], [14]

Multivariate Normal Approximation for the Stochastic Simulation Algorithm: Limit Theorem and Applications. We present a central limit theorem for the Gillespie stochastic trajectories when the living system has reached a steady-state, that is when the internal bio-molecules concentrations are assumed to be at equilibrium. It appears that the stochastic behavior in steady-state is entirely characterized by the stoichiometry matrix of the system and a single vector of reaction probabilities. We propose several applications of this result such as deriving multivariate confidence regions for the time course of the system and a constraints-based approach which extends the flux balance analysis framework to the stochastic case. [*Jérémie Bourdon, Vincent Picard, Anne Siegel*] [20], [12]

A Logic for Checking the Probabilistic Steady-State Properties of Reaction Networks. Designing probabilistic reaction models and determining their stochastic kinetic parameters are major issues in systems biology. In order to assist in the construction of reaction network models, we introduce a logic that allows one to express asymptotic properties about the steady-state stochastic dynamics of a reaction network. Basically, the formulas can express properties on expectancies, variances and co-variances. We demonstrate that deciding the satisfiability of a formula is NP-hard. [*Jérémie Bourdon, Vincent Picard, Anne Siegel*] [28], [12]

7.3. Sequence and structure annotation

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

Amplitude Spectrum Distance: measuring the global shape divergence of protein fragments. We introduce here the Amplitude Spectrum Distance (ASD), a novel way of comparing protein fragments based on the discrete Fourier transform of their C_α distance matrix. Defined as the distance between their amplitude spectra, ASD can be computed efficiently and provides a parameter-free measure of the global shape dissimilarity of two fragments. ASD inherits from nice theoretical properties, making it tolerant to shifts, insertions,

deletions, circular permutations or sequence reversals while satisfying the triangle inequality. The practical interest of ASD with respect to RMSD, $RMSD_d$, BC and TM scores is illustrated through zinc finger retrieval experiments and concrete structure examples. The benefits of ASD are also illustrated by two additional clustering experiments: domain linkers fragments and complementarity-determining regions of antibodies. [*Clovis Galiez, François Coste*] [19]

Structural conservation of remote homologues: better and further in contact fragments. We address a basic question on sequence-structure relationships in proteins: does a protein sequence depict a structure with a uniform faithfulness all along the sequence ? We investigate this question by defining contact fragments. This study suggests that sequence homologs of CF are significantly more faithful to structure than randomly chosen fragments, so that CF carry a strong sequence-structure relationship, allowing them to be used as accurate building blocks for structure prediction. [*Clovis Galiez, François Coste*] [26]

VIRALpro: a tool to identify viral capsid and tail sequences. Not only sequence data continues to outpace annotation information, but the problem is further exacerbated when organisms are underrepresented in the annotation databases. This is the case with non human-pathogenic viruses which occur frequently in metagenomic projects. Thus there is a need for tools capable of detecting and classifying viral sequences. We describe VIRALpro a new effective tool for identifying capsid and tail protein sequences, which are the cornerstones toward viral sequence annotation and viral genome classification. [*Clovis Galiez, François Coste*] [18]

Finding Optimal Discretization Orders for Molecular Distance Geometry. The Molecular Distance Geometry Problem (MDGP) is the problem of finding the possible conformations of a molecule by exploiting available information about distances between some atom pairs. Under minimal assumptions the MDGP can be discretized so that the search domain of the problem becomes a tree that can be explored by using an interval Branch & Prune (iBP) algorithm. In this context, the discretization assumptions are strongly dependent on the atomic ordering, which can also impact the computational cost of the iBP algorithm. In this work, we propose a new partial discretization order for protein backbones. This new atomic order optimizes a set of objectives that aim at improving the iBP performances. The optimization of the objectives is performed by Answer Set Programming (ASP), which allows to express the problem by a set of logical constraints. The comparison with previously proposed orders for protein backbones shows that this new discretization order makes iBP perform more efficiently. [*Jacques Nicolas*] [34]

From formal concepts to analogical complexes. Reasoning by analogy is an important component of common sense reasoning whose formalization has undergone recent improvements with the logical and algebraic study of the analogical proportion. The starting point of this study considers analogical proportions on a formal context. We introduce analogical complexes, a companion of formal concepts formed by using analogy between four subsets of objects in place of the initial binary relation. They represent subsets of objects and attributes that share a maximal analogical relation. We show that the set of all complexes can be structured in an analogical complex lattice and give explicit formulae for the computation of their infimum and supremum. [*Laurent Miclet, Jacques Nicolas*] [27]

Comparison of the targets obtained by a scoring matrix and by a regular expression. Application to the search for LXR binding sites. In bioinformatics, it is a common task to search for new instances of a pattern built from a set of reference sequences. For the simplest and most frequent cases, patterns are represented in two ways : regular expression or scoring matrix. Since both representations seem to be used indifferently in pratice, one may wonder if they have any impact on the result. This study compares hits obtained with scoring matrices or by regular expressions allowing up to two substitutions. It shows that, in our LXR study, sequences found by a scoring matrix are closer to the targeted hits than sequences found by a regular expression. [*Aymeric Antoine-Lorquin, Jacques Nicolas, Catherine Belleannée*] [29]

Finding and Characterizing Repeats in Plant Genomes. Plant genomes contain a particularly high proportion of repeated structures of various types. This chapter proposes a guided tour of available softwares that can help biologists to look for these repeats and check some hypothetical models intended to characterize their structures. Since transposable elements are a major source of repeats in plants, we have provided a whole section on this topic as well as a selection of the main existing softwares. In order to better understand how

they work and how repeats may be efficiently found in genomes, the rest of the chapter is devoted to the foundations of the search for repeats and more complex patterns. We first introduce the key concepts that are useful for understanding the current state of the art in playing with words, applied to genomic sequences. In fact, biologists need to represent more complex entities where a repeat family is built on more abstract structures, including direct or inverted small repeats, motifs, composition constraints as well as ordering and distance constraints between these elementary blocks. The last section introduces concepts and practical tools that can be used to reach this syntactic level in biological sequence analysis. [*Jacques Nicolas*] [35]

ERABLE Project-Team

6. New Results

6.1. General comments

We present in this section the main results obtained in 2015. Some were already in preparation or submitted at the end of 2014. It will be indicated whenever this is the case.

We tried to organise the results following four of the five main axes of research of the team. Clearly, in some cases, a result obtained overlaps more than one axis. We chose the one that could be seen as the main concerned by such results. As concerns the Axis "Going towards control", a work is in preparation that fits it. It will be presented in 2016.

We did not indicate here the results on more theoretical aspects of computer science if it did not seem for now that they could be relevant in contexts related to computational biology. Actually, we do believe those on scheduling (by, among others, A. Marchetti-Spaccamela and/or L. Stougie) [3], [38], [39], [10], [31], [23], [44], [43] or even one result related to context-free grammars (by, among others, P. Crescenzi) [11] could in the future become relevant for the life sciences (biology or ecology). However, we preferred for now to only indicate the theoretical results related to problems closely resembling questions that have already been addressed by us in computational biology.

Notice that such CS results concern not only cross-fertilising issues among different computational approaches, and we therefore extended the title of this axis for the purpose of presenting such results, for now purely theoretical.

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

6.2. Identifying the molecular elements

Genomic / NGS data management
Next-generation sequencing (NGS) technology has led the life sciences into the big data era. Today, sequencing genomes takes little time and cost, but yields terabytes of data to be stored and analysed. The biologists are often exposed to excessively time consuming and error-prone data management and analysis hurdles. We therefore proposed a database management system (DBMS) based approach to accelerate and substantially simplify genome sequence analysis [9]. To that aim, we extended MonetDB, an open-source column-based DBMS (urlhttps://www.monetdb.org), with a BAM module, which enables easy, flexible, and rapid management and analysis of sequence alignment data stored as Sequence Alignment/Map (SAM/BAM) files. The main features of MonetDB/BAM were described using a case study on Ebola virus.

We also designed and realised a knowledge base for collecting, elaborating, and extracting analytical results of genomic, proteomic, biochemical, morphological investigations from animal models of cerebral stroke [45]. Data analysis techniques are tailored to make the data available for processing and correlation, in order to increase the predictive value of the preclinical data, to perform bio-simulation studies, and to support both academic and industrial research in the area of cerebral stroke therapy. The low reliability of animal models in replicating the human disease is one of the most serious problems in the field of medical and pharmaceutical research about stroke. The standard models for the study of ischaemic stroke are often poorly predictive as they simulate only partially the human disease. This work aims therefore at investigating animal models with diseases typically associated with the onset of stroke in human patients. A first statistical analysis of the retrieved information led to the validation of our animal models and suggested a predictive and translational value for parameters related to a specific model. In particular, concerning gene expression data, we applied a data analysis pipeline that initially takes into account an initial set of 64,000 genes and brought down the focus on a few tens of them.

NGS data analysis

The problem of enumerating bubbles with length constraints in directed graphs arises in transcriptomics where the question is to identify all alternative splicing events present in a sample of mRNAs sequenced by RNA?seq. We presented a new algorithm for enumerating bubbles with length constraints in weighted directed graphs [30]. This is the first polynomial delay algorithm for this problem and we showed that in practice, it is faster than previous approaches. This settled one of the main open questions from previous literature. Moreover, the new algorithm allows us to deal with larger instances and possibly detect longer alternative splicing events.

We also developed CIDANE, a novel framework for genome-based transcript reconstruction and quantification from RNA-seq reads [37]. CIDANE assembles transcripts with significantly higher sensitivity and precision than existing tools, while competing in speed with the fastest methods. In addition to reconstructing transcripts *ab initio*, the algorithm also allows to make use of the growing annotation of known splice sites, transcription start and end sites, or full-length transcripts, which are available for most model organisms. CIDANE supports the integrated analysis of RNA-seq and additional gene-boundary data and recovers splice junctions that are invisible to other methods.

SNPs (Single Nucleotide Polymorphisms) are genetic markers used in many areas of biology. Their precise identification is a prerequisite for association studies, which associate genotypes to phenotypes. Methods are currently developed for model species, but rely on the availability of a (good) reference genome, and cannot be applied to non-model species. They are also mostly tailored for whole genome (re-)sequencing experiments, whereas in many cases, transcriptome sequencing can be used as a cheaper alternative which already enables to identify SNPs located in transcribed regions. We proposed a method that identifies, quantifies and annotates SNPs without any reference genome, using RNA-seq data only. Individuals can be pooled prior to sequencing, if not enough material is available for sequencing from one individual. This pooling strategy still enables to allelotype loci and to associate them to phenotypes. Using human RNA-seq data, we first compared the performance of our algorithm, KISSSPLICE, with GATK, a well established method that requires a reference genome. We showed that both methods perform similarly in terms of precision and recall. We then validated experimentally the predictions of our method using RNA-seq data from two non-model species. The method can be used for any species to annotate SNPs and to predict their impact on proteins. It can further be used to assess variants that are associated to a particular phenotype within a population, when replicates are provided for each biological condition. This work was submitted at the end of 2015.

Sequence alignment (full genomes or NGS data)

Sequence comparison is a fundamental step in many important tasks related to biology. Traditional algorithms for measuring approximation in sequence comparison are based on the notions of distance or similarity, and are generally computed through sequence alignment techniques. As circular genome structure is a common phenomenon in nature, a caveat of specialised alignment techniques for circular sequence comparison is that they are computationally expensive, requiring from super-quadratic to cubic time in the length of the sequences. We introduced a new distance measure based on q-grams, and showed how it can be computed efficiently for circular sequence comparison [41]. Experimental results, using real and synthetic data, demonstrated orders-of-magnitude superiority of our approach in terms of efficiency, while maintaining an accuracy very competitive to the state of the art.

Burrows-Wheeler Transform (BWT) has been successfully used to reduce the memory requirement for sequence alignment. We improved on previous results related to the problem of computing the Burrows-Wheeler Transform (BWT) using small additional space [12]. Our in-place algorithm does not need the explicit storage for the suffix sort array and the output array, as typically required in such previous work. It relies on the combinatorial properties of the BWT, and runs in $O(n^2)$ time in the comparison model using $O(1)$ extra memory cells, apart from the array of n cells storing the n characters of the input text. We then discussed the time-space trade-off when $O(k\sigma k)$ extra memory cells are allowed with σk distinct characters, providing an $O((n^2/k + n)\log?k)$-time algorithm to obtain (and invert) the BWT. In real systems where the alphabet size is a constant, for any arbitrarily small $\epsilon > 0$, the BWT of a text of n bytes can be computed in $O(n\sigma^{-1}\log n)$ time using just σn extra bytes.

Genome assembly problems

The human genome is diploid, which requires assigning heterozygous single nucleotide polymorphisms (SNPs) to the two copies of the genome. The resulting haplotypes, lists of SNPs belonging to each copy, are crucial for downstream analyses in population genetics. Currently, statistical approaches, which are oblivious to direct read information, constitute the state-of-the-art. Haplotype assembly, which addresses phasing directly from sequencing reads, suffers from the fact that sequencing reads of the current generation are too short to serve the purposes of genome-wide phasing. While future-technology sequencing reads will contain sufficient amounts of SNPs per read for phasing, they are also likely to suffer from higher sequencing error rates. Currently, no haplotype assembly approaches exist that allow for taking both increasing read length and sequencing error information into account. We developed WHATSHAP, the first approach that yields provably optimal solutions to the weighted minimum error correction problem in runtime linear in the number of SNPs [25]. WHATSHAP is a fixed parameter tractable (FPT) approach with coverage as the parameter. We demonstrated that WHATSHAP can handle datasets of coverage up to 20x, and that 15x are generally enough for reliably phasing long reads, even at significantly elevated sequencing error rates. We also find that the switch and flip error rates of the haplotypes we output are favourable when comparing them with state-of-the-art statistical phasers. By using novel combinatorial properties of Minimum Error Correction (MEC) instances, we were then able to provide new results on the fixed-parameter tractability and approximability of MEC [35]. In particular, we showed that MEC is in FPT when parameterised by the number of corrections, and, on "gapless" instances, it is in FPT also when parameterised by the length of the fragments, whereas the result known in the literature forces the reconstruction of complementary haplotypes. We then showed that MEC cannot be approximated within any constant factor while it is approximable within factor $O(\log nm)$ where nm is the size of the input. Finally, we provided a practical 2-approximation algorithm for the Binary MEC, a variant of MEC that has been applied in the framework of clustering binary data. Finally, by exploiting a feature of future-generation technologies – the uniform distribution of sequencing errors – we designed an exact algorithm, called HAPCOL, that is exponential in the maximum number of corrections for each SNP position and that minimises the overall error-correction score [26]. We performed an experimental analysis, comparing HAPCOL with the current state-of-the-art combinatorial methods both on real and simulated data. On a standard benchmark of real data, we showed that HAPCOL is competitive with state-of-the-art methods, improving the accuracy and the number of phased positions. Furthermore, experiments on realistically-simulated datasets revealed that HAPCOL requires significantly less computing resources, especially memory. Thanks to its computational efficiency, HAPCOL can overcome the limits of previous approaches, allowing to phase datasets with higher coverage and without the traditional all-heterozygous assumption.

Completing the genome sequence of an organism is an important task in comparative, functional and structural genomics. However, this remains a challenging issue from both a computational and an experimental viewpoint. Genome scaffolding (*i.e.* the process of ordering and orientating contigs) of de novo assemblies usually represents the first step in most genome finishing pipelines. We developed MEDUSA (Multi-Draft based Scaffolder), an algorithm for genome scaffolding [6]. MEDUSA exploits information obtained from a set of (draft or closed) genomes from related organisms to determine the correct order and orientation of the contigs. MEDUSA formalises the scaffolding problem by means of a combinatorial optimisation formulation on graphs and implements an efficient constant factor approximation algorithm to solve it. In contrast to currently used scaffolders, it does not require either prior knowledge on the microrganisms dataset under analysis (*e.g.* their phylogenetic relationships) or the availability of paired end read libraries. This makes usability and running time two additional important features of our method. Moreover, benchmarks and tests on real bacterial datasets showed that MEDUSA is highly accurate and, in most cases, outperforms traditional scaffolders. The possibility to use MEDUSA on eukaryotic datasets has also been evaluated, leading to interesting results. medusa/releases.

Genome annotation problems

Repetitive DNA, including transposable elements (TEs), is found throughout eukaryotic genomes. Annotating and assembling the "repeatome" during genome-wide analysis often poses a challenge. To address this problem, we developed DNAPIPETE – a new pipeline that uses a sample of raw genomic reads [20]. It produces precise estimates of repeated DNA content and TE consensus sequences, as well as the relative ages of TE families. We showed that DNAPIPETE performs well using very low coverage sequencing in different

genomes, losing accuracy only with old TE families. We applied this pipeline to the genome of the Asian tiger mosquito *Aedes albopictus*, an invasive species of human health interest, for which the genome size is estimated to be over 1 Gbp. Using DNAPIPETE, we showed that this species harbours a large (50% of the genome) and potentially active repeatome with an overall TE class and order composition similar to that of *Aedes aegypti*, the yellow fever mosquito. However, intra-order dynamics showed clear distinctions between the two species, with differences at the TE family level. Our pipeline's ability to manage the repeatome annotation problem will make it helpful for new or ongoing assembly projects, and our results will benefit future genomic studies of *A. albopictus*.

On another topic, we developed a reliable, robust, and much faster method for the prediction of pre-miRNAs [22]. With this method, we aimed mainly at two goals: efficiency and flexibility. Efficiency was made possible by means of a quadratic algorithm. Since the majority of the predictors use a cubic algorithm to verify the pre-miRNA hairpin structure, they may take too long when the input is large. Flexibility relies on two aspects, the input type and the organism clade. MIRINHO can receive as input both a genome sequence and small RNA sequencing (sRNA-seq) data of both animal and plant species. To change from one clade to another, it suffices to change the lengths of the stem-arms and of the terminal loop. Concerning the prediction of plant miRNAs, because their pre-miRNAs are longer, the methods for extracting the hairpin secondary structure are not as accurate as for shorter sequences. With MIRINHO, we also addressed this problem, which enabled to provide premiRNA secondary structures more similar to the ones in MIRBASE than the other available methods. MIRINHO served also as the basis to two other issues we addressed. The first issue led to the treatment and analysis of sRNA-seq data of *Acyrthosiphon pisum*, the pea aphid. The goal was to identify the miRNAs that are expressed during the four developmental stages of this species, allowing further biological conclusions concerning the regulatory system of such an organism. For this analysis, we developed a whole pipeline, called MIRINHOPIPE, at the end of which MIRINHO was aggregated. A paper is currently being prepared that presents this work.

6.3. Inferring and analysing the networks of molecular elements

Protein structure comparison

We proposed a new distance measure for comparing two protein structures based on their contact map representations [1]. We showed that our novel measure, which we refer to as the maximum contact map overlap (max-CMO) metric, satisfies all properties of a metric on the space of protein representations. Having a metric in that space allows one to avoid pairwise comparisons on the entire database and, thus, to significantly accelerate exploring the protein space compared to no-metric spaces. We showed on a gold standard superfamily classification benchmark set of 6759 proteins that our exact k-nearest neighbour ($k-NN$) scheme classifies up to 224 out of 236 queries correctly and on a larger, extended version of the benchmark with 60; 850 additional structures, up to 1361 out of 1369 queries. Our $k-NN$ classification thus provides a promising approach for the automatic classification of protein structures based on flexible contact map overlap alignments.

Metabolic network analysis

Flux balance analysis (FBA) is one of the most often applied methods on genome-scale metabolic networks. Although FBA uniquely determines the optimal yield, the pathway that achieves this is usually not unique. The analysis of the optimal-yield flux space has been an open challenge. Flux variability analysis is only capturing some properties of the flux space, while elementary mode analysis is intractable due to the enormous number of elementary modes. However, it had been previously found that the space of optimal-yield fluxes decomposes into flux modules. These decompositions allow a much easier but still comprehensive analysis of the optimal-yield flux space. Using the mathematical definition of module introduced by Müller and Bockmayr in 2013, we discovered that flux modularity is rather a local than a global property which opened connections to matroid theory [28]. Specifically, we showed that our modules correspond one-to-one to so-called separators of an appropriate matroid. Employing efficient algorithms developed in matroid theory we are now able to compute the decomposition into modules in a few seconds for genome-scale networks. Using that every module can be represented by one reaction that corresponds to its function, we also presented a method that uses this

decomposition to visualise the interplay of modules. We expect the new method to replace flux variability analysis in the pipelines for metabolic networks.

Integrated network analysis

Data on molecular interactions is increasing at a tremendous pace. Since biological functionality primarily operates at the network level, there is a clear need for topology-aware comparison methods. We developed one such method for global network alignment that is fast and robust and can flexibly deal with various scoring schemes taking both node-to-node correspondences as well as network topologies into account [18]. We exploited that network alignment is a special case of the well-studied quadratic assignment problem (QAP). We focused on sparse network alignment, where each node can be mapped only to a typically small subset of nodes in the other network. This corresponds to a QAP instance with a symmetric and sparse weight matrix. We obtained strong upper and lower bounds for the problem by improving a Lagrangian relaxation approach and introduce the open source software tool NATALIE 2.0, a publicly available implementation of our method (https://github.com/ls-cwi/natalie). In an extensive computational study on protein interaction networks for six different species, we found that our new method outperforms alternative established and recent state-of-the-art methods.

Integrative network analysis methods provide robust interpretations of differential high-throughput molecular profile measurements. They are often used in a biomedical context-to generate novel hypotheses about the underlying cellular processes or to derive biomarkers for classification and subtyping. The underlying molecular profiles are frequently measured and validated on animal or cellular models. Therefore the results are not immediately transferable to human. In particular, this is also the case in a study of the recently discovered interleukin-17 producing helper T cells (Th17), which are fundamental for anti-microbial immunity but also known to contribute to autoimmune diseases. We proposed a mathematical model for finding active subnetwork modules that are conserved between two species [19]. These are sets of genes, one for each species, which (1) induce a connected subnetwork in a species-specific interaction network, (2) show overall differential behaviour and (iii) contain a large number of orthologous genes. We proposed a flexible notion of conservation, which turns out to be crucial for the quality of the resulting modules in terms of biological interpretability. We developed an algorithm that finds provably optimal or near-optimal conserved active modules in our model. We applied our algorithm to understand the mechanisms underlying Th17 T cell differentiation in both mouse and human. As a main biological result, we found that the key regulation of Th17 differentiation is conserved between human and mouse.

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

Computationally investigating co-phylogenetic reconstructions and co-evolution

Despite an increasingly vast literature on co-phylogenetic reconstructions for studying host-symbiont associations, understanding the common evolutionary history of such systems remains a problem that is far from being solved. Most algorithms for host-symbiont reconciliation use an event-based model, where the events include in general (a subset of) co-speciation, duplication, loss, and host-switch. All known parsimonious event-based methods then assign a cost to each type of event in order to find a reconstruction of minimum cost. The main problem with this approach is that the cost of the events strongly influences the reconciliation obtained.To deal with this problem, we developed an algorithm, called COALA, for estimating the frequency of the events based on an approximate Bayesian computation approach [4]. The benefits of this method are twofold: (1) it provides more confidence in the set of costs to be used in a reconciliation, and (2) it allows estimation of the frequency of the events in cases where the dataset consists of trees with a large number of taxa. We evaluated our method on simulated and on biological datasets. We showed that in both cases, for the same pair of host and parasite trees, different sets of frequencies for the events lead to equally probable solutions. Moreover, often these solutions differ greatly in terms of the number of inferred events. It appears crucial to take this into account before attempting any further biological interpretation of such reconciliations. More generally, we also showed that the set of frequencies can vary widely depending on the input host and parasite trees. Indiscriminately applying a standard vector of costs may thus not be a good strategy. This work had been indicated as submitted in 2014.

Once such a cost vector has been inferred, one can proceed analysing the possible co-evolution of host-symbiont associations, phylogenetic tree reconciliation is the approach of choice for investigating the co-evolution of sets of organisms such as hosts and parasites. It consists in a mapping between the parasite tree and the host tree using event-based maximum parsimony. Given a cost model for the events, many optimal reconciliations are however possible. Only two algorithms existed that attempted such enumeration; in one case not all possible solutions are produced while in the other not all cost vectors are currently handled. We developed a polynomial-delay algorithm, EUCALYPT, for enumerating all optimal reconciliations that address these two issues [15]. We showed that in general many solutions exist. We gave an example where, for two pairs of host-parasite trees having each less than 41 leaves, the number of solutions is 5120, even when only time-feasible ones are kept. To facilitate their interpretation, those solutions are also classified in terms of how many of each event they contain. The number of different classes of solutions may thus be notably smaller than the number of solutions, yet they may remain high enough, in particular for the cases where losses have cost 0. In fact, depending on the cost vector, both numbers of solutions and of classes thereof may increase considerably (for the same instance, to respectively 4080384 and 275). To further deal with this problem, we introduced and analysed a restricted version where host-switches are allowed to happen only between species that are within some fixed distance along the host tree. This restriction allowed us to reduce the number of time-feasible solutions while preserving the same optimal cost, as well as to find time-feasible solutions with a cost close to the optimal in the cases where no time-feasible solution is found. This work had been indicated as submitted in 2014.

Evolution and metabolic complementation of organisms leaving inside the cells of another (endosymbionts)

Insect cells host many endosymbiotic bacteria, which are in general classified according to their importance for the host: "primary" symbionts are by definition mandatory and synthesise essential nutrients for the insects that feed on poor or unbalanced food sources, while "secondary" symbionts are optional and use mutualistic strategies and/or manipulation of reproduction to invade and persist within insect populations.

Hamiltonella defensa is a secondary endosymbiont that established two distinct associations with phloemophagous insects. In aphids, it protects the host against parasitoid attacks. Its ability to infect many host tissues, notably the hemolymph, could promote its contact with parasitoid eggs. Despite this protective phenotype, the high costs associated with its presence within the host prevent its fixation in the population. In the whitefly *Bemisia tabaci* however, this symbiont is found only in cells specialised in hosting endosymbionts, the bacteriocytes. In these cells, it cohabits with other symbiotic species, such as the primary symbiont *Portiera aleyrodidarum*, a proximity that favours potential exchanges between the two symbionts. It is fixed in populations of *B. tabaci*, which suggests an important role for the consortium, probably nutritious.

We studied the specificities of each of these systems [27]. First, in the bacteriocytes of *B. tabaci*, we identified a partitioning of the synthetic capacities of two endosymbionts, *H. defensa* and *P. aleyrodidarum*, in addition to a potential metabolic complementation between the symbionts and their host for the synthesis of essential amino acids. We proposed a key nutritive role for *H. defensa*, which would indicate a transition to a mandatory status in relation to the host and would explain its fixation in the population.

We also focused on the genomic evolution of the genus *Hamiltonella*, by comparing the strains infecting *B. tabaci* with a strain infecting the aphids [29]. We highlighted the specialisation of the symbionts to their hosts, and found that the genomes of the endosymbionts reflected their respective ecology. The aphid strain thus possesses many virulence factors and is associated with two partners, a bacteriophage and a recombination plasmid. These systems, inactive in the symbiont of *B. tabaci*, are directly related to the protection against and arms race with parasitoids. Conversely, the presumed avirulence of whitefly endosymbionts is consistent with their nutritional phenotype and a transition to a mandatory status to the host.

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

insight into the evolution of the species *H. defensa*, since the last ancestor to the present species, by establishing a link between bacterial.

These results were part of the PhD of Pierre-Antoine Rollat-Farnier, co-supervised by Laurence Mouton (LBBE, UMR5558), Marie-France Sagot (Inria and LBBE, UMR5558) and Fabrice Vavre (LBBE, UMR5558) and defended on November 24th, 2014. The results had been indicated as submitted in 2014.

Insights on the virulence of swine respiratory tract mycoplasmas through genome-scale metabolic modelling

The respiratory tract of swines is colonised by several bacteria among which are three *Mycoplasma* species: *Mycoplasma flocculare*, *Mycoplasma hyopneumoniae* and *Mycoplasma hyorhinis*. While colonisation by *M. flocculare* was shown to be virtually asymptomatic, *M. hyopneumoniae* is known to be the causative agent of enzootic pneumonia and *M. hyorhinis* to be present in cases of pneumonia, polyserositis and arthritis. Nonetheless, the elevated genomic resemblance among these three mycoplasmas combined with their different levels of pathogenicity is an indication that they have unknown mechanisms of virulence and differential expression. We performed whole-genome metabolic network reconstructions for these three mycoplasmas and were able to show that overall they have similar metabolic capabilities. The metabolic differences that were observed include a wider range of carbohydrate uptake in *M. hyorhinis*, which in turn may also explain why this species is a widely known contaminant in cell cultures. Moreover, the myo-inositol catabolism is exclusive to *M. hyopneumoniae* and may be an important trait for virulence. However, the most important difference seems to be related to glycerol conversion to dihydroxyacetone-phosphate, which produces toxic hydrogen peroxide. This activity, missing only in *M. flocculare*, may be directly involved in cytotoxicity, as already been described for two lung pathogenic mycoplasmas, namely *Mycoplasma pneumoniae* in human and *Mycoplasma mycoides* subsp. mycoides in ruminants. Metabolomic data suggest that even though these mycoplasmas are extremely similar in terms of their genome and metabolism, different products and reaction rates may be the result of differential expression in each of them. We were able to infer from the reconstructed networks that the lack of pathogenicity of *M. flocculare* if compared to the highly pathogenic *M. hyopneumoniae* may be related to its incapacity to produce cytotoxic hydrogen peroxide. Moreover, the ability of *M. hyorhinis* to grow in diverse sites and even in different hosts may be a reflection of its enhanced and wider carbohydrate uptake. Altogether, the metabolic differences highlighted in silico and in vitro provide important insights to the different levels of pathogenicity observed in each of the studied species.

These results were part of the PhD of Mariana Galvão Ferrarini, co-supervised by Arnaldo Zaha (Federal University of Rio Grande do Sul and Marie-France Sagot (Inria and LBBE, UMR5558). and defended on December 10th, 2015. These results have been submitted to a journal. The PhD manuscript will be made available in HAL in early 2016.

6.5. Cross-fertilising different computational approaches

Tree matching

We considered the following problem related to tree matching, that we called the Tree-Constrained Bipartite Matching problem. Given a bipartite graph $G = (V_1, V_2, E)$ with edge weights $w : E \to R^+$, a rooted tree T_1 on the set V_1 and a rooted tree T_2 on the set V_2, find a maximum weight matching M in G, such that none of the matched nodes is an ancestor of another matched node in either of the trees [8]. This generalisation of the classical bipartite matching problem appears, for example, in the computational analysis of live cell video data. We showed that the problem is APX-hard and thus, unless P = NP, disproved a previous claim that it is solvable in polynomial time. Furthermore, we gave a 2-approximation algorithm based on a combination of the local ratio technique and a careful use of the structure of basic feasible solutions of a natural LP-relaxation, which we also show to have an integrality gap of $2 - o(1)$. We then considered a natural generalisation of the problem, where trees are replaced by partially ordered sets (posets). We showed that the local ratio technique gives a $2k\sigma$- approximation for the k-dimensional matching generalisation of the problem, in which the maximum number of incomparable elements below (or above) any given element in each poset is bounded by σ. We finally gave an almost matching integrality gap example, and an inapproximability result showing that the dependence on σ is most likely unavoidable.

Graph measures

We proposed a new algorithm that computes the radius and the diameter of a weakly connected digraph $G = (V, E)$, by finding bounds through heuristics and improving them until they are validated [5]. Although the worst-case running time is $O(|V||E|)$, we experimentally showed that it performs much better in the case of real-world networks, finding the radius and diameter values after 10-100 BFSs instead of $|V|$ BFSs (independently of the value of |V|), and thus having running time $O(|E|)$ in practice. As far as we know, this is the first algorithm able to compute the diameter of weakly connected digraphs, apart from the naive algorithm, which runs in time $O(|V||E|)$ performing a BFS from each node. In the particular cases of strongly connected directed or connected undirected graphs, we compared our algorithm with known approaches by performing experiments on a dataset composed by several real-world networks of different kinds. These experiments showed that, despite its generality, the new algorithm outperforms all previous methods, both in the radius and in the diameter computation, both in the directed and in the undirected case, both in average running time and in robustness. Finally, as an application example, we used the new algorithm to determine the solvability over time of the "Six Degrees of Kevin Bacon" game, and of the "Six Degrees of Wikipedia" game. As a consequence, we computed for the first time the exact value of the radius and the diameter of the whole Wikipedia digraph.

The closeness and the betweenness centralities are two well-known measures of importance of a vertex within a given complex network. Having high closeness or betweenness centrality can have positive impact on the vertex itself: hence, we considered the problem of determining how much a vertex can increase its centrality by creating a limited amount of new edges incident to it [40]. We first proved that this problem does not admit a polynomial-time approximation scheme (unless P=NP), and we then proposed a simple greedy approximation algorithm (with an almost tight approximation ratio), whose performance is then tested on synthetic graphs and real-world networks.

The (Gromov) hyperbolicity is a topological property of a graph, which has been recently applied in several different contexts, such as the design of routing schemes, network security, computational biology, the analysis of graph algorithms, and the classification of complex networks. Computing the hyperbolicity of a graph can be very time consuming: indeed, the best available algorithm has running-time $O(n^{3.69})$, which is clearly prohibitive for big graphs. We provided a new and more efficient algorithm: although its worst-case complexity is $O(n^4)$, in practice it is much faster, allowing, for the first time, the computation of the hyperbolicity of graphs with up to 200,000 nodes [36]. We experimentally showed that the new algorithm drastically outperforms the best previously available algorithms, by analyzing a big dataset of real-world networks. Finally, we applied the new algorithm to compute the hyperbolicity of random graphs generated with the Erdös-Renyi model, the Chung-Lu model, and the Configuration Model.

Hypergraph problems

It had been previously proved independently and with different techniques that there exists an incremental output polynomial algorithm for the enumeration of the minimal edge dominating sets in graphs, *i.e.*, minimal dominating sets in line graphs. We provided the first polynomial delay and polynomial space algorithm for the problem [42]. We proposed a new technique to enlarge the applicability of Berge's algorithm that is based on skipping hard parts of the enumeration by introducing a new search strategy. The new search strategy is given by a strong use of the structure of line graphs.

We also studied some average properties of hypergraphs and the average complexity of algorithms applied to hypergraphs under different probabilistic models [14]. Our approach is both theoretical and experimental since our goal is to obtain a random model that is able to capture the real-data complexity. Starting from a model that generalizes the Erdös-Renyi model and we obtain asymptotic estimations on the average number of transversals, irredundants and minimal transversals in a random hypergraph. We use those results to obtain an upper bound on the average complexity of algorithms to generate the minimal transversals of a hypergraph. Then we make our random model more complex in order to bring it closer to real-data and identify cases where the average number of minimal transversals is at most polynomial, quasi-polynomial or exponential.

The hypergraph transversal problem has been intensively studied, both from a theoretical and a practical point of view. In particular, its incremental complexity is known to be quasi-polynomial in general and polynomial for bounded hypergraphs. Recent applications in computational biology however require to solve a generalisation of this problem, that we call bi-objective transversal problem. The instance is in this case composed of a pair of hypergraphs (A, B), and the aim is to enumerate minimal sets which hit all the hyperedges of A while intersecting a minimal set of hyperedges of B. We formalised this problem and related it to the enumeration of minimal hitting sets of bundles [32]. We showed cases when under degree or dimension contraints, these problems remain NP-hard, and gave a polynomial algorithm for the case when A has bounded dimension, by building a hypergraph whose transversals are exactly the hitting sets of bundles.

GENSCALE Project-Team

7. New Results

7.1. HTS data processing

7.1.1. Genome Analysis Tool Box Optimization

Participants: C. Deltel, P. Durand, E. Drezen, D. Lavenier, C. Lemaitre, P. Peterlongo, G. Rizk

Among the GATB library, the kmer-counting procedure is one of the most useful building block to speed-up development of new NGS tools. It is the first step of many NGS tools developed with GATB : Leon, Bloocoo, MindTheGap, DiscoSnp, Simka, TakeAbreak. This procedure has been optimized to be less limited by disks I/O. It relies on the use of kmer minimizers that help quickly partition the whole set of kmers into compact subsets. The kmer-counting procedure has also been re-worked to be more versatile, it is now able to count separately many input files and allows easy parametrization of the output, from simple kmer-count to the creation of custom user-defined kmer measures. At the core of the GATB library is also the manipulation and traversal of the de Bruijn Graph. The implementation has been optimized, leading to graph traversal twice fast as before. We introduced a new type of bloom filters, that are specially optimized for the manipulation of kmers. In these bloom filters neighboring kmers in the graph are close together in the bloom filter bit array, leading to better data locality, less cache misses and better overall performance [38].

7.1.2. NGS Data Compression

Participants: G. Benoit, E. Drezen, D. Lavenier, C. Lemaitre, G. Rizk

A novel reference-free method to compress data issued from high throughput sequencing technologies has been developed. Our approach, implemented in the LEON software, employs techniques derived from assembly principles. The method is based on a reference probabilistic de-Bruijn Graph, built de novo from the set of reads and stored in a Bloom filter. Each read is encoded as a path in this graph, by memorizing an anchoring kmer and a list of bifurcations. The same probabilistic de-Bruijn Graph is used to perform a lossy transformation of the quality scores allowing higher compression rates to be obtained without loosing pertinent information for downstream analyses. Leon was run on various real sequencing datasets (whole genome, exome, RNA-seq or metagenomics). In all cases, LEON showed higher overall compression ratios than state-of-the-art compression software. On a *C. elegans* whole genome sequencing dataset, LEON divided the original file size by more than 20 [16].

7.1.3. Multistep global optimization approach for the scaffolding problem

Participants: R. Andonov, D. Lavenier, I. Petrov

Our overall goal here is to address the computational hardness of the scaffolding problem by designing faster algorithms for global optimization that combine the branch-and-bound method which is able to find the global optimum but is usually slow for accuracy, with the use of massive parallelism and exploiting the special properties of the data–for scalability. A new two step scaffolding modeling strategy is in development. It tries to break the problem complexity by first solving a graph containing only large unitigs building something that can be compared to a trustworthy genomic frame. In our preliminary works [40] we developed integer programming optimization models that have been successfully applied on synthetic data generated from small chloroplast genomes. For computation we uses the Gurobi optimization solver.

7.1.4. Mapping reads on graph

Participants: A. Limasset, C. Lemaitre, P. Peterlongo

Next Generation Sequencing (NGS) has dramatically enhanced our ability to sequence genomes, but not to assemble them. In practice, many published genome sequences remain in the state of a large set of contigs. Although many subsequent analyses can be performed, one may ask whether mapping reads on the contigs is as informative as mapping them on the paths of the assembly graph. We proposed a formal definition of mapping on a de Bruijn graph, analysed the problem complexity which turned out to be NP-complete, and provided a practical solution. We proposed a pipeline called GGMAP (Greedy Graph MAPping). Its novelty is a procedure to map reads on branching paths of the graph, for which we designed a heuristic algorithm called BGREAT (de Bruijn Graph REAd mapping Tool). For the sake of efficiency, BGREAT rewrites a read sequence as a succession of unitigs sequences. GGMAP can map millions of reads per CPU hour on a de Bruijn graph built from a large set of human genomic reads. Surprisingly, results show that up to 22% more reads can be mapped on the graph but not on the contig set. Although mapping reads on a de Bruijn graph is a complex task, our proposal offers a practical solution combining efficiency with an improved mapping capacity compared to assembly-based mapping even for complex eukaryotic data [43].

7.1.5. Improving discoSnp features

Participants: C. Riou, C. Lemaitre, P. Peterlongo

NGS data enable to detect polymorphisms such as SNPs and indels. Their detection in NGS data is now a routine task. The main methods for their prediction usually need a reference genome. However, non-model organisms and highly divergent genomes such as in cancer studies are more and more investigated. The discoSnp tool has been successfully applied to predict isolated SNPs from raw read set(s) without the need of a reference genome. We improved discoSnp which became discoSnp++ [44]. DiscoSnp++ benefits from a new software design that reduces time and memory consumption, and from a new algorithmic design that detects all kinds of SNP and small indels, adds genotype information and outputs a VCF (Variant Calling Format) file. Moreover, when a reference genome may be used, discoSnp++ predictions are automatically mapped to this reference and the VCF file shows up location information of each prediction. This step also provides a way to filter out false predictions due to genomic repeats. Using discoSnp++ even when a reference is available has multiple advantages: it is several order of magnitude faster and uses much less memory. We are currently working in showing that it also provides better predictions than methods based on read mapping.

7.1.6. HLA genotyping

Participant: D. Lavenier

The human leukocyte antigen (HLA) system drives the regulation of the Human immune system. Genotyping the HLA genes involved in the immune system consists first in a deep sequencing of the HLA region. Next, a NGS analysis is performed to detects SNP variations from which correct haplotypes are computed. We have developed a fast method that outperforms standard approaches which, generally, require exhaustive database searches. Instead, the method extracts a few significant k-mers from all the haplotypes referenced in the HLA database. Each haplotype is then characterized by a small set of informative k-mers. By comparing these k-mer sets with the HLA sequencing data of a specific person, we can rapidly determine its HLA genotype.

7.1.7. Identification of long non-coding RNAs in insects genomes

Participant: F. Legeai

The development of high throughput sequencing technologies (HTS) has allowed researchers to better assess the complexity and diversity of the transcriptome. Among the many classes of non-coding RNAs (ncRNAs) that were identified during the last decade, long non-coding RNAs (lncRNAs) represent a diverse and numerous repertoire of important ncRNAs, reinforcing the view that they are of central importance to the cell machinery in all branches of life. Although lncRNAs have been involved in essential biological processes such as imprinting, gene regulation or dosage compensation especially in mammals, the repertoire of lncRNAs is poorly characterized for many non-model organisms [23]. In collaboration with the Institut de Génétique et de Développement de Rennes (IGDR) we participate in the development of a software for extracting long non coding RNA from high throughput data (https://github.com/tderrien/FEELnc).

7.1.8. Data-mining applied to GWAS

Participants D. Lavenier, Pham Hoang Son

Discriminative pattern mining methods are powerful techniques for discovering variant combinations related to diseases. The aim is to find a set of patterns that occur with disproportion frequency in case-control data sets, and a real challenge is to select a complete set of variant combinations that are biologically significant. There are various measurement methods for evaluating the discriminative power of individual combination in two-class data sets. Our research activity on this topic attempts to compare the statistical discriminative power measurements in genetic case-control data sets in order to evaluate the effectiveness of detecting variants associated with diseases.

7.2. Sequence comparison

7.2.1. Amplicon alignment

Participants: S. Brillet, C. Deltel, P. Durand, D. Lavenier, I. Petrov

Many metagenomics projects identify species by the studying16S-RNA sequences. This is mainly done by comparing the amplicons with 16S-RNA bacterial banks (amplicons are short fragments sequenced from very specific genome areas). As these sequences share a lot of similarities, immediate blast-like heuristics achieve poor performances. To speed up the process, we first select informative k-mers, from both the amplicon dataset and in the RNA16 bank (informative k-mers are defined as under represented k-mers). An index is built from this reduced set of k-mers and a "seed-and-extend" procedure is run. This strategy avoids many non-useful computation and accelerate the overall computation by two orders of magnitude. This new approach is currently implemented in the PLAST software (Regional KoriPlast2 project).

7.2.2. Metagenomics datasets comparison

Participants: G. Benoit, D. Lavenier, C. Lemaitre, P. Peterlongo, G. Rizk

We developp a new method, called Simka, to compare simultaneously numerous large metagenomics datasets. The method computes pairwise distances based on the amount of shared k-mers between datasets. The method scales to a large number of datasets thanks to an efficient kmer-counting step that processes all datasets simultaneoulsy. Additionnally, several distance definitions were implemented and compared, including some originating from the ecological domain. The method is currently applied to the TARA oceans project (more than 500 datasets) which aims at comparing worldwide sea water samples (ANR HydrGen project) [39].

7.3. Protein 3D structure

7.3.1. Discovering protein conformations by distance geometry

Participant: A. Mucherino

The distance geometry asks whether a simple weighted undirected graph G can be embedded in a Euclidean space having a predefined dimension $K > 0$, so that distances between pairs of embedded vertices are the same as the weights on graph edges. One of the most important applications of the distance geometry can be found in biology, where experimental techniques are able to find estimates of certain distances between atom pairs in molecules. Even if the scientific community is used to employ standardized techniques for the solution of this problem, which are essentially based on heuristic searches, we have recently shown that our combinatorial approach to this problem can be in fact employed for solving biological instances of the distance geometry [17]. This work is in collaboration with international people and researchers from the Pasteur Institut in Paris.

7.3.2. Discretization orders for distance geometry

Participant: A. Mucherino

The concept of discretization order is fundamental for the discretization of the distance geometry, i.e. for reducing the search space of a given distance geometry instance to a discrete (and finite) space. A discretization order is an order on the vertices of the graph G representing an instance of the distance geometry that is able to satisfy the discretization assumptions. Recent research was focused on the problem of finding, for a given distance geometry instance, a suitable discretization order that allows for its discretization [32]. The problem is tackled from a purely theoretical point of view in [33], while a special order for protein backbones was identified in [27] by creating a path on a "pseudo" de Bruijn graph. In [36], additional requirements are included during the search for a vertex order, in order to identify discretization orders that are also "optimal". In this work, we used Answer Set Programming (ASP) for identifying optimal partial orders that ensure the discretization of distance geometry instances related to proteins. This work is in collaboration with the Dyliss team, as well as international people.

7.3.3. Structure Similarity Detection

Participants: M. Le Boudic-jamin, R. Andonov

The most commonly used among the various measures of alignment similarity are the internal distances root mean squared deviation (RMSDd) and the coordinate root mean squared deviation (RMSDc) . In the paper [18] we introduce a novel approach to find similarities between protein structures. Our algorithm is both internal-distances based and Euclidean-coordinates based (i.e., it uses a rigid transformation to optimally superimpose the two structures). Resulting alignments are guaranteed to score well for both RMSDd and RMSDc , while remaining polynomial. We also replace the goal of finding the largest clique by the one of returning several very dense “near-clique” subgraphs. This choice is strongly justified by the observation that distinct solutions to the structural alignment problem that are close to the optimum are all equally viable from the biological perspective, and hence are all equally interesting from the computation standpoint. Our tool is suitable for detecting similar domains when comparing multi-domain proteins, as well to detect structural repetitions within a single protein and between related proteins [12].

7.3.4. Automatic Classification of Protein Structure

Participants: M. Le Boudic-jamin, R. Andonov

In this paper [15] we propose a new distance measure for comparing two protein structures based on their contact map representations . We show that our novel measure, which we refer to as the maximum contact map overlap (max-CMO) metric, satisfies all properties of a metric on the space of protein representations. Having a metric in that space allows one to avoid pairwise comparisons on the entire database and, thus, to significantly accelerate exploring the protein space compared to no-metric spaces. We show on a gold standard superfamily classification benchmark set of 6759 proteins that our exact k-nearest neighbor (k-NN) scheme classifies up to 224 out of 236 queries correctly and on a larger, extended version of the benchmark with 60 850 additional structures, up to 1361 out of 1369 queries. Our k-NN classification thus provides a promising approach for the automatic classification of protein structures based on flexible contact map overlap alignments.

7.3.5. Detection of structure repeats in proteins

Participant: M. Le Boudic-jamin, R. Andonov

Almost 25% of proteins contain internal repeats, these repeats may have a major role in the protein function. Furthermore some proteins actually are the same substructure repeated many times, these proteins are solenoids. However, very few protein repeats detection programs exist today. In the paper [29] we present a simple and efficient tool for discovering protein repeats. Our tool is based on protein fragment comparison and clique detection. We show that our tool is able to detect different levels of repetitions and to successfully identify protein tiles.

7.4. Parallelism

7.4.1. Processor in Memory

Participants: C. Deltel, D. Lavenier

The concept of PIM (Processor In Memory) aims to dispatch the computer power near the data. Together with the UPMEM company, which is currently developing a DRAM enhanced with computing units, we investigate the parallelization of several bioinformatics algorithms for this new types of memory. The first results show that blast-like algorithms or mapping algorithms can highly benefit of such memory. But the core algorithms must be revisited in order to better suite the PIM architecture.

7.4.2. Alignment search tools on cloud

Participants: S. Brillet, D. Lavenier, I. Petrov

PLAST is an alternative version of Blast to target intensive sequence comparison (bank-to-bank comparison). The multicore version offers a speed from 5 to 10 compared to Blast. In 2015, we deploy PLAST in the IFB cloud infra-structure (French Bioinformatics Institute) and demonstrate that an Hadoop implementation provides a very good scalability [34].

7.4.3. Bioinformatics Workflow

Participants: D. Lavenier, F. Moorews

Bioinformatics workflows play an important role in the development of new methodologies for analyzing sequencing data. Optimizing this activity brings the questions of how workflow can be efficiently captured and how technical tasks integration can be simplified. Thus, we define an expressive graphic worfklow language, adapted to the quick capture of workflows. This graphical input is then interpreted by a workflow engine based on a new model of computation with high performances obtained by the use of multiple levels of parallelism. A Model-Driven design approach is associated to facilitate the data parallelism generation and the production of suitable implementations for different execution contexts. In the case of the cloud model Container as a Service (CaaS), a workflow specification intrinsically re-executable and readily disseminatable has been developed. The adoption of this kind of model could lead to an acceleration of exchanges and a better availability of data analysis workflows [25] [31] [13].

7.4.4. Graph processing

Participants: D. Lavenier, R. Andonov

In the paper [20] we present a new approach for solving the all-pairs shortest-path (APSP) problem for planar graphs that exploits the massive on-chip parallelism available in today's Graphics Processing Units (GPUs). We describe two new algorithms based on our approach. Both algorithms use Floyd-Warshall method, have near optimal complexity in terms of the total number of operations, while their matrix-based structure is regular enough to allow for efficient parallel implementation on the GPUs. By applying a divide-and-conquer approach, we are able to make use of multi-node GPU clusters, resulting in more than an order of magnitude speedup over fastest known Dijkstra-based GPU implementation and a two-fold speedup over a parallel Dijkstra-based CPU implementation.

7.4.5. Analytical models and Optimization for GPUs

Participants: R. Andonov

In [28] we develop a methodology for modeling the energy efficiency of tiled nested-loop codes running on a graphics processing unit (GPU) and use it for energy efficiency optimization. We use the polyhedral model, and we assume that a highly optimized and parametrized version of a tiled nested – loop code, either written by an expert programmer or automatically produced by a polyhedral compilation tool – is given to us as an input. We then model the energy consumption as an analytical function of a set of parameters characterizing the software and the GPU hardware. Our approach develops analytical models based on (i) machine and architecture parameters, (ii) program size parameters as found in the polyhedral model and (iii) tiling parameters, such as those that are chosen by auto-or manual tuners. Our model therefore allows efficient optimization of the energy efficiency with respect to a set of parameters of interest.

7.5. Applications

7.5.1. CAMI: Critical Assessment of Metagenomic Interpretation

Participants: C. Deltel, D. Lavenier, C. Lemaitre, P. Peterlongo, G. Rizk

The interpretation of metagenomes relies on sophisticated computational approaches such as short read assembly, binning and taxonomic classification. All subsequent analyses can only be as meaningful as the outcome of these initial data processing methods [0]. The CAMI initiative aims to evaluate these methods independently, comprehensively and without bias. The goal is to supply users with exhaustive quantitative data about the performance of methods in many relevant scenarios. In 2015, we participate to CAMI within the "assembly" category using the Minia assembly pipeline. Results are provided here: https://data.cami-challenge.org/. For the medium challenge datasets, our assemblies are refered under the identifiers *goofy-wilson* and *fervent-blackwell*.

7.5.2. Assembly and Annotation of Arthropods Genomes

Participants: A. Gouin, F. Legeai, C. Lemaitre

Within a large international network of biologists, GenScale has contributed to various projects for identifying important components such as protein coding or non coding genes involved in the adaptation of major agricultural pests to their environment. We provided the assembly and the annotation of 4 new aphids, 3 parasitic wasps, and improved the assembly of 2 variants of fall army worm by removing unwanted sequences due to heterozygosity [41], [42]. Following specific agreement or policy, these new genomes and annotations are available for a restricted consortium or a large community through the BioInformatics platform for Agro-ecosystems Arthropods (http://bipaa.genouest.org/is). These results, and further analyses led to a better understanding of the biology, evolution and life history traits of *Spodoptera frugiperda* [19], the identiifcation and characterization of new genome of pea aphid symbionts [22] and the identification of differentially expressed genes in the sensory system of Sesamia nonagrioides [21].

7.5.3. Study of the rapeseed genome structure

Participants: D. Lavenier, C. Lemaitre, S. Letort, P. Peterlongo

In collaboration with IGEPP (Institut de Génétique, Environnement et Protection des Plantes), INRA, and through two national projects, PIA Rapsodyn and France-Génomique Polysuccess, we are involved in the genome analysis of several rapeseed varieties. The Rapsodyn project has the ambition to insure long-term competitiveness of the rapeseed production through improvement of the oil yield and reduction of nitrogen inputs during the crop cycle. Rapeseed varieties must thus be selected from genotypes that favor low nitrogen input. DiscoSNP++ is here used to locate new variants among the large panel of rapeseed varieties which have been sequenced during the project. The PolySuccess project aims to answer the following question: how a polyploid, such as the oilseed rape plant, becomes a new species? Oilseed rape (Brassica napus) beeing a natural hybrid between B.rapa and B.oleracea, different genomes of these three species have been sequenced to study their structures. The Minia assembly pipeline provides a fast way to generate contigs that are used for studying gene specificities.

[0] http://www.cami-challenge.org/

IBIS Project-Team

6. New Results

6.1. Inference of bacterial regulatory networks from reporter gene data

The use of fluorescent and luminescent reporter genes allows real-time monitoring of gene expression, both at the level of individual cells and cell populations (Section 3.2). In order to fully exploit this technology, we need methods to rapidly construct reporter genes, both on plasmids and on the chromosome, mathematical models to infer biologically relevant quantities from the primary data, and computer tools to achieve this in an efficient and user-friendly manner. For instance, in a typical microplate experiment, 96 cultures are followed in parallel, over several hours, resulting in 10,000-100,000 measurements of absorbance and fluorescence and luminescence intensities. Over the past few years, we put into place an experimental platform and data analysis software, notably the WELLREADER program (Section 5.4), to allow biologists to make the most out of the information contained in reporter gene expression data. An invited review on the analysis of fluorescent reporter gene data was published in the proceedings of the Third International Workshop on Hybrid Systems Biology (HSB 14) [25].

Valentin Zulkower, in the framework of his PhD thesis, has developed novel methods for the analysis of reporter gene data, based on the use of regularized linear inversion. This allows a range of estimation problems in the analysis of reporter gene data, notably the inference of growth rate, promoter activity, and protein concentration profiles, to be solved in a mathematically sound and practical manner. We have evaluated the validity of the approach using *in-silico* simulation studies, and observed that the methods are more robust and less biased than indirect approaches usually encountered in the experimental literature based on smoothing and subsequent processing of the primary data, like in WELLREADER. We have applied the methods to the analysis of fluorescent reporter gene data acquired in kinetic experiments with *Escherichia coli*. The methods were shown capable of reliably reconstructing time-course profiles of growth rate, promoter activity, and protein concentration from weak and noisy signals at low population volumes. Moreover, they captured critical features of those profiles, notably rapid changes in gene expression during growth transitions. The linear inversion methods have been implemented in the Python package WELLFARE, and integrated by Michel Page in the web application WELLINVERTER (Section 5.3). This work was presented at the major bioinformatics conference ISMB/ECCB 2015 and published in the special issue of *Bioinformatics* associated with the conference [24]. The Institut Français de Bioinformatique (IFB) accepted a proposal to extend WellInverter into a scalable and user-friendly web service providing a guaranteed quality of service, in terms of availability and response time. This web service will be deployed on the IFB platform and accompanied by extensive user documentation, online help, and a tutorial.

Over the years, the above tools have been used in several studies in IBIS directed at the experimental mapping of gene regulatory networks in *E. coli*. An example is the motility network of *E. coli*, studied by Diana Stefan in the context of her PhD thesis. The main thrust of this work lies in clarifying and solving methodological issues in the automated inference of quantitative models of gene regulatory networks from time-series gene expression data, also called reverse engineering in the bioinformatics literature. The application of existing reverse engineering methods is commonly based on implicit assumptions on the biological processes under study. First, the measurements of mRNA abundance obtained in transcriptomics experiments are taken to be representative of protein concentrations. Second, the observed changes in gene expression are assumed to be solely due to transcription factors and other specific regulators, while changes in the activity of the gene expression machinery and other global physiological effects are neglected. While convenient in practice, these assumptions are often not valid and bias the reverse engineering process. In her PhD thesis, Diana Stefan systematically investigated, using a combination of models and experiments, the importance of this bias and possible corrections. She measured with the help of fluorescent reporter genes the activity of genes involved in the FliA-FlgM module of the *E. coli* motility network. From these data, protein concentrations and global physiological effects were estimated by means of kinetic models of gene expression. The results indicate

that correcting for the bias of commonly-made assumptions improves the quality of the models inferred from the data. Moreover, it was shown by simulation that these improvements are expected to be even stronger for systems in which protein concentrations have longer half-lives and the activity of the gene expression machinery varies more strongly across conditions than in the FliA-FlgM module. The approach proposed in this study is broadly applicable when using time-series transcriptome data to learn about the structure and dynamics of regulatory networks. The paper describing the work was published in *PLoS Computational Biology* [23].

In addition to reporter gene data, a variety of other experimental data can be used for the mapping of gene regulatory networks. For example, using Chromatin Immunoprecipitation-sequencing (ChIP-seq) experiments, Stéphan Lacour and colleagues have identified a large number of target promoters of the sigma factor σ^S during the transition from exponential to stationary phase. Sigma factors are accesory subunits of RNA polymerase, allowing the recognition of specific promoter sequences by the transcriptional machinery, and σ^S is known to specifically accumulate in a variety of stress conditions. The study, published in *Scientific Reports* [21], has confirmed the importance of σ^S for redirecting RNA polymerase to promoters that drive the expression of genes necessary for the survival of *E. coli* after nutrient exhaustion. Furthermore, the results highlight the role of σ^S in the regulation of several noncoding RNAs.

6.2. Models of carbon metabolism in bacteria

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

The complexity of this integrated system, with regulatory mechanisms cutting across metabolism, gene expression, signaling and subject to global physical and physiological constraints, has motivated important modeling efforts over the past four decades, especially in the enterobacterium Escherichia coli. Different hypotheses concerning the dynamic functioning of the system have been explored by a variety of modeling approaches. In an article in *Trends in Microbiology* [19], which was initiated during the sabbatical of Andreas Kremling in Grenoble in 2013, we have reviewed these studies and summarized their contributions to the quantitative understanding of CCR, focusing on diauxic growth in *E. coli*. Moreover, we have proposed a highly simplified representation of diauxic growth that makes it possible to bring out the salient features of the models proposed in the literature and confront and compare the explanations they provide. In parallel, specific aspects of CCR, in particular a better understanding of the role of the signalling molecule cyclic adenosine monophosphate (cAMP) in the dynamic regulation of promoters during growth transitions in *E. coli*, have been studied in the context of the PhD thesis of Valentin Zulkower, using both models and experimental data.

Beside CCR and the multiple regulatory systems controlling the metabolism of *E. coli*, the involvement of post-transcriptional regulation is uncertain. The post-transcriptional factor CsrA is stated as being the only regulator essential for the use of glycolytic substrates, but its impact on the functioning of central carbon metabolism has not been demonstrated. In the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique INRA-Inria, the collaboration of Delphine Ropers, Muriel Cocaign-Bousquet and Brice Enjalbert from LISBP at INSA Toulouse has resulted in a multi-scale analysis of a wild-type strain and its isogenic mutant attenuated for CsrA. A variety of experimental data has been acquired for these two strains in relevant conditions, including growth parameters, gene expression levels, metabolite pools, enzyme

activities and metabolic fluxes. Data integration, metabolic flux analysis and regulation analysis revealed the pivotal role of post-transcriptional regulation for reshaping carbon metabolism. In particular, the work has shed light on *csrA* essentiality and has provided an explanation for the glucose-phosphate stress observed in the mutant strain. A paper summarizing the work has been submitted for publication in a microbiology journal.

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

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

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

In the context of the response of yeast cells to osmotic shocks, in collaboration with the LIFEWARE project-team and colleagues from Université Paris Descartes and University of Pavia (Italy), Eugenio Cinquemani has investigated the use of mixed effects-modelling and identification techniques to characterize individual cell dynamics in isogenic cell populations. Mixed-effects models are hierarchical models where parametric response profiles of individuals is subject to inter-individual parameter variability following a common population distribution. Starting from identification approaches in pharmacokinetics, we have developed and applied inference methods to microfluidics data, with a focus on the response of budding yeast to osmotic shocks. First results presented at conference in 2013 and the identification and validation work performed with Andres Gonzales, who visited IBIS for a few months in 2014 during his PhD at the University of Pavia, have been finalized into a journal article recently accepted for publication in *PLoS Computational Biology* [20].

Started with a study of the arabinose uptake dynamics in *E. coli*, work on identification and state estimation for single-cell intrinsic noise models of gene networks has focused on the reconstruction of promoter activity profiles from fluorescent reporter data. In the single-cell stochastic context, given population snapshots of fluorescence levels at subsequent experimental instants, the problem becomes that of inferring promoter activity statistics over a cell population such as mean, variance or even higher-order moments from analogous statistics of the reporter output. This nontrivial extension of the deterministic deconvolution of promoter activity from population-average data requires knowledge of the stochastic reporter dynamics and of the relation between promoter and fluorescence statistical moments. In two conference papers, Eugenio Cinquemani investigated identifiability and identification of the kinetic parameters of the stochastic reporter dynamics [28] and proposed parametric and nonparametric methods for the reconstruction of the desired promoter activity statistics [27], [28], demonstrating their effectiveness *in silico*. Further developments of these methods and application to experimental data for addressing relevant biological questions will be the subject of future journal publications.

In parallel, collaboration of Eugenio Cinquemani with Marianna Rapsomaniki, post-doctoral researcher at at IBM Zurich Research Lab (Switzerland), Zoi Lygerou at the University of Patras (Greece) and John Lygeros at ETH Zurich (Switzerland) has been devoted to the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments and the inference of kinetic parameters of protein dynamics in single

eukaryotic cells. As an alternative to current approximate analytical methods, we have explored inference methods based on simulation of biological processes in realistic environments at a particle level. We introduced and demonstrated a new method for the inference of kinetic parameters of protein dynamics, where a limited number of *in-silico* FRAP experiments is used to construct a mapping from FRAP recovery curves to the parameters sought. Parameter estimates from experimental data are then computed by applying the mapping to the observed recovery curves, at virtually no additional price for any number of experiments, along with the application of a bootstrap procedure for determining identifiability of the parameters and confidence intervals for their estimates. After validation on synthetic data, the method was successfully applied to the analysis of the nuclear proteins Cdt1, PCNA and GFPnls in mammalian cells, also sheding light on cell-to-cell variability of the protein kinetics. Method and results have been published in *Bioinformatics* this year [22].

6.4. Growth control in bacteria and biotechnological applications

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

A foundation result for growth control in bacteria was published in the journal *Molecular Systems Biology* this year [18]. In this publication, which is based on the PhD thesis of Jérôme Izard and post-doctoral work of Cindy Gomez Balderas, we describe an engineered *E. coli* strain where the transcription of a key component of the gene expression machinery, RNA polymerase, is under the control of an inducible promoter. By changing the inducer concentration in the medium, we can adjust the RNA polymerase concentration and thereby switch bacterial growth between zero and the maximal growth rate supported by the medium. We have shown that our synthetic growth switch functions in a medium-independent and reversible way, and we have provided evidence that the switching phenotype arises from the ultrasensitive response of the growth rate to the concentration of RNA polymerase. In parallel, Delphine Ropers in collaboration with Jean-Luc Gouzé and Stefano Casagrande of the BIOCORE team are developing a quantitative model of the gene expression machinery to account for this surprising observation.

The publication in *Molecular Systems Biology* also presents a biotechnological application of the growth switch in which both the wild-type *E. coli* strain and our modified strain are endowed with the capacity to produce glycerol when growing on glucose. Cells in which growth has been switched off continue to be metabolically active and harness the energy gain to produce glycerol at a twofold higher yield than in cells with natural control of RNA polymerase expression. Remarkably, without any further optimization, the improved yield is close to the theoretical maximum computed from a flux balance model of *E. coli* metabolism. The synthetic growth switch is thus a promising tool for gaining a better understanding of bacterial physiology and for applications in synthetic biology and biotechnology. We submitted a patent for such applications at the European Patent Office.

Whereas the synthetic growth switch has been designed for biotechnological purposes, the question can be asked how resource allocation is organized in wild-type strains that have naturally evolved. Recent work has shown that coarse-grained models of resource allocation can account for a number of empirical regularities relating the the macromolecular composition of the cell to the growth rate. Some of these models hypothesize control strategies enabling microorganisms to optimize growth. While these studies focus on steady-state growth, such conditions are rarely found in natural habitats, where microorganisms are continually challenged by environmental fluctuations. The aim of the PhD thesis of Nils Giordano is to extend the study of microbial growth strategies to dynamical environments, using a self-replicator model. In a recently submitted paper, we have formulated dynamical growth maximization as an optimal control problem that can be solved using Pontryagin's Maximum Principle. We compare this theoretical gold standard with different possible

implementations of growth control in bacterial cells. This study has been carried out in collaboration with Jean-Luc Gouzé and Francis Mairet of the BIOCORE project-team.

LIFEWARE Project-Team

7. New Results

7.1. Hybrid Simulation of Heterogeneous Biochemical Models in SBML

Participants: Katherine Chiang, François Fages, Sylvain Soliman.

Models of biochemical systems presented as a set of formal reaction rules can be interpreted in different for- malisms, most notably as either deterministic Ordinary Differential Equations, stochastic continuous-time Markov Chains, Petri nets or Boolean transition systems. While the formal composition of reaction systems can be syntactically defined as the (multiset) union of the reactions, the composition and simulation of models in different formalisms remains a largely open issue. In [5], we show that the combination of reaction rules and events, as already present in SBML, can be used in a non-standard way to define stochas- tic and boolean simulators and give meaning to the hybrid composition and simulation of heterogeneous models of biochemical processes. In particular, we show how two SBML reaction models can be composed into one hybrid continuous-stochastic SBML model through a high-level interface for composing reaction models and specifying their interpretation. Furthermore, we describe dynamic strategies for automatically partitioning reactions with stochastic or continuous interpretations according to dynamic criteria. The per- formances are then compared to static partitioning. The proposed approach is illustrated and evaluated on several examples, including the reconstructions of the hybrid model of the mammalian cell cycle regulation of Singhania et al. as the composition of a Boolean model of cell cycle phase transitions with a continuous model of cyclin activation, the hybrid stochastic-continuous models of bacteriophage T7 infection of Alfonsi et al., and the bacteriophage λ model of Goutsias, showing the gain in both accuracy and simulation time of the dynamic partitioning strategy.

7.2. Theoretical and Practical Complexities of Enumerating Minimal Siphons in Petri Nets

Participants: François Fages, Thierry Martinez, Sylvain Soliman.

Petri nets are a simple formalism for modeling concurrent computa- tion. They are also an interesting tool for modeling and analysing biochemical reaction systems, bridging the gap between purely qualitative and quantita- tive models. Biological networks can indeed be complex, large, and with many unknown kinetic parameters, which makes the development of quantitative models difficult. In [9], we focus on the Petri net representation of biochemical reactions and on two structural properties of Petri nets, siphons and traps, that bring us information about the persistence of some molecular species, independently of the kinetics. We first study the theoretical time com- plexity of minimal siphon decision problems in general Petri nets, and present three new complexity results: first, we show that the existence of a siphon of a given cardinality is NP-complete; second, we prove that deciding the Siphon- Trap property is co-NP-complete; third, we prove that deciding the existence of a minimal siphon containing a given set of places, deciding the existence of a siphon of a given cardinality and deciding the Siphon-Trap property can be done in linear time in Petri nets of bounded tree-width. Then, we present a Boolean model of siphons and traps, and two methods for enumerating all minimal siphons and traps of a Petri net, by using a SAT solver and a Con- straint Logic Program (CLP) respectively. On a benchmark of 345 Petri nets of hundreds of places and transitions, extracted from biological models from the BioModels repository, as well as on a benchmark composed of 80 Petri nets from the Petriweb database of industrial processes, we show that both the SAT and CLP methods are overall faster by one or two orders of magnitude com- pared to the state-of-the-art algorithm from the Petri net community, and are in fact able to solve all the enumeration problems of our practical benchmarks. We investigate why these programs perform so well in practice, and provide some elements of explanation related to our theoretical complexity results.

7.3. Abstraction-based Parameter Synthesis for Multiaffine Systems

Participant: Grégory Batt.

Multiaffine hybrid automata (MHA) represent a powerful formalism to model complex dynamical systems. This formalism is particularly suited for the representation of biological systems which often exhibit highly non-linear behavior. In [10], we consider the problem of parameter identification for MHA. We present an abstraction of MHA based on linear hybrid automata, which can be analyzed by the SpaceEx model checker. This abstraction enables a precise handling of time-dependent properties. We demonstrate the potential of our approach on a model of a genetic regulatory network and a myocyte model.

7.4. Tropical Algebra Methods for Model Reduction

Participants: François Fages, Jonas Sénizergues, Sylvain Soliman.

Jonas Sénizergues has just started a PhD Thesis on the design of model reduction techniques for systems biology based on tropical algebra. The idea is to reason on the orders of magnitude of both kinetic parameters and molecular concentrations in order to determine particular regimes exhibiting fast-slow decomposition and amenable to model reductions. Such model reductions generalize the quasi steady-state (QSSA) and quasi-equilibrium (QE) criteria, and lead to hybrid automata for chaining the reduced dynamics. The solving of tropical equilibration equations rely on previous work using constraint programming techniques [0] with collaboration with Ovidier Radulescu (Univ. Montpellier) and Andreas Weber (University of Bonn, Germany).

7.5. Modeling the Effect of the Cell Cycle on the Circadian Clock in Mouse Embryonic Fibroblasts

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

Experimental observations have put in evidence autonomous self-sustained circadian oscillators in most mammalian cells, and proved the existence of molecular links between the circadian clock and the cell cycle. Several models have been elaborated to assess conditions of control of the cell cycle by the circadian clock, in particular through the regulation by clock genes of Wee1, an inhibitor of the mitosis promoting factor, responsible for a circadian gating of mitosis and cell division period doubling phenomena. However, recent studies in individual NIH3T3 fibroblasts have shown an unexpected acceleration of the circadian clock together with the cell cycle when the milieu is enriched in FBS, the absence of such acceleration in confluent cells, and the absence of any period doubling phenomena. In [14], we try to explain these observations by a possible entrainment of the circadian clock by the cell cycle through the inhibition of transcription during mitosis. We develop a differential model of that reverse coupling of the cell cycle and the circadian clock and investigate the conditions in which both cycles are mutually entrained. We use the mammalian circadian clock model of Relogio et al. and a simple model of the cell cycle by Qu et al. which focuses on the mitosis phase. We show that our coupled model is able to reproduce the main observations reported by Feillet et al. in individual fibroblast experiments and use it for making some predictions. In [17], those hypothesis are revised in order to reproduce the phase data in addition to the period data and make new predictions.

7.6. Effects of repeated osmotic stress on gene expression and growth: from cell-to-cell variability to cellular individuality in the budding yeast Saccharomyces cerevisiae

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

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

When shifted to a stressful environment, cells are capable of complex response and adaptations. Although the cellular response to a single stress has been studied in great detail, very little is known when it comes to dynamically fluctuating stressful environments. In addition, in the context of stress response, the role of cell-to-cell variability in cellular processes and more specifically in gene expression is still unclear.

In his PhD thesis [3], Artémis Llamosi uses a systems and synthetic biology approach to investigate osmotic stress in *S. cerevisiae* at the single cell level. Combining microfluidics, fluorescent microscopy and advanced image analysis, we are able to subject cells to precise fluctuating osmolarity and monitor single-cell temporal response.

While much previous research in gene expression heterogeneity focused on its stochastic aspect, we consider here long-lasting differences between cells regarding expression kinetics. Using population models and state-of-the-art statistical analysis, we manage to represent both population and single-cell dynamics in a single concise modelling framework. This quantitative approach capturing stable individuality in gene expression dynamics can define a form of non-genetic cellular identity.

To improve our understanding of the biological interpretation of such identity, we investigate the relation between single-cell specificities in their gene expression with their phenotype and micro-environment. We then take a lineage based perspective and find this form of identity to be partially inherited.

Understanding the evolutionary consequences of inheritable non-genetic cellular identity requires a better knowledge of the impact of fluctuating stress on cell proliferation. Dissecting quantitatively the consequences of repeated stress on cell-cycle and growth gives us an overview of the energetic and temporal consequences of repeated stress. At last, technical and theoretical developments needed to carry this investigation further are presented.

7.7. Resistance to anti-cancer drugs by non-mutational mechanisms: insights from a cell-based multi-scale model of TRAIL-induced apoptosis

Participants: Virgile Andréani, Grégory Batt, François Bertaux.

The fact that tumors can acquire drug resistance by non-mutational mechanisms is increasingly gaining attention (Sharma et al., 2010; Pisco et al., 2013; Flusberg et al., 2013). Stochastic fluctuations in cellular states of different resistance and proliferative potential could play an important role in such resistance acquisition. Thus, to enable a quantitative, molecular-level understanding of those phenomena, modeling approaches that go beyond traditional, deterministic kinetic models of biological pathways are required.

An interesting and well-studied example of non-mutational resistance acquisition concerns the response of cancer cells to the agent TRAIL, a selective inducer of apoptotic cell death. In a previous work (Bertaux et al., 2014), we have developped a single-cell model of TRAIL-induced apoptosis that accounts for (1) protein-protein signaling reactions linking TRAIL exposure to commitment to apoptosis, (2) stochastic gene expression for the proteins involved in this signaling and (3) protein degradation. Under parsimonious and realistic assumptions for parameter values, fractional killing and transient resistance acquisition readily emerged from model simulations. Those two properties relating to TRAIL resistance are observed in-vitro for many different cancer cell lines.

Here, again in collaboration with Dirk Drasdo and Szymon Stoma, we investigate the long-term response of proliferating cancer cell populations repeatedly treated by TRAIL by integrating our single-cell model of TRAIL-induced apoptosis into a multi-cellular simulation framework. We predict that the long-term killing efficiency of repeated treatments is strongly reduced compared to the first treatment. A detailed analysis showed that resistance acquisition is caused mainly by the targeted degradation of activated pro-apoptotic proteins and an imbalance between the turnover of pro- and anti- apoptotic proteins. In addition, simulations of the treatment of multi-cellular spheroids suggested that limited TRAIL penetration is unlikely to be a driving cause of resistance, but that it can exacerbate the impact of cell-intrinsic resistance acquisition.

7.8. Controlling a genetic inverted pendulum

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

The ability to routinely control complex genetic circuits in vivo and in real-time promises quantitative understanding of cellular processes of unprecedented precision, quality, and richness. With combined efforts in microfluidic design, microscope automation, image segmentation and analysis, and control theory, we propose a platform for real-time, single-cell, externalized in silico control and monitoring of genetic networks in E. coli. Computational framework and hardware are optimized for parallelizing the experiments and we use the platform to test and control an entire library of synthetic genetic circuits. The circuits we are trying to control are based on the genetic toggle switch, a foundational circuit in synthetic biology, which consists of two genes that repress each other. This genetic system features two stable equilibrium points where one of the genes has taken over. Our objective is to dynamically balance the circuit in single cells around a third, unstable equilibrium point at which no gene dominates and their mutual repression strengths are balanced. This is similar to the landmark problem in control theory of stabilizing an inverted pendulum. Although our work indicates that this real-time control approach can drive convoluted genetic networks towards states that are inaccessible to traditional genetic perturbations such as knock-outs and promoter induction, the a priori quantitative knowledge of the system required for achieving this control is minimal. We show that even a simple Proportional-Integral controller can stabilize the unstable point of the toggle switch in single cells. Finally, we demonstrate that manipulation, or even inversion, of the stability map of the network is possible, though counter intuitive, via the simultaneous stabilization of an entire population of toggle switch cells around their unstable point with a common dynamic input.

7.9. Synthesizing Configurable Biochemical Implementation of Linear Systems from Their Transfer Function Specifications

Participants: Katherine Chiang, François Fages, Sylvain Soliman.

The ability to engineer synthetic systems in the biochemical context is constantly being improved and has a profound societal impact. Linear system design is one of the most pervasive methods applied in control tasks, and its biochemical realization has been proposed by Oishi and Klavins and advanced further in recent years. However, several technical issues remain unsolved. Specifically, the design process is not fully automated from specification at the transfer function level, systems once designed often lack dynamic adaptivity to environmental changes, matching rate constants of reactions is not always possible, and implementation may be approximative and greatly deviate from the specifications. In [6], building upon the work of Oishi and Klavins, we overcome these issues by introducing a design flow that transforms a transfer-function specification of a linear system into a set of chemical reactions, whose input-output response precisely conforms to the specification. This system is implementable using the DNA strand displacement technique. The underlying configurability is embedded into primitive components and template modules, and thus the entire system is adaptive. Simulation of DNA strand displacement implementation confirmed the feasibility and superiority of the proposed synthesis flow.

7.10. Reconfigurable Neuromorphic Computation in Biochemical Systems

Participants: Katherine Chiang, François Fages.

Implementing application-specific computation and con- trol tasks within a biochemical system has been an important pursuit in synthetic biology. Most synthetic designs to date have focused on realiz- ing systems of fixed functions using specifically engineered components, thus lacking flexibility to adapt to uncertain and dynamically-changing environments. To remedy this limitation, an analog and modularized approach to realize reconfigurable neuromorphic computation with biochemical reactions is presented in [11]. We propose a biochemical neural network consisting of neuronal modules and interconnects that are both reconfigurable through external or internal control over the concentrations of certain molecular species. Case studies on classification and machine learning applications using the DNA strain displacement technology demonstrate the effectiveness of our design in both reconfiguration and autonomous adaptation.

7.11. Search by Constraint Propagation

Participants: François Fages, Thierry Martinez, Sylvain Soliman.

Constraint programming is traditionally presented as the combination of two components: a constraint model and a search procedure. In [13] we show that tree search procedures can be fully internalized in the constraint model with a fixed enumeration strategy. This approach has several advantages: 1) it makes search strategies declarative, and modeled as constraint satisfaction problems; 2) it makes it possible to express search strategies in existing front-end modeling languages supporting reified constraints without any extension; 3) it opens up constraint propagation algorithms to search constraints and to the implementation of novel search procedures based on constraint propagation. We illustrate this approach with a Horn clause extension of the MiniZinc modeling language and the modeling in this language of a variety of search procedures, including dynamic symmetry breaking procedures and limited discrepancy search, as constraint satisfaction problems. We show that this generality does not come with a significant overhead, and can in fact exhibit exponential speedups over procedural implementations, thanks to the propagation of the search constraints.

7.12. Execution models for Constraint Programming and Semantics Equivalence

Participants: François Fages, Thierry Martinez, Sylvain Soliman.

Logic programming and constraint programming are two declarative programming paradigms which rely on the identification of programs to theories, and programming to modeling. Execution models result from the operational interpretation of logical provability in logic programming, and of constraint propagation in constraint programming. However, the control of execution is crucial for the practicability of these schemes and extra-logical traits are thus added in those programming systems, with the classical slogans "logic program = logical theory + control", "constraint program = constraint model + search".

In his thesis [4], Thierry Martinez investigates execution models in which control and search can be shifted into the logic or the constraint model, while preserving the semantics. The three parts of the thesis correspond to the three semantics equivalence that are showed: the first between two committed-choice forward-chaining logic languages, the second between constraint logic programs and constraint models, and the third between guard semantics in angelic settings. Each of these equivalence is constructive in the sense that there exists an encoding that enables the compilation from one of the paradigm to the other.

7.13. On Translating MiniZinc Constraint Model into Fitness Functions: Application to Continuous Placement Problems.

Participants: François Fages, Thierry Martinez, Bao Duy Tran.

MiniZinc is a solver-independent constraint modeling lan- guage which is increasingly used in the constraint programming com- munity. It can be used to compare different solvers which are currently based on either constraint programming, Boolean satisfiability or mixed integer linear programming. In [12], we show how MiniZinc models can be compiled into fitness functions for evolutionary algorithms. More specifically, we describe the translation of FlatZinc models into fitness functions over the reals and their use in the Covariance Matrix Adapta- tion Evolution Strategy (CMA-ES) solver. We illustrate this approach, and evaluate it, on the modeling and solving of complex shape continuous placement problems.

MORPHEME Project-Team

6. New Results

6.1. Exact continuous penalties for ℓ_2-ℓ_0 minimization

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

We consider the following ℓ_0-regularized least squares problem

$$\widehat{\mathrm{x}} \in \arg\min_{\mathrm{x}\in\mathrm{R}^N} G_{\ell_0}(x) := \frac{1}{2}\|Ax - d\|^2 + \lambda\|\mathrm{x}\|_0\,, \tag{3}$$

where $A \in \mathrm{R}^{M\times N}$, $d \in \mathrm{R}^M$ represents the data and $\lambda > 0$ is an hyperparameter characterizing the trade-off between data fidelity and sparsity. This problem finds a wide range of applications in signal/image processing, learning and coding areas among many others. We proposed a unified framework for exact continuous penalties approximating the ℓ_0-norm. In other words, we are concerned by the design of a class of continuous relaxations of G_{ℓ_0}, preserving all its global minimizers, and for which any local minimal point is also one of the initial functional. Hence, we highlight five *necessary and sufficient* conditions on the continuous penalty approximating the ℓ_0-norm ensuring that the minimizers of the underlying continuous relaxation of G_{ℓ_0} are consistent with those of G_{ℓ_0}. However, some local minimizer of the relaxed functional are not minimizer of G_{ℓ_0} which is an interesting point for such highly non-convex functional. This work offers a new way to compare penalties approximating the ℓ_0-norm. Finally, it is worth noting that the CEL0 penalty [1], [14], [17] is the inferior limit of the obtained class of penalties and seems to be the best choice to do in order to obtained an equivalent continuous reformulation of (1).

6.2. Application of the Continuous Exact ℓ_0 relaxation to Channel and DOA sparse estimation problems

Participants: Emmanuel Soubies, Laure Blanc-Féraud.

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

This work is devoted to two classical sparse problems in array processing: Channel estimation and DOA (Direction Of Arrivals) estimation. We show how our results on ℓ_0 optimization [1], [14], [17] can be used, at the same computational cost, in order to obtain improvement in comparison with ℓ_1 optimization (usually used) for sparse estimation. Moreover, for the DOA case, we show that our analysis conducted in the Single Measurement Vector (SMV) case [1] can be generalized to the Multiple Measurement Vectors (MMV) case. In that case, the variable x is not a vector of $\mathbb{R}^N$ but a matrix of $\mathbb{R}^{N\times K}$ where N is the signal length and K the number of snapshots. Hence, one wants to apply sparsity to the rows of x, i.e. x must have a small number of nonzero rows, instead of applying the sparsity on all the components of x. This results in a row-structured sparsity penalty which is modelled using a mixed ℓ_2-ℓ_0 norm.

Finally, numerical experiments demonstrate the efficiency of the proposed approach compared to classical methods as ℓ_1 relaxation, Iterative Hard Thresholding or MUSIC algorithms and that it can reach the Cramer Rao Bound in some cases [4].

6.3. From TIRF microscope calibration to 3D biological reconstructions

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

This work is made in collaboration with Agata Radwanska, Ellen Van Obberghen-Schilling (iBV).

Total Internal Reflection Fluorescence microscopy (TIRF) is a method of choice to visualize membrane-substrate interactions. The principle of this device relies on the total internal reflection phenomenon generating an evanescent wave capable of producing a selective excitation of the dye molecules within a single layer of 100 to 500nm. The fast decay of the evanescent wave varies with respect to the incident angle of the light beam. Hence, intensity variations on TIRF images, occurring when changing the incident angle, are, in part, due to the axial positions of the observed structures. While a direct interpretation of Multi-Angle TIRF (MA-TIRF) images in terms of axial structure positions is not an easy task, reconstruction algorithms can be dedicated to compute a quantitative depth map with high axial resolution. However, the success of such reconstruction methods strongly depends on the system calibration.

We have proposed a pipeline for MA-TIRF calibration. Considering back focal plane (BFP) images of several solutions differing by their refractive indices, we validate the theoretical relation linking the tension applied to the galvanometric mirror (which controls the laser beam orientation) and the incident angle of the beam on the specimen. Then it is crucial to verify if the simple exponential decaying model of the evanescent wave is sufficient to describe our setup. To this end we propose to build a phantom sample (for which the geometry is known) using a large lens placed into a homogeneous fluorescent solution (Fig. 1 top). Based on a least square estimation, we showed a good agreement between the estimated slope of the lens (we assume the lens to be linear near the border) and the expected one up to 400nm depth (Fig. 1 bottom-left). To complete the validation procedure, we use a sample for which the structures of interest are labeled using two different fluorescent proteins sensitive to different wavelengths and emitting respectively green and red fluorescence. Then, using standard variational approaches, we obtain a co-localization of the reconstructed structures with a precision around 30-40nm (Fig. 1 bottom-middle) over at least 170nm depth showing the precision of the method. Finally, once this calibration step is achieved, we perform color-coded depth representation of 3D biological structures living in the vicinity of the cell membrane (Fig. 1 bottom-right). All these experiments have been made on an experimental TIRF system develloped at iBV lab in Valrose.

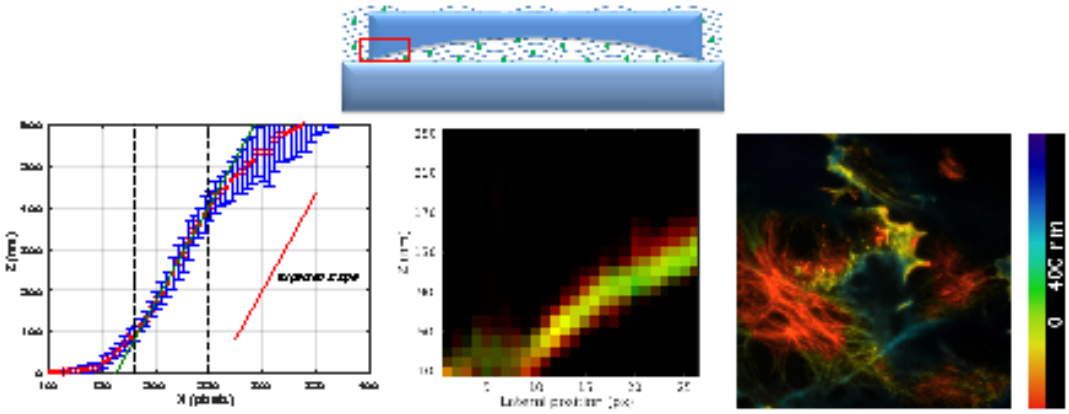


Figure 1. Top: Phantom sample constructed from a large lens and an homogeneous fluorescent solution. The red rectangle represents the observed region through the MA-TIRF setup. Bottom: estimated (green) and expected (red) slope of the lens within the red zone of the top figure (left), results of the co-localization experiment along a XY-line (middle) and a color-coded depth representation of a 3D biological reconstruction (right).

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

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

We present a gradient-based optimization method for the estimation of a specimen phase function from polychromatic DIC images. The method minimizes the sum of a nonlinear least-squares discrepancy measure and a smooth approximation of the total variation. A new formulation of the gradient and a recent updating rule for the choice of the step size are both exploited to reduce computational time. Numerical simulations on two computer-generated objects show significant improvements, both in efficiency and accuracy, with respect to a more standard choice of the step size.

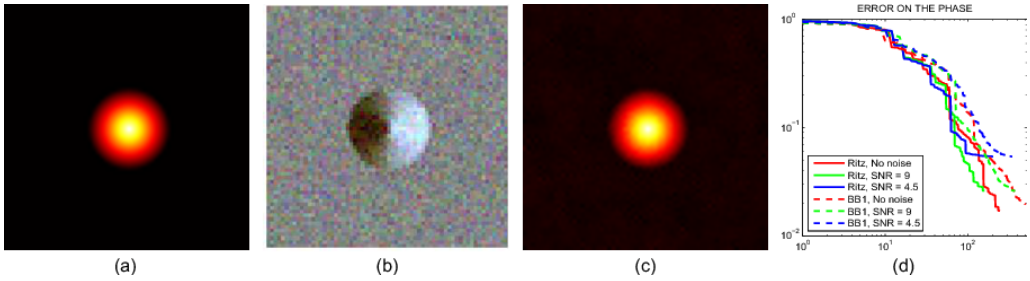


Figure 2. Data and results for the cone object. From left to right: true object, noisy DIC color image taken at angle $\tau_0 = 0°$ and SNR = 4.5, reconstructed phase and the relative error versus the number of iterations.

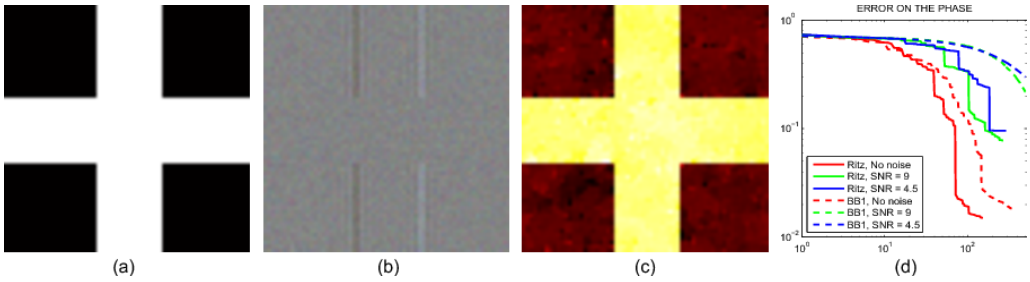


Figure 3. Data and results for the cross object. From left to right: true object, noisy DIC color image taken at angle $\tau_0 = 0°$ and SNR = 4.5, reconstructed phase with and the relative error versus the number of iterations.

6.5. Spatio-temporal registration of 3D microscopy image sequences of Arabidopsis floral meristems

Participants: Gaël Michelin, Grégoire Malandain.

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

The shoot apical meristem (SAM) is at the origin of all the plant above-ground organs (including stems, leaves and flowers) and is a biological object of interest for the understanding of plant morphogenesis. The quantification of tissue growth at a cellular level requires the analysis of 3D microscopic image sequences of developing meristems. To address inter-individual variability, it is also required to compare individuals. This obviously implies the ability to process inter-individual registration, i.e. to compute spatial and temporal correspondences between sequences from different meristems.

In [8], we propose a spatial registration method dedicated to microscopy floral meristem (FM) images, based on the registration of both the outer and the inner surfaces of the L1 layer (the epidermal cell layer). A given meristem (one timepoint) can be compared to a sequence (several timepoints) of an other meristem: the goodness-of-fit criterion allows to identify the best corresponding time-point in this sequence of an other individual, achieving the temporal registration (see figure 4). Since the morphological deformations remain important between successive images of a sequence, images interpolation between time-points is also performed in order to refine the sequence temporal resolution and thus to ensure a precise temporal registration.

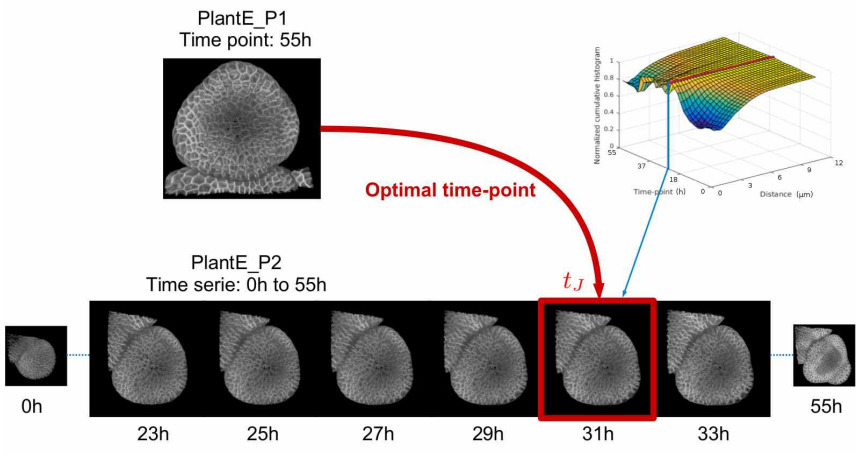


Figure 4. Inter-individual temporal registration result with 3D views of the registered meristem and the interpolated movie at several time-points.

6.6. Epidermal cell layer thickness variability in Arabidopsis floral meristems

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Yassin Refahi (Sainsbury Lab., University of Cambridge) and Jan Traas (ENS Lyon), within the Morphogenetics Inria Project Lab.

Flowers from the same species display a great robustness in their global shape and their developing stage can be theoretically identifiable to their size. The cells in epidermal (L1) and sub-epidermal (L2) layers of the floral meristem divide anticlinally, i.e. in a sideway fashion that ensures that L1 and L2 remain distinct. Thus a goodness-of-fit criterion on L1 and L2 layers is considered as an adequate registration quality measure in the inter-individual spatio-temporal registration framework developed in [8].

The aim of the present work is to investigate the variability of L1 layer thickness over development stages of an individual and between individuals. The study results may impact the way we process the inter-individual spatial registration. Therefore we measured the thickness distribution (histogram) of the L1 cells and we plotted the distribution of cells thickness (see figure 5) on images provided from three distinct floral meristems at acquisition time-points. Our results tend towards showing that L1 thickness increases over time non-uniformly, with a higher L1 thickness on sepals for advanced developing stages. We also observed an inter-individual thickness variability of about 15% for developing floral meristems at close developing stages. Future investigations will consist in taking a larger set of data to assess our first observations, in providing a biological interpretation of these observations and in using this knowledge to propose a refined spatial registration method.

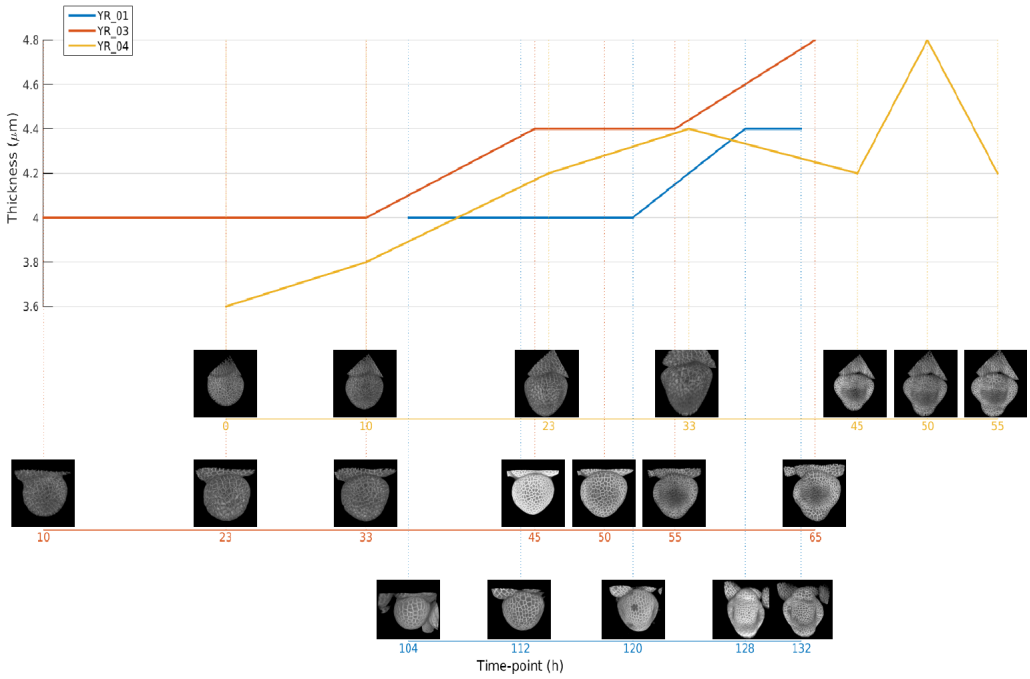


Figure 5. Epidermal cell layer thickness distribution over floral meristem development. Each row relates the measures at different developing times of a floral meristem.

6.7. Statistical Characterization, Modelling and Classification of Morphological Changes in imp Mutant Drosophila Gamma Neurons

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

In Drosophila brain, gamma neurons in the mushroom body are involved in higher functions such as olfactory learning and memory. During metamorphosis, they undergo remodelling after which they adopt their adult shape. Some mutations alter remodelling and therefore neuronal final morphology, causing behavioural dysfunctions. The RNA binding protein Imp, for example, was shown to control this remodelling process at least partly by regulating profilin expression. This work aims at precisely characterizing the morphological changes observed upon imp knockdown in order to further understand the role of this protein. We developed a methodological framework that consists in the selection of relevant morphological features (axon length and shape and branch length distribution and density), their modelling and parameter estimation. We thus perform a statistical comparison and a likelihood analysis to quantify similarities and differences between wild type and mutated neurons. The data was a set of 3D images showing a single neuron taken with a confocal microscope and provided by F. Besse group, IBV. The workflow from raw data to the likelihood analysis is summarized on figure6 . We show that imp mutant neurons can be classified into two phenotypic groups (called Imp L and Imp Sh) that differ in several morphological aspects. We also demonstrate that, although Imp L and wild-type neurons show similarities, branch length distribution is discriminant between these populations. Finally, we study biological samples in which Profilin was reintroduced in imp mutant neurons, and show that defects in main axon and branch lengths are partially suppressed.

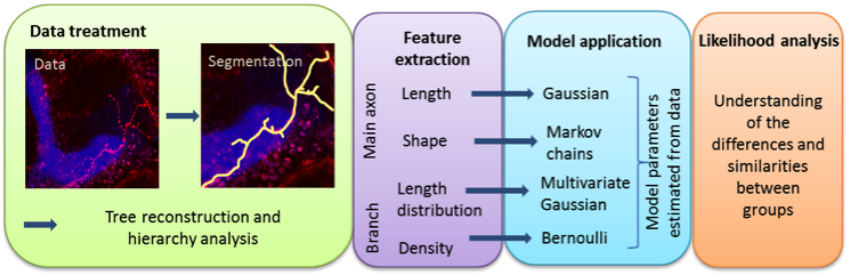


Figure 6. Summary of the workflow from raw data to the likelihood analysis.

6.8. Genome-wide search for factors that control the assembly of RNA granules

Participants: Wei Shen, Nicolas Cedilnik, Florence Besse, Xavier Descombes.

This work has been done in collaboration with Fabienne De Graeve from iBV

In vivo, mRNAs are packaged together with regulatory proteins into ribonucleoprotein particles (RNP) that control their fate and undergo extensive remodeling in response to developmental cues or environmental stresses. Cytoplasmic RNPs of different sizes, composition and regulatory properties have been described, including large macromolecular complexes such as P-bodies, stress granules, or germ cell granules. We aim at studying the different RNA granules distribution within the cytoplasm depending on genomic factors.

Before considering a spatial statistics analysis of the granules, it is necessary to detect them on confocal microscopy images of the cells. Therefore, we have studied a first pipeline for detecting these granules in confocal microscopy images. We have marked cells wih DAPI for detecting nuclei. These nuclei are then classified into "dead" or "alive" by a support vector machine (SVM) using intensity and shape criteria. In the

second step we consider GFP marked images to segment the cytoplasm and detect the granules within the cytoplasm. The cytoplasm segmentation is performed using an active contour whereas the granule detection is based on a marked point process model optimized by the multiple births and cut algorithm.

The full pipeline has been validated on a few samples from a pilot study. The next step will consists of a validation on the full study before considering a genome-wide screening.

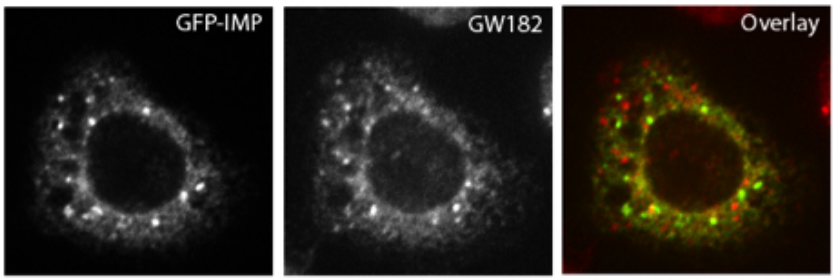


Figure 7. GFP-IMP particles are distinct from P-bodies. S2R+ cells expressing GFP-IMP fusions (left: green in the overlay) and stained with α-GW182 antibodies (middle: red in the overlay). GW182 is a well-described marker of P-bodies. Experiment performed in F. Besse lab at iBV (unpublished).

6.9. Cells detection using segmentation competition

Participants: Sen Wang, Emmanuel Soubies, Xavier Descombes.

Marked point processes have proved to be very efficient for segmenting a collection of objects from digital images. The multiple birth and death algorithm provides an optimization framework that allows reasonable time computation. This algorithm has to be embedded in a simluated annealing framework which involves parameters tuning (initial temperature and cooling scheme). This tedious task can be overcame considering a graph cut algorithm instead of the probabilistic death step. The algorithm then consists in successively adding new random objects in the configuration and selecting the most relevant using the graph cut algorithm. In the graph construction a node is associated to each object. In the orignal algorithm proposed by [21] the regularity condition imposed by the graph cut prevents to consider attractive interactions such as clustering or alignment constraints, which restricts the model to repulsive properties such as non overlap between objects. To overcome this restriction we have investigated new graph constructions by considering nodes defined by clusters of interacting objects. Different strategies have been compared to avoid being tracked in local minima defined by clusters while minimizing the number of required iterations. We have applied this new algorithm on different bioimagery problems such as axon extraction or cells detection (see figure 8).

6.10. Vesicles trajectory detection and analysis

Participant: Xavier Descombes.

This work has been done in collaboration with Maximilian Furthauer and Thomas Juan from iBV.

In many species, the left right asymetry of organs location is initiated in a ciliated cavity called Kupffer's vesicle in zebrafish. The cills beating induce a non symetrical flow in the cavity that can be studied by following the trajectory of exovesicles in the Kupffer's vesicle. The goal of this project is to automatically track these exovesicles and to perform a statistical analysis of theses trajectories in different conditions.

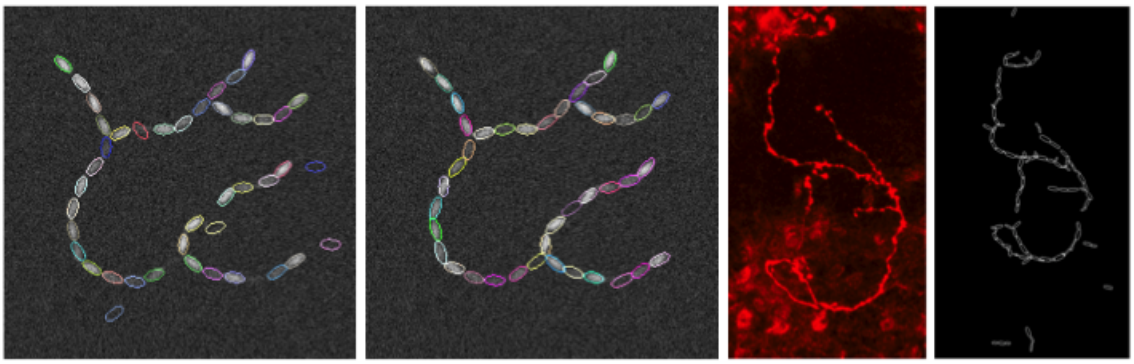


Figure 8. Axon detection on a synthetic image without (left) and with (middle left) attractive interactions and on real image (middle right) with attractive interactions (right).

We consider 2D time sequences of images. To extract the vesicles from the time sequence we first remove the background by substracting a local time mean. We then detect the cell border using an active contour computed on the spatial derivative of the images. The vesicles are then simply detected using a threshold followed by a morphological opening to remove the noise. The trajectory are finally obtained using a morphological closing in time. We aim at statistically comparing populations. In order to aggregate trajectories from several samples, we project the datasets into the same space using a continuous transformation of each cell into a reference disk. We thus project all the obtained trajectories from a given population into this disk. We compute the speed vector on each time point of each detected trajectory. To obtain a dense representation of the norm and the orientation of the vector speed in the reference disk, we extrapolate the obtained vector speed to a regular lattice with a Gaussian Markov random field. Finally, we obtain two spatial maps of respectively the norm and the orientation of the speed 9 .

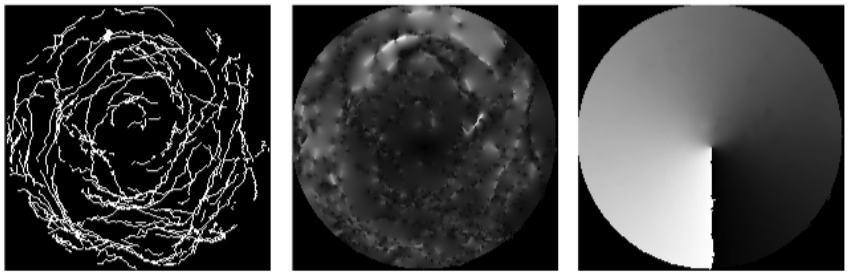


Figure 9. Exovesicles trajectories detected on the control population (left), the speed norm map (middle) and the angle map (left) showing an anticlockwise movement (black=0, white=2π).

6.11. Extraction and Analysis of the Vascular Network to Classify and Grade Kidney Tumors in Histological Imaging

Participants: Alexis Zubiolo, Eric Debreuve, Xavier Descombes.

This work is made in collaboration with Philippe Pognonec (Team TIRO, CEA/UNS), Damien Ambrosetti (Histopathology department, CHU Pasteur, Nice).

The renal carcinoma is the most frequent type of kidney cancer (between 90% and 95% of all cases). Twelve classes of carcinoma can be distinguished, among which the clear cell carcinoma (CCRCC) and the papillary carcinoma (PRCC) are the two most common (75% and 10% of the cases, respectively). After the carcinoma has been diagnosed, the tumor is ablated and prepared for histological examination (fixation, staining, slicing, observation with a microscope). Along with genetic tests and protein reactions, the histological study allows to classify and grade the tumor in order to make a prognosis and to take decisions for the subsequent patient treatment. Digital histology is a recent domain (routinely, histological slices are studied by MDs directly on the microscope). The pioneer works deal with the automatic analysis of cells. However, one crucial factor for carcinoma classification is the structure of the vascular network. Coarsely, CCRCC is characterized by a "fishnet" structure while the PRCC has a tree-like structure.

In this context, our goal was to extract the vascular network from a given histological slice, compute features of the underlying graph structure, and classify the tumor into CCRCC or PRCC based on these features. The histological images being huge (typically, 100k x 100k pixels), they must be split into tiles (with some overlap to ease the combination of results) and processed tile-wise. The first step is to combine the color channels so that the vessels are as highlighted as possible. Then, the vascular network is detected by a processing pipeline including tailored, Gabor-like filtering, thresholding, and extraction of the skeleton. Small gaps in the skeleton are filled and some pruning is performed. Finally, the skeleton is converted to a graph representation. Based on the medical interpretation procedure, we focused our analysis of the graph on the following elements: the number of terminal and junction nodes, and the terminal branches. We proposed to compute the ratio between the number of terminal nodes and the number of junctions (T/J ratio), and the average length of terminal branches. Both features seem to be adapted to classification, especially the T/J ratio which, on the fairly small database of cases we currently have, exhibits an average value 65% higher for PRCC.

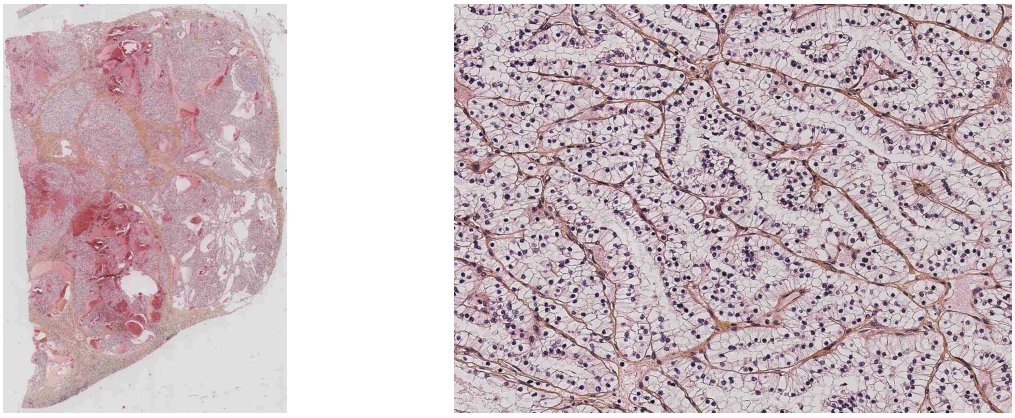


Figure 10. A histological slice through a kidney tumor: the whole slice (left) and a close-up (right).

PLEIADE Team

6. New Results

6.1. Inference of metabolic networks

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

We are particularly interested in incremental modeling of metabolic networks, where the target organism to be modeled is demonstrably similiar to other organisms for which whole or partial models are available. The other organisms are typically strains of the same species as the target, or species with a close phylogenetic relation to the target species. The similarity is measured genomically at different scales: sequence polymorphisms, expansions and contractions in conserved protein families, and genome rearrangements. We have defined and refined two complementary methods for inferring metabolic models for target species.

In the same way that comparative analysis of genomes and proteomes makes it possible to define protein families that summarize protein-coding genes into phyletic patterns [24], comparative analysis of related metabolic models makes it possible to define network generalizations [26] that factor families of reactions and metabolites into summary graphs that preserve stoichiometry. These summaries can be used for expert curation and visualization [5]. An online demonstration tool is made available at http://mimoza.bordeaux.inria.fr/.

Starting from an existing reference metabolic network and measures of similarity between the reference and the target organisms' genomes, we can use knowledge-based inference to rewrite the reference network based on these differences, and thus obtain a draft network for the metabolisms of the target organism [2]. This rewriting, formalized in the Pantograph system, can be extended to an abductive logic framework as described in Razanne Issa's thesis [19]. Current work aims at extending the Pantograph and ab-Pantograph frameworks to leverage reaction classifications obtained by network generalization.

6.2. Bio-medicine and biotechnology

Participants: Pascal Durrens [correspondant], David Sherman.

6.2.1. Genome assembly for bio-medicine

We performed the assembly of the *Clavispora lusitaniae* (aka *Candida lusitaniae*) genome. Yeasts from the genus *Candida* are opportunistic human pathogens in immunocompromised patients, linked to a high mortality rate. Although *Candida albicans* is the major pathogen, related species are more and more isolated, such as *Clavispora lusitaniae* which is responsible for candidaemia in newborn babies and in onco-hematology patients.

Even though the genome of a *Clavispora lusitaniae* strain (ATCC 42720), isolated from a patient, has already been sequenced by the Broad Institute, we achieved the genome assembly of the wild type reference strain (CBS 6936) as patient isolates tend to harbor genome modifications. The assembly was computed from Illumina reads with a coverage of 30X, using the MINIA assembler from Inria GENSCALE team. We also looked for single nucleotide polymorphismw (SNPs) in the reads coming from 3 hypovirulent mutants impaired in the beta oxidation metabolic pathway. Some detected SNPs are now under experimental validation and we are going to make a Genome Announcement for the CBS 6936 genome.

6.2.2. Transcriptome assembly for bio-technology

We carried out the assembly of transcriptomes from different tissues of the African oil palm tree *Elaeis guineensis*. The goal of this project is twofold: (i) Select the most relevant genes involved in oil synthesis in order to implement heterologous expression of some of these genes in a cultivated plant recipient such as tobacco. Preliminary results on heterologous expression of 2-3 key genes/factors ended in 15% of dry weight of oil synthesis. New expression technology allows for simultaneous expression of 15-20 genes. Identifiying the best candidates for co-expression will permit efficient heterologous oil synthesis. (ii) Identify the polymorphism of genes in a panel of 25 wild type isolates and of 5 production lineages of *Elaeis guineensis* in relation to the oil yield in different environmental conditions. In addition to a high variability of oil quantity (1-12 tons/ha/year), the relative amount of unsaturated fatty acids spans widely (15-55% dry weight) among the 30 *Elaeis guineensis* strains. Identification of polymorphisms will pave the way to genome-wide association genetics (GWAS) for the improvement of the oil resource.

In a first step, we produced assembled transcriptomes of ca. 300 million reads from 3 tissues (leaf, mesocarp, kernel) ocoming form a single tree, using state-of-the-art assembler TRINITY. Tuning of the software parameters was performed on the Inria PLAFRIM computation platform. About 20% of the assembled sequences revealed to be tissue-specific. Computation of the protein sequences deduced from the assembled transcripts gave a protein repertoire which was annotated using related sequences available in public databases. These transcript and protein sets will be used as a framwork in the polymorphism studies.

6.3. Biodiversity and ecology

Participants: Alain Franc [correspondant], Jean-Marc Frigerio, Philippe Chaumeil, Razanne Issa, Leyla Mirvakhabova.

Our main activity has been code development in the framework of a research project with ONEMA, and preparing future development. Code development has been fostered with the work of Razanne Issa (CDD ONEMA) in the last three months of 2015, and preparation of further development has been fostered by welcoming Leyla Mirvakhabova (L3, National Research University Higher School of Economics (NRU HSE), Moscow, mathematics) . Declic is a python library providing tools for analysing molecular data for biodiversity stydies. The main object is a distance matrix, from which one can either build a point cloud in a Euclidean space with distances between points as close as possible from distances between reads (Multidimensional scaling), or to build a graph with edges between reads when their distance is smaller than a given threshold. Meanwhile, the team has developed the network around the Galaxy server where an early verison of tools has been installed and made available, especially with SLU at Uppsala (Maria Khalert). Alain Franc has developed a collaboration with Olivier Coulaud and Pierre Blanchard (Hiepacs) for efficient computation of eigenvectors and eigenvalues of large, dense and symmetric matrices, needed for scaling in Multidimensional scaling.

The work of Razanne Issa has made it possible to extend the declic library in the direction of machine learning, by incorporating tools from support vector machines through library sklearn. This development will be pursued in 2016. The work of Leyla Mirvakhabova has permitted a first incursion into topological data analysis as a possible approach for studying the shape of point clouds produced by multidimensional scaling. The collaboration with NRU HSE on this topic will be pursued in 2016.

SERPICO Project-Team

7. New Results

7.1. Statistical aggregation methods for image denoising and estimation

Participants: Charles Kervrann, Frédéric Lavancier.

We have already proposed a general statistical aggregation method which combines image patches denoised with several commonly-used algorithms [10]. We showed that weakly denoised versions of the input image obtained with standard methods, can serve to compute an efficient patch-based aggregated estimator. In our approach, the Stein's Unbiased Risk Estimator (SURE) is used to evaluate each denoised candidate image patch and to compute the exponential weighted aggregation (EWA) estimator. This year, we adapted this framework (PEWA) to denoise images corrupted by mixed Gaussian-Poisson in 2D fluorescence image sequences.

In this range of work, we have also introduced in [24] a general method to combine estimators in order to produce a better estimate. From a theoretical point of view, we proved that this method is optimal in some sense. It is illustrated on standard statistical problems in parametric and semi-parametric models where the averaging estimator outperforms the initial estimators in most cases. As part of an on-going work, we are applying this method to improve patch-based image denoising algorithms.

Reference: [24]

Collaborators: Paul Rochet (Laboratoire de Mathématiques Jean Leray (LMJL), University of Nantes).

7.2. Image deconvolution algorithms for tagged-RNA and gene localization in live yeast

Participant: Charles Kervrann.

In fluorescence microscopy, the image quality is limited by out-of-focus blur and high noise. Traditionally, image deconvolution is needed to estimate a good quality version of the observed image. The result of deconvolution depends heavily on the choice of the regularization term. The regularization functional should be designed to remove noise while retaining the image structure. In this study, we investigated non quadratic regularization terms to preserve fine details of underlying structures and we studied appropriate optimization algorithms. The deconvolution method has been especially dedicated for 3D high-precision gene localization in cell nuclei [47]. For illustration, tagged gene (green marker) and tagged nucleoporins/nuclear periphery (red marker) are shown in Fig. 3 . A noisy and blurred image can affect the nuclear membrane estimation and gene detection and, consequently, the computed related distances.

Collaborators: Giovanni Petrazzuoli (Inserm U944, CNRS UMR 9212, Hôpital Saint-Louis, Paris), Catherine Dargemont (Inserm U944, CNRS UMR 9212, Hôpital Saint-Louis, Paris), Jean Salamero (UMR 144 CNRS-Institut Curie, PICT-IBiSA).

7.3. Estimation of the reference point giving the most uniform angular distribution

Participants: Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

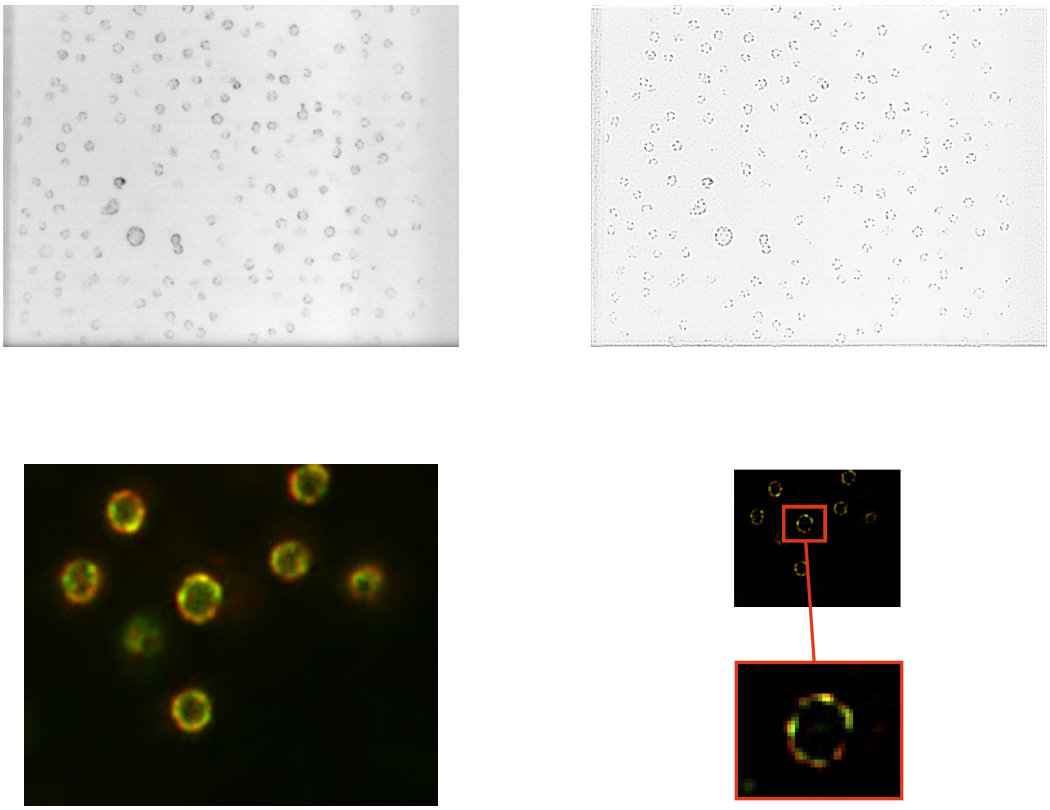


Figure 3. Deconvolution of 3D image depicting tagged gene and tagged nucleoporins/nuclear periphery. First row: deconvolution (right) of a tagged nucleoporin image (left). Second row: blurred image of tagged gene and nucleoporins (left) and zoom-in view of the deblurred image.

Rab11 proteins are trafficking from the Endosomal Recycling Compartment (ERC) to locations in the cell membrane where they eventually fuse. In this study, we assume that the Rab11 positive membranes are uniformly distributed around the ERC at the cell membrane. To test this hypothesis, we estimate the angular distribution of Rab11 positive membranes from several image sequences acquired with a TIRF microscope at the cell membrane level by considering all the points located in the cell as a reference point. We then compute the entropy of angular distribution for each point and estimate the ERC location as the reference point that gives the maximum entropy for the angular distribution (see Fig. 4). These results are very close to the ERC locations manually annotated by experts.

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

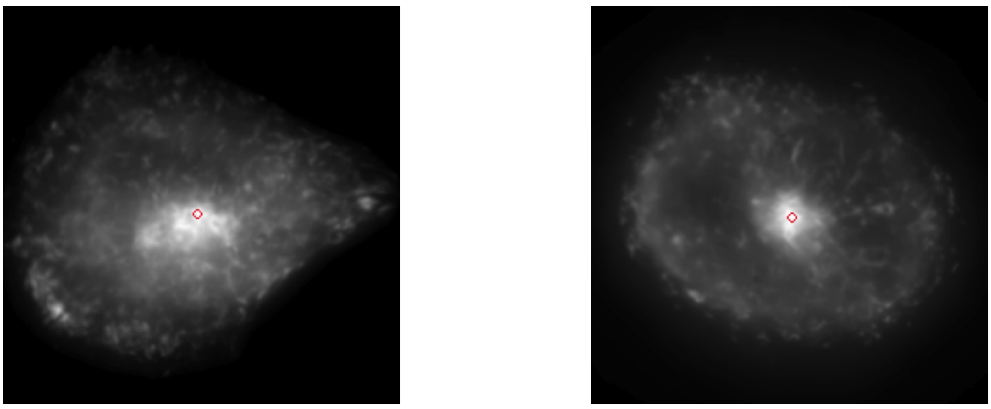


Figure 4. Estimation of reference points (red circles) for Rab11 traffic comparison, registration and quantification.

7.4. Modeling and estimation of protein release and diffusion in TIRFM

Participants: Antoine Basset, Charles Kervrann, Patrick Bouthemy.

We have pursued our work on membrane dynamics, still following a local approach in space and time. We have proposed a new model to account for the full behavior of cargo transmembrane proteins during the vesicle fusion to the plasma membrane at the end of the exocytosis process (see Fig. 5). It combines release and diffusion steps. The former is represented by an exponential decay to account for a continuous release of the proteins from the vesicle to the plasma membrane. We can relax the usual point source assumption, and we name our model the "Small-extent Source with Exponential Decay release" (SSED). An iterative minimization method is used to estimate simultaneously both biophysical parameters, i.e., the release rate and the diffusion coefficient, for every active vesicle detected in the total internal reference fluorescence microscopy (TIRFM) image sequence. Quantitative evaluation has demonstrated the efficiency of the method, which has also allowed us to exhibit differences in the behaviors of Transferrin receptor (TfR) and Langerin proteins.

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

7.5. Counting-based particle flux estimation for traffic analysis in live cell imaging

Participants: Thierry Pécot, Charles Kervrann.

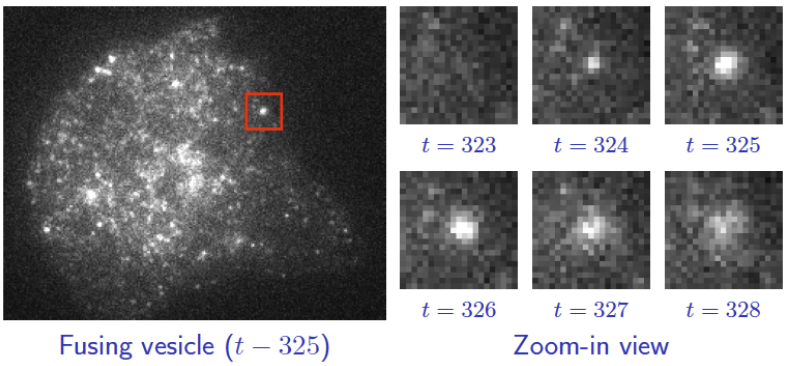


Figure 5. Left: Fusing vesicle (frame in red) in a TIRFM (UMR 144 CNRS-Institut Curie, PICT-IBiSA) sequence (frame 325, 50ms/frame). Right: Zoom-in view of the temporal evolution of the fusing vesicle.

In this study, we have proposed an original traffic analysis approach based on the counting of particles from frame to frame. Object tracking methods or optical flow methods are generally considered to analyze the dynamic contents of intracellular video-microscopy. The suggested method lies between these two well-known approaches. Instead of tracking each moving particle, we estimate fluxes of particles between predefined and adjacent regions. Our three-step counting-based approach is as follows:

- The cell is uniformly partitioned into fixed-size and fixed-shape regions.
- The moving particles are automatically detected using an appropriate algorithm.
- The fluxes are estimated with sparse constraints from an image pair at each time step from temporal variations of the number of particles in each region of the uniform tessellation. Except for some trivial cases, the flux estimation is actually an ill-posed problem and additional constraints are necessary to find the optimal solution.

The problem is formulated as the minimization of a global cost function and the approach allows us to process image sequences with a high number of particles and a high rate of particle appearances and disappearances. We studied the influence of object density, image partition scale, motion amplitude and particle appearances/disappearances in a large variety of simulations. The potential of the method has been demonstrated on real image sequences showing GFP-tagged Rab6 trafficking in confocal microscopy.

Reference: [26]

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

7.6. Tracking of astral microtubules at the cell cortex

Participants: Frédéric Logé-Munerel, Thierry Pécot, Antoine Basset, Charles Kervrann.

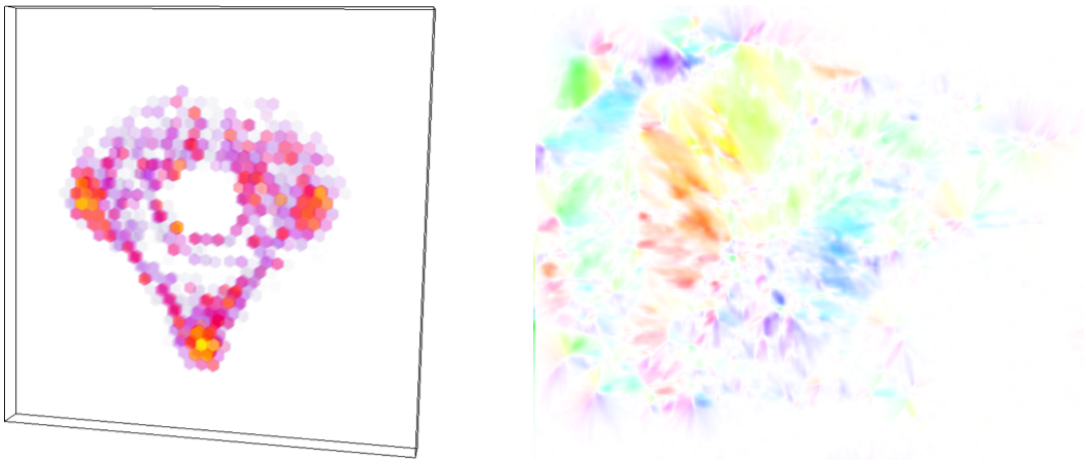


Figure 6. Left: Spatial distribution of GFP-tagged Rab6 vesicle numbers estimated when considering a regular 3D tesselation. Right: Estimation of directional/intensity flows of actin filaments for an image sequence acquired in TIRF microscopy (UMR 144 CNRS-Institut Curie, PICT-IBiSA).

In this study, we are currently interested in the influence of the mechanical properties of astral microtubules in the centering mechanisms of the mitotic spindle, giving it a robust positioning. In their previous studies, the CeDRE group (IGDR Rennes) identified two subpopulations of astral microtubules that either push or pull the cell cortex. To better understand these mechanisms, they acquired image sequences at the cortex level where astral microtubules extremities come to exert forces. In order to characterize the two subpopulations of astral microtubules during the mitosis in the unicellular embryos of C. Elegans life span, that is the period during which the microtubule is touching the cell cortex, has to be measured for every single microtubule. A short life span corresponds to a pulling force and a long life span corresponds to a pushing force. Detecting and tracking microtubules at the cell cortex has to be done to collect these measures. This year, F. Logé-Munerel (internship Master 1, supervisors: T. Pécot and C. Kervrann) improved the analysis workflow and calibrated the parameters of the algorithms to successfully track the microtubules. This workflow is composed of the ND-SAFIR denoising algorithm [4], the ATLAS detection algorithm [12] and the ASTRE tracking algorithm [56]. The experimental results are currently compared with results obtained by the CeDRE group using the U-track platform [50] (see Fig. 7).

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

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

Participants: Ancageorgiana Caranfil, Antoine Basset, Charles Kervrann.

The dynamics of the plasma membrane of the cell is not fully understood yet; one of the crucial aspects to clarify is the diffusion process during exocytosis. Several observation methods exist, including TIRFM (Total Internal Reflection Fluorescence Microscopy), that has successfully been used to determine the successive steps of exocytosis. However, computing characteristic values for plasma membrane dynamics is problematic, as the experimental conditions have a strong influence on the obtained data and a global model cannot be determined. The goal of this study was to build a correlation-like method to estimate local diffusion parameters in TIRFM images. Using a correlation approach similar to TICS (Temporal Image Correlation Spectroscopy) with an adapted local model, we have developed a novel correlation-based method to estimate the diffusion

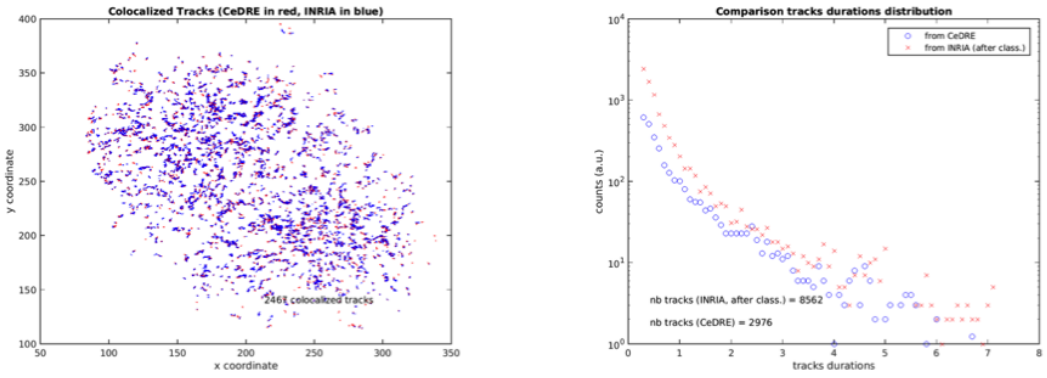


Figure 7.

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

coefficient for every diffusion event in TIRFM images. We turned the non-linear model of the TICS method into a linear one, and made it rely on less parameters than the other estimation methods. Results are excellent for sequences with a good signal-to-noise ratio (see Fig. 8); however, time and space dependencies are introduced with the presence of moderate-to-strong image noise. Although only synthetic images have been used so far, studies of real-life TIRFM images are forthcoming, along with refinements to make the method robust to noise.

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

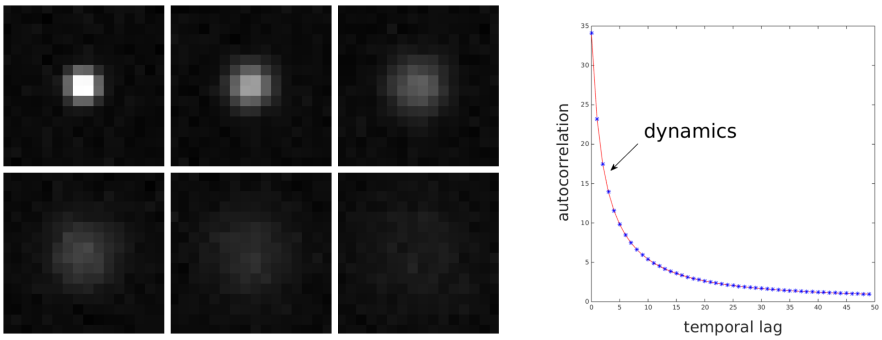


Figure 8. Left: first six instances of a TIRFM image sequence showing a diffusion event. Right: the correlation-based method is applied on the TIRFM sequence; both the computed values of the autocorrelation, for different values of the temporal lag, and the fitting function for these values are represented.

7.8. Co-localization between proteins : testing procedure and generative models

Participants: Frédéric Lavancier, Thierry Pécot, Charles Kervrann.

In the context of bioimaging, co-localization refers to the detection of emissions from two or more fluorescent molecules within the same pixel of the image. This approach enables to quantify the protein-protein interactions inside the cell, just at the resolution limit of the microscope. In statistics, this amounts to characterizing the joint spatial repartition and the spatial overlap between different fluorescent labels. An illustration of the co-localization of green (Langerin protein) and red (Rab11 GTPase protein) fluorescence is shown in Fig. 9 (the images were segmented by applying the ATLAS algorithm [12]). In our framework, the spatial repartition of proteins in the same cell is modeled by a union of random balls, possibly overlapping, and a Gibbs interaction is introduced to take into account the possible interaction between the two co-expressed proteins. A simulation algorithm is described and an inference procedure, based on the Takacs-Fiksel method, is proposed to estimate the interaction parameter. This estimation allows us to determine the presence of co-localization and to quantify the degree of interactions. On the other hand, this model can be used as a generator for synthesized images of co-localized proteins, in a view to assess testing procedures as the one explained below.

In an on-going project, we are developing a non-parametric testing procedure for co-localization. It is mainly based on the overlap area, corresponding to yellow spots as displayed in the right-hand side image of Fig. 9 . Our first experiments on synthesized images showed that our procedure is more powerful than all existing methods to detect co-localization. Moreover this testing procedure turns out to be robust to different shapes and sizes of objects segmented by any competitive algorithm.

Reference: [36]

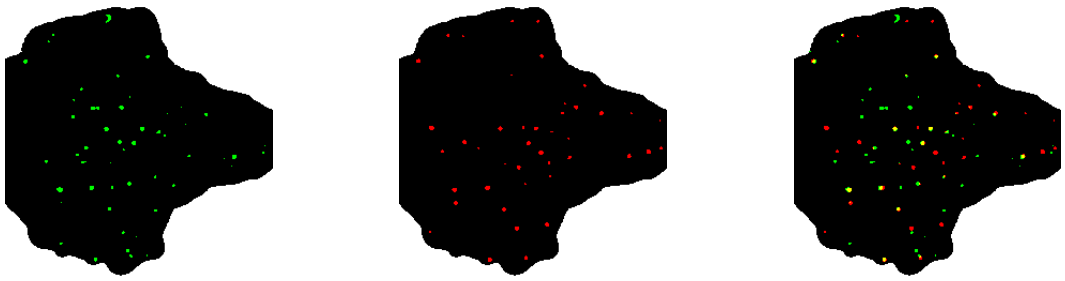


Figure 9. M10 cell showing Langerin proteins (left, in green) and Rab11 GTPase proteins (middle, in red). Right: superposition of the two previous images resulting in some possible yellow spots (co-expression of proteins within the same pixel).

7.9. Classification of diffusion dynamics from particle trajectories

Participants: Vincent Briane, Charles Kervrann.

In this study, we are currently interested in describing the dynamics of particles inside live cell. We assume that the motions of particles follow a certain class of random process: the diffusion processes. We have proposed a statistical method able to classify the motion of the observed trajectories into three groups: “confined”, “directed” and “free diffusion” (namely Brownian motion). This method is an alternative to the commonly used Mean Square Displacement (MSD) analysis. We assessed our procedure on both simulations and real cases; an example of confined diffusion is the Ornstein-Uhlenbeck process while an example of directed diffusion is the Brownian motion with constant drift. The method is currently applied to investigate membrane trafficking (Rab11/Langerin (see Fig. 10) and Rab11/TfR protein sequences) using the following procedure:

1. Tracking of particles with any competitive algorithm.
2. Statistical test /classification applied on tracks longer than ten time points.
3. Estimation of diffusion parameters (e.g. drift, diffusion, ...).

Each trajectory is labelled with the most likely process and the parameters of the underlying process are estimated. Future work will concern the detection of change of motion dynamic over time. Some results of our test on the Langerin protein sequence are shown in Fig. 10 .

Collaborator: Myriam Vimond (ENSAI Rennes).

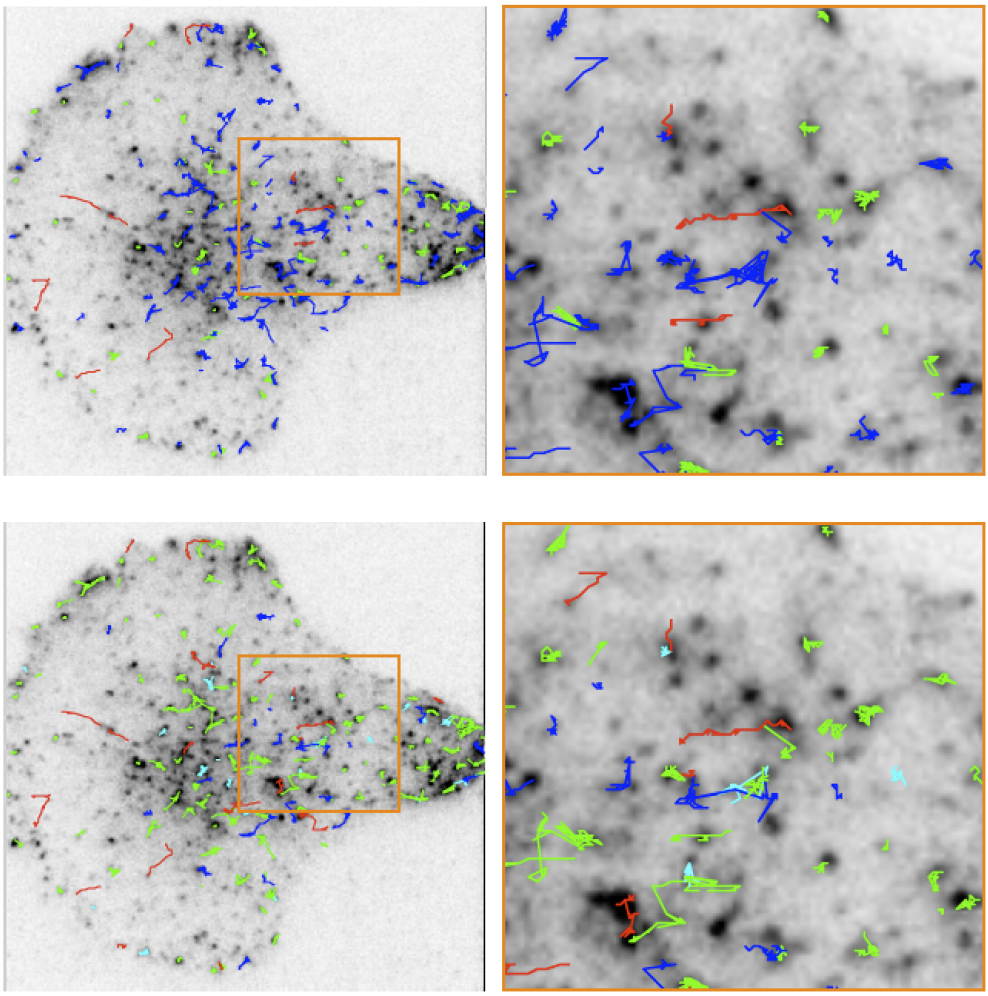


Figure 10. Labelling of the dynamics of trajectories on the Langerin protein sequence (Courtesy of UMR 144 CNRS-Institut Curie and PICT IBiSA). We display only the trajectories appearing on the first 100 frames. The color code is red for directed Brownian, green for Ornstein-Uhlenbeck, blue for Brownian, cyan for motionless. Top panel is labelled with our test, bottom panel with the MSD method.

7.10. Inference for spatial Gibbs point processes

Participant: Frédéric Lavancier.

Gibbs point processes are popular and widely used models in spatial statistics to describe the repartition of points or geometrical structures in space. They initially arose from statistical physics where they are models for interacting particles. They are now used in as different domains as astronomy, biology, computer science, ecology, forestry, image analysis and materials science. Assuming a parametric form of the Gibbs interaction, the natural method to estimate the parameters is likelihood inference. Since its first use in the 80's, this method is conjectured to be consistent and efficient. However the theoretical properties of maximum likelihood for Gibbs point processes remain largely unknown. In [39], we have partly solved this 30 years old conjecture by proving the consistency of the likelihood procedure for a large class of Gibbs models. As important examples, we deduced the consistency of the maximum likelihood estimator for all parameters of the Strauss model, the hardcore Strauss model, the Lennard-Jones model and the area-interaction model, which are commonly used models in practice.

A practical issue of likelihood estimation yet is that this method depends on an intractable normalizing constant that has to be approximated by simulation. To avoid this problem, other methods of estimation have been introduced, including pseudo-likelihood estimation. The theoretical properties of the pseudo-likelihood method are fairly well known in the case of finite-range Gibbs interactions. However, this setting rules out some major Gibbs models as the Lennard-Jones model. In [15], we have extended the pseudo-likelihood procedure to infinite range Gibbs interactions and proved its consistency and its asymptotic normality.

References: [15], [39]

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

7.11. Statistical aspects of Determinantal Point Processes

Participant: Frédéric Lavancier.

Determinantal point processes (DPPs) have been introduced in their general form by Macchi (1975) and have been extensively studied from a probabilistic point of view in the 2000's (one of the main reason being their central role in random matrix theory). In [23], we have demonstrated that DPPs provide useful models for the description of spatial point pattern datasets where nearby points repel each other. We have exploited the appealing probabilistic properties of DPPs to develop parametric models, where the likelihood and moment expressions can be easily evaluated and realizations can be quickly simulated. We have discussed how statistical inference is conducted using the likelihood or moment properties of DPP models, and we provided freely available software for simulation and statistical inference.

In [13], we have addressed the question of how repulsive a stationary DPP can be, in order to assess the range of practical situations this promising class of models may model. We determine the most repulsive DPP (in some sense) and we introduce new parametric families of stationary DPPs that can cover a large range of DPPs, from the stationary Poisson process (the case of no interaction) to the most repulsive DPP. Some theoretical aspects of inference for stationary DPPs are tackled in [37] and [38]. In the former study we have established the Brillinger mixing property of stationary DPPs, a first important step toward asymptotic inference. In the latter contribution, we have exploited this result to deduce the consistency and asymptotic properties of contrast estimators for stationary DPPs.

References: [23], [13], [37], [38]

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

7.12. Modelling aggregation and regularity in spatial point pattern datasets

Participant: Frédéric Lavancier.

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

In [22], we have considered a dependent thinning of a regular point process with the aim of obtaining aggregation on the large scale and regularity on the small scale in the resulting target point process of retained points. Various parametric models for the underlying processes are suggested and the properties of the target point process are studied. Simulation and inference procedures have been discussed when a realization of the target point process is observed, depending on whether the thinned points are also observed or not. Some typical simulations of the target processes are shown in Fig. 11 .

Reference: [22]

Collaborator: Jesper Møller (Department of Mathematical Sciences, Aalborg University, Denmark).

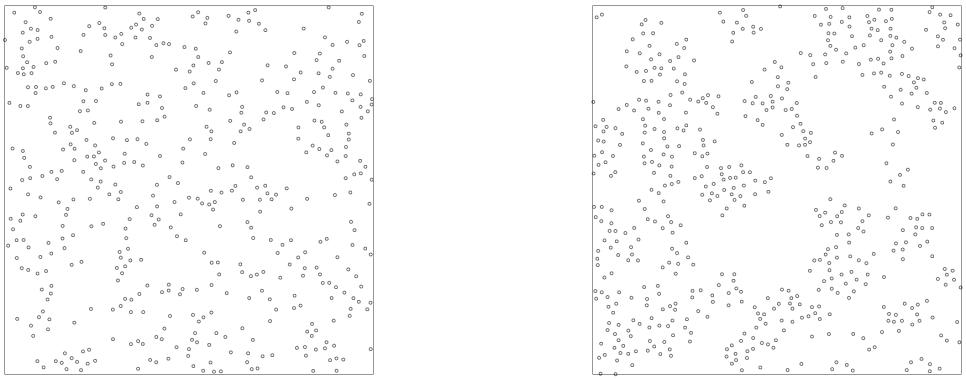


Figure 11. Examples of simulations with aggregation on the large scale and regularity on the small scale.

7.13. Retracing and registration for Correlative light-electron microscopy (CLEM)

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

Correlative light-electron microscopy (CLEM) enables to relate cell dynamics visualized in light microscopy (LM) with cell structure provided by electron microscopy (EM) for a better understanding of cell mechanisms. Registration of LM and EM modalities is then a timely, important but difficult open problem, which still requires some manual assistance. LM and EM images are indeed of very different size, spatial resolution, field of view, and appearance. We have investigated an original automated approach for the retracing-and-registration stage of the overall CLEM workflow (see Fig. 12). Pairing between the LM region of interest (ROI) and the corresponding EM patch relies on a common representation for both images, based on the LoG (Laplacian of Gaussian) transform with an adaptive associated scale (or blurring). We exploit histograms of the LoG values or histograms supplied by the LDP (Local Directional Pattern) texture descriptor, with associated histogram distances, to solve the EM patch search issue. The search step supplies a pre-registration, which is

refined by the estimation of an affine motion model to overlay the EM image onto the LM image around the ROI. Preliminary results on real CLEM images provided by UMR 144 CNRS-Institut Curie demonstrated the interest and efficiency of the proposed method.

Collaborators: Perrine Paul-Gilloteaux and Xavier Heiligenstein (UMR 144 CNRS-Institut Curie).

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

Participants: Emmanuel Moebel, Charles Kervrann.

In this study, we have addressed two important issues in cryo electron tomography (CET) images: the low signal-to-noise ratio and the presence of a missing wedge (MW) of information in the spectral domain. Indeed, according to the Fourier slice theorem, limited angle tomography results into an incomplete sampling of the Fourier domain. Therefore, the Fourier domain is separated into two regions: the known spectrum (KS) and the unknown spectrum, the latter having the shape of a missing wedge (see Fig. 13). The proposed method tackles both issues jointly, by iteratively applying a denoising algorithm in order to fill up the MW, and proceeds as follows:

1. Excitation step: Add noise into the MW.
2. Denoising step: Apply a patch-based denoising algorithm.
3. Repeat steps 1 and 2, by keeping KS constant through the iterations.

The excitation step is used to randomly initialize the coefficients of the MW, whereas the denoising step acts as a spatial regularization. The employed denoising algorithm, which exploits the self-similarity of the image, filters out coefficient values which are dissimilar to KS, thereby keeping similar ones. By iterating these steps, we are able to diffuse the information contained in KS into the MW.

An application example on experimental data can be seen on Fig. 13 , which shows the data in both spectral and spatial domain. The data contains a spherical gold particle, deformed by MW induced artifacts: elongation of the object, side- and ray-artifacts. From the residue image it can be seen that noise and MW artifacts have been reduced, while preserving the details of the image. Experiments are being performed to verify if particle detection and alignment are enhanced by using the method as a pre-processing step.

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

7.15. Algorithms for row registration to improve quality of Tissue MicroArray (TMA) images

Participants: Hoai Nam Nguyen, Charles Kervrann.

Row jittering is a common problem arising in medical imaging devices such as CT (Computer Tomography) and MRI (Magnetic Resonance Imaging) scanners due to errors of synchronization during image acquisition process. On scanners designed and developed by Innopsys, the problem becomes more challenging mainly because the pixel displacement is non constant along each row (Fig. 14) and possibly sub-pixel (i.e. non integer translation). To overcome this drawback, we first proposed a window-based algorithm to approximate the translation at each pixel by selecting the value that best minimizes a matching criteria over a finite set of possible sub-pixel translations. We obtained satisfying results with this method on real data with fast computation time (see Fig. 14). Furthermore, this matching criteria has been considered as a data fidelity term and was combined to a regularization term to promote a smooth solution and correct small artifacts which were not removed with the window-based method. To minimize the energy functional, we have adopted the quadratic relaxation technique and proximal method. This algorithm is slower and is initalized by the window-based algorithm to produce very encouraging results and elimination of all undesirable artifacts (see Fig. 14).

Collaborators: Vincent Paveau and Cyril Cauchois (Innopys).

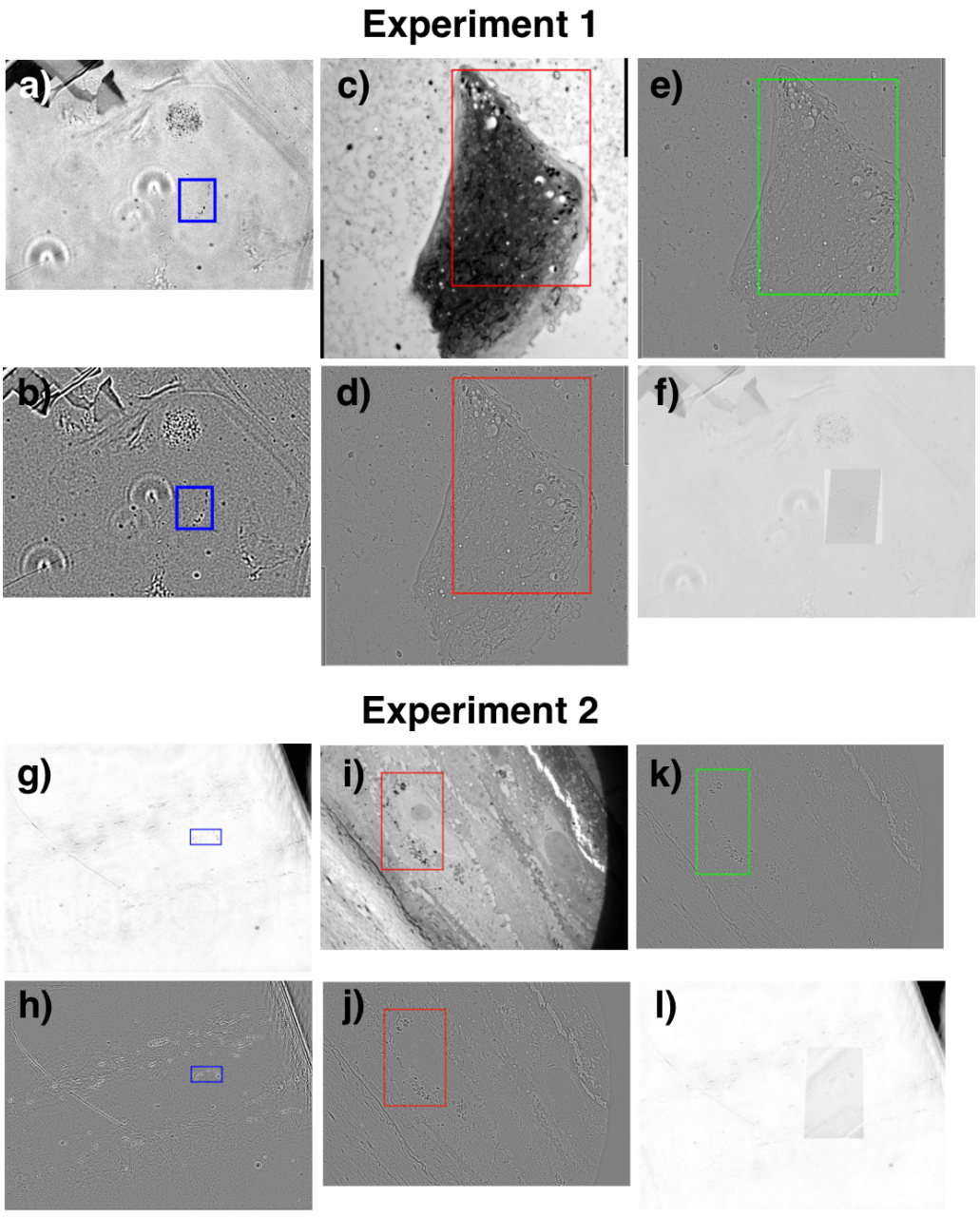


Figure 12. CLEM experiment #1: a) LM image with Region of Interest (ROI) framed in blue; b) same ROI delineated in the LoG-LM image; c) ground-truth location of the corresponding EM patch framed in red; d) the same but in the LoG-EM image; e) selected patch (SP) in the LoG-EM image in green; f) overlay (after registration) of the (decimated) EM image on the LM image around the ROI. CLEM experiment #2: g) LM-ROI in blue; h) LoG-LM-ROI; i) EM-GT in red; j) LoG-EM-GT; k) LoG-EM selected patch; l) Overlay of EM on LM around ROI.

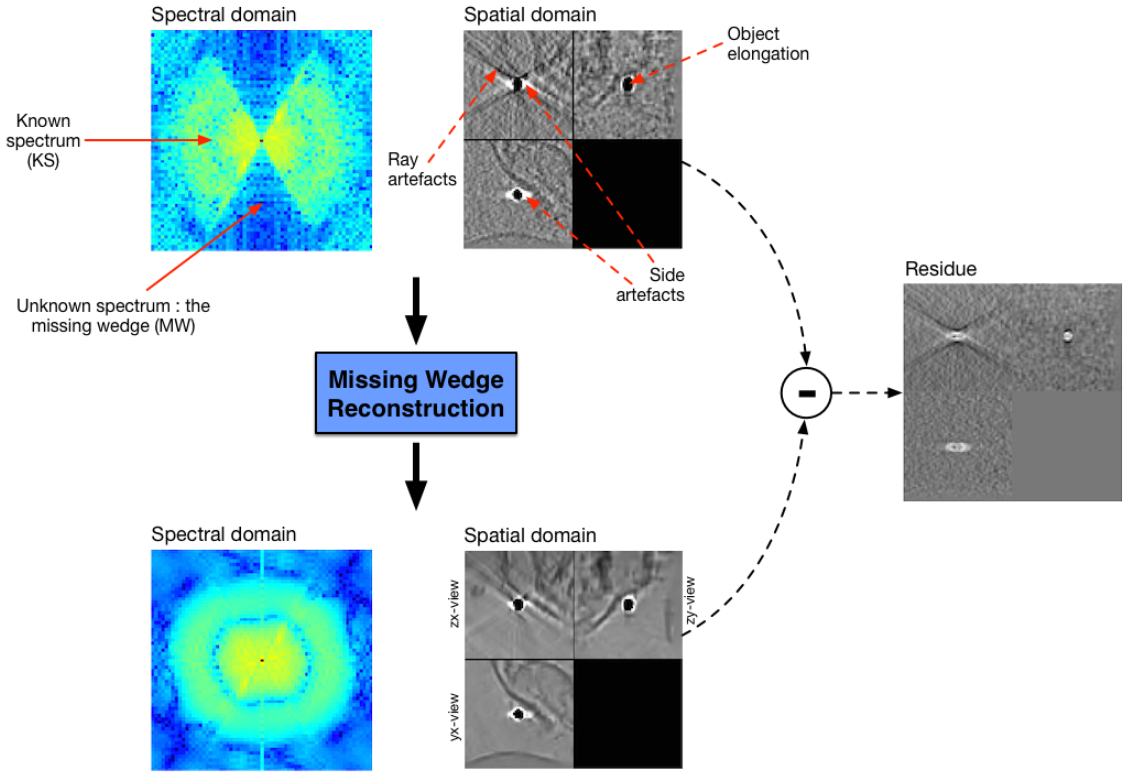


Figure 13. Experimental result of denoising and compensation of the missing wedge in cryo electron tomography.

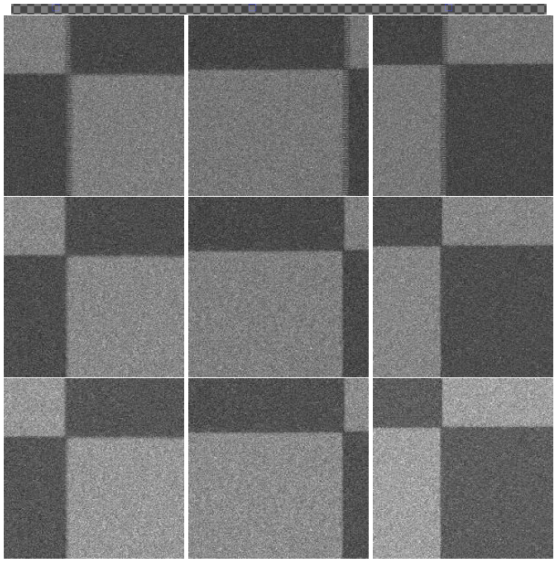


Figure 14. Illustration of the two-row registration algorithms. First row : full width input image (by courtesy of Innopsys). Second row : zooms on input image (blue boxes). Third row : corrected with window-based algorithm. Fourth row : corrected with variational method.

7.16. Robust motion model selection

Participants: Patrick Bouthemy, Bertha Mayela Toledo Acosta.

Parametric motion models are commonly used in image sequence analysis for different tasks. A robust estimation framework is usually required to reliably compute the motion model. However, choosing the most appropriate model in that estimation context is still an open issue. Indeed, penalizing the model complexity while maximizing the size of the inlier set may be contradictory. In this study, we proposed a robust motion model selection method which relies on the Fisher statistic. We also derived an interpretation of it as a robust C_P-Mallows criterion. The resulting criterion is straightforward to compute and explicitly involves the aforementioned trade-off between maximizing the size of the inlier set and minimizing the complexity (i.e., the number of parameters) of the selected motion model. We have conducted a comparative experimental evaluation on synthetic and real image sequences demonstrating that our criterion outperforms the RBIC criterion.

Collaborator: Bernard Delyon (IRMAR Rennes).

7.17. Anomaly detection in crowded scenes

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

We have defined an original motion-based method to detect and localize abnormal events in videos of crowded scenes. The algorithm relies on so-called labeled affine flows, involving both affine motion types and affine velocity vectors, and on view-based crowd motion classes. At every pixel the crowd motion class is inferred from the affine motion model selected among a set of candidate models estimated over a collection of windows. Then, the image is subdivided in blocks where local crowd motion class histograms weighted by the affine motion vector magnitudes are computed. They are block-wise compared to histograms of normal behaviors with a combined distance. More specifically, we introduce the so-called local outlier factor (LOF) to detect anomalous blocks. LOF is a local flexible measure of the relative density of data points in a feature space,

here the space of crowd motion class histograms. By thresholding the LOF value, we can detect an abnormal event in a given block at a given time. Comparative experiments on several real datasets demonstrated that our method is competitive with methods relying on far more elaborated models and exploiting both appearance and motion, while yielding superior performance over motion-based anomaly detection methods.

7.18. Occlusion detection in image sequences

Participants: Juan Manuel Perez Rua, Patrick Bouthemy.

The problem of localizing occlusions between consecutive frames of a video is important but rarely tackled on its own. In most works, it is tightly interleaved with the computation of accurate optical flows, which leads to a delicate chicken-and-egg problem. With this in mind, we proposed a novel approach to occlusion detection where visibility or not of a point in next frame is formulated in terms of visual reconstruction. The key issue is now to determine how well a pixel in the first image can be "reconstructed" from co-located colors in the next image. We first exploited this reasoning at the pixel level with a new detection criterion. Contrary to the ubiquitous displaced-frame-difference, the proposed alternative does not critically depend on a pre-computed, dense displacement field, while being shown to be more effective. We then leveraged this local modeling within an energy-minimization framework that delivers occlusion maps. An easy-to-obtain collection of parametric motion models is exploited within the energy to provide the required level of motion information. Our approach outperforms state-of-the-art detection methods on the challenging MPI Sintel dataset.

Collaborators: Tomas Crivelli and Patrick Pérez (Technicolor).

VIRTUAL PLANTS Project-Team

6. New Results

6.1. Analysis of structures resulting from meristem activity

6.1.1. Acquisition and design of plant geometry

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

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (ligth, rain, wind, pests, ...), to set up initial conditions of growth models or to assess their output against real data. In the past decade, methods have been developed to digitize plant architectures in 3D [76], [63]. These methods are based on direct measurements of position and shape of every plant organ in space. Although they provide accurate results, they are particularly time consuming. More rapid and automated methods are now required in order to collect plant architecture data of various types and sizes in a systematic way. In this aim, we explore the use of laser scanner and pictures.

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

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

  For the branching systems, we previously proposed a reconstruction method to reconstruct plausible branching structures from laser scanner data based on the concept of space colonization [73]. Additionally, a number of automatic methods were proposed in the literature. The question of their comparison and relative accuracy is however critical for further exploitation in biological applications. To address such problem, we developed an evaluation pipeline that takes two plant structures as input and compares their organization using two indices of geometrical and structural similarities [55]. A first comparative evaluation of the different methods of the literature has been designed and conducted. A graphical editor has been developed and makes it possible to test the different methods and correct manually the reconstruction. A procedure to automatically determine phyllotactic angles from scans of small plants has been added to the reconstruction pipeline and has been tested on database of 150 scans of arabidopsis thaliana with different genotypes. The editor has also been tested on apple trees and large African trees.

  In the context of the PhD of M. Balduzzi, we also investigated the reconstruction of tree foliage from 3D scans. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers in the silhouette of the scans that make the meshing of the pointset extremely difficult. New generation of laser scanners provide intensity of the laser reflected on the surface of scanned objects. This intensity depends on the distance to the object, its optical property and the incidence angle. A first work on this topic showed that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we developed a reconstruction pipeline using the scan intensities and based on Shape-From-Shading. Outliers being along the edge of the surface point cloud, we chose to develop a propagation SFS method initialized with points of the scans with high quality. We proved that surface with constant intensity are necessarily surfaces of constant slope or sand-pile surfaces. Using this result, a propagation method along iso-intensity regions was developed. These surfaces can then be sampled to provide a smooth point set without outliers.
- *Reconstruction of annual plants from multi-view images.* (Simon Artzet, Jerome Chopard, Christian Fournier, Christophe Pradal, Christophe Godin, Xavier Sirault [CSIRO-HRPPC, Canberra])

Image-based phenotyping platforms in semi-controlled conditions offer large possibilities to perform genetic analyses of plant growth, architecture, light interception, and biomass accumulation over large time series for thousands of plants. However, methods for image analyses currently available are still very crude and need improvement and robustness to process huge amount of data. We are developping an integrated pipeline allowing assessment of growths of individual organs, of plant geometry, and of derived variables such as light interception. The pipeline currently consists of 2D image analysis workflows built with standard image libraries (OpenCV, Scikit.Image), algorithms for 3D reconstruction, segmentation and tracking of plant organs for maize (under development), and workflows for estimation of light interception by plants during their growth.A 3D FSPM model for maize architectural development, is used to help segmenting plant images and to automate the mapping between segmented 3D objects and plant organs defined in the model.

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

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

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

  1. We integrated a validation step in the workflow based on the comparison method presented in [55].
  2. We developed a standard file format for root architecture (RSML) described in [19] during an international collaboration with the niversité Catholique de Louvain (Belgium), the CPIB of the University of Notthingham (UK), the University of Vienna (Austria), the Jülich research center (Germany) and INRA.

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

  This research theme is supported by the Agropolis project MecaFruit3D.

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function. To this end, a modeling pipeline has been developed in the OpenAlea platform. It involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex vascular network that is embedded within this mesh using the concept of space colonization [75]. Previous studies demonstrated the possibility to create species-specific models of fruit structure with relatively low effort [57]. We focus now on validating the vascular networks by comparing them to experimental data from the literature. This work has been presented at the ISHS symposium in Montpellier [38]

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

6.1.2. Modeling the plant ontogenic programme

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

This research theme is supported by two PhD programmes.

The remarkable organization of plants at macroscopic scales may be used to infer particular aspects of meristem functioning. The fact that plants are made up of the repetition of many similar components at different scales, and the presence of morphological gradients, e.g. [52], [65], [66], [62], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as "morphogenetic programme" [71], "age state" [61] or "physiological age" [54]. All these concepts state that meristem fate changes according to position within the plant structure and during its development. Even though these changes in meristem fate are specific to each species and lead to the differentiation of axes, general rules can be highlighted [61], [54]. Here we develop computational methods to decipher these rules.

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

  Shoot branching structures often take the form of a succession of homogeneous branching zones and have been analyzed using segmentation models such as hidden semi-Markov chains. Axillary meristem fates are influenced by local properties of the parent shoot such as for instance its growth rate or local curvature. The objective of this work, which was part of the PhD subject of Jean Peyhardi, is to develop statistical models that generalize hidden semi-Markov chains with the capability to incorporate explanatory variables that vary along the parent shoot (e.g. leaf growth rate, leaf surface, internode length, local curvature of the parent shoot). More precisely, the simple multinomial distributions that represent the axillary productions observed in the different branching zones are replaced by multinomial generalized linear models (GLMs). Since the two classical categories of multinomial GLMs that correspond either to nominal or ordinal categorical response variables were not appropriate, we chose to develop a new family of multinomial GLMs called partitioned conditional GLMs [72] that enable to tackle hierarchically-structured categorical response variables. Typically, we need to distinguish different timing of branching events (e.g. immediate shoot, one-year-delayed shoot and latent bud), different categories of offspring shoots (e.g. among one-year-delayed shoots, vegetative short shoot, vegetative long shoot and flowering shoot) and to specialize the explanatory variables for certain categories of offspring shoots (e.g. the growth of the parent shoot influence the immediate offspring shoots but not the one-year-delayed offspring shoots). The resulting integrative models are called semi-Markov switching partitioned conditional GLMs and have been applied to apple and pear tree branching structures.
- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Alix Allard [AGAP, AFEF team], Jean Peyhardi, Baptiste Guitton [AGAP, AFEF team], Yan Holtz [AGAP, AFEF team] Catherine Trottier, Evelyne Costes [AGAP, AFEF team], Yann Guédon)

  A first study was published to characterize genetic determinisms of the alternation of flowering in apple tree progenies [58]. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot, which corresponds to portions of stem that were grown during the same year). Two replications of each genotype were available.

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

  For better interpretation of the relationships of alternation at both scales, new models and indices were developped for sequences of flowering events at axis scale. New data sets where collected in other F1 progenies. Ancestral relationships between parents of different progenies were taken into account to enhance the power of QTL detection, and other QTL were found using these new indices.
- *Identifying and characterizing patterns in tree-structured data* (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).

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

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

In multitype branching processes, the plant development is viewed as a demographic process, a parent entity of a given type generating child entities of different types (e.g. vegetative and flowering entities). Formally, the botanical entity properties are summarized as a categorical state variable. The number of child entities in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious generation distributions. We developed an approach based on probabilistic graphical models to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution. The graph building stage was based on exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs [59]. To relax the strong constraints regarding dependencies induced by parametric distributions, mixture of graphical models were also considered [60]. Multitype branching processes were applied to the analysis of the patchiness pattern (consisting of canopies made of clumps of either vegetative or flowering growth units) in mango trees. To identify the clumps, a novel approach based on tree-segmentation was developed [35].

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

  Mango is a tropical tree characterized by strong asynchronisms within and between trees. To study more precisely the interplay between the plant structural components, we built an integrative model to simulate the plant development based on the L-system formalism and GLM to model the dependencies between events. With such model, we showed the importance of architectural and temporal factors in the development of the units of the trees. The model also simulates the phenology of shoots and inflorescences. For this, the sizes of the different organs is modelled by statistical laws estimated from measurements that depends on their locations in the architecture. The growth speed of organs is modulated by the temperature. This structural and phenological model has been presented at the ISHS symposium on Montpelier [37].

  This year, the model has been extended, during the intership of S. Persello to take into account fruiting probabilities and coupled with an ecophysiological model of fruit growth [68], [69]. The global aim is to have a crop simulation model to predict fruit yield and quality on mango tree. An overview of this global model based on the coupling of different structural or ecophysiological submodels has been also presented in different ISHS symposia [40], [50]

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

  Plant growth, i.e. the increase of organ dimensions over time, and development, i.e. the change in plant structure, are often studied as two separate processes. However, there is structural and functional evidence that these two processes are strongly related. Our aim was to investigate the co-ordination between growth and development using mango trees, which have well-defined developmental stages. Developmental stages, determined in an expert way, and organ sizes, determined from objective measurements, were collected during the vegetative growth and flowering phases of two cultivars of mango. For a given cultivar and growth unit type (either vegetative or flowering), a multi-stage model based on absolute growth rate (AGR) sequences deduced from the measurements was first built, and then growth stages deduced from the model were compared with developmental stages. Strong matches were obtained between growth stages and developmental stages, leading to a consistent definition of integrative developmental growth stages [14]. The growth stages highlighted growth asynchronisms between two topologically connected organs, namely the vegetative axis and its leaves. Integrative developmental growth stages emphasize that developmental stages are closely

related to organ growth rates and can be interpreted in terms of the possible physiological processes (hydraulics, biomechanics and carbohydrate partitioning) underlying these stages. We also explore growth stages deduced from relative growth rate (RGR) sequences applying the same methodology. AGR and RGR have different meanings regarding plant metabolism since AGR represents net sink strength whereas RGR represents net sink activity. For vegetative growth units, the match rates between RGR-based stages and developmental stages were rather similar to the match rates between AGR-based stages and developmental stages, because of the rich information provided by the four organs modeled (the axis and three selected leaves). The match rates were far lower for the inflorescences where only the main axis was modeled. This is related to the fact that, compared to AGRs, RGRs amplify the variations at the beginning of growth of an organ while damping the variations at the end of growth.

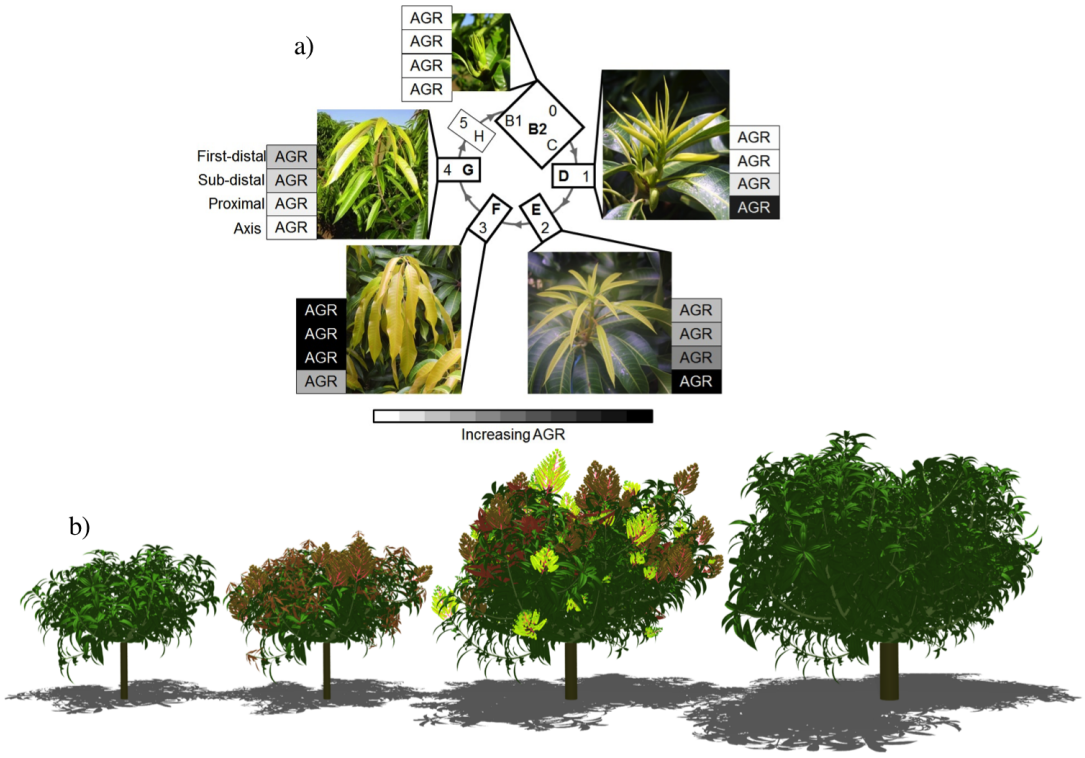


Figure 1. a) Correspondences between developmental stages of mango growth units determined from morphological observations in an expert way (inside the central circle) and growth stages obtained using segmentation models (outside the circle) [14]. Main developmental stages in bold are illustrated by the photographs. Variations in absolute growth rate for axis and leaves are illustrated using a white to black scale. b) Simulation of the development of a mango tree over two cycles [37]. The first and last image corresponds to the end of the vegetative period of the 3rd and 5th growing cycle (June), respectively while the second and third images correspond to the flowering phase (August) of the 3rd and 4th cycles, respectively. The different colours of the inflorescences of the 3rd image show different developmental stages and the flowering asynchronism over the tree.

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

Our aim was to characterize the successive flowering phases of perpetual flowering strawberry genotypes, which is of particular importance for better predicting fruit production. We applied multiple change-point models for the synchronous segmentation of the individuals of a given genotype in successive flowering phases. We identified two groups of genotypes that differ by the intensity of the flowering at the end of the flowering period. Using a genetic approach, we identified a locus controlling the flowering intensity at the end of the flowering period that likely explain these two groups of genotypes. A multivariate generalization of the synchronous segmentation approach is developed in the context of Marc Labadie's PhD, the idea being to characterize not only the flowering pattern as in our first study but more generally the developmental pattern combining vegetative development, branching and flowering.

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

  In a previous work [7], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG (Directed Acyclic Graph, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree.

  The method proposed in [7] thus compresses a tree in width, but not in height. In a new work, we designed an extension of this compression method in which a tree is compressed in both width and height. The method is based on the detection of so-called *quasi-isomorphic paths* in a tree and on the compression of these paths in height. A paper describing the corresponding algorithms has been recently accepted in the Journal of Theoretical Biology (To appear).

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

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

This research theme is supported by three PhD programs.

The ontogenetic programme of a plant is actually sensitive to environmental changes. If, in particular cases, we can make the assumption that the environment is a fixed control variable (see section 6.1.2), in general the structure produced by meristem results from a tight interaction between the plant and its environment, throughout its lifetime. Based on observations, we thus aim to trace back to the different components of the growth (ontogenetic development and its modulation by the environment). This is made using two types of approaches. On the one hand, we develop a statistical approach in which stochastic models are augmented with additional time-varying explanatory variables that represent the environment variations. The design of estimation procedures for these models make it possible to separate the plant ontogenetic programme from its modulation by the environment. On the other hand, we build reactive models that make it possible to simulate in a mechanistic way the interaction between the plant development and its environment.

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

  In the context of a collaboration with Claudia Negron and Ted DeJong, we studied the influence of water availability and pruning practices [21] on the branching and axillary flowering structures of different categories of almond shoots. Stochastic models (hidden semi-Markov chains) were built for the branching and axillary flowering structures of different categories of almond shoots corresponding to different genetic backgrounds, levels of irrigation and pruning practices.

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

We identified robust indicators that summarize the respective importance of ontogeny and environmental constraints (mainly related to light environment) in forest tree development [26]. In this context, tree growth data correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models to identify tree growth phases. These segmentation models, which are hidden semi-Markov chains, were compared with simple hidden Markov chains that correspond to the environment-driven development assumption. This statistical modelling approach was applied to both evergreen (Corsican pine and silver fir) and deciduous (sessile oak and Persian walnut) tree species growing in contrasted conditions ranging from managed forest stands to unmanaged understoreys. Growth phase duration distributions estimated within these segmentation models characterize the respective importance of ontogeny and environmental constraints in tree development at the population scale and have very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. These characteristics may change over tree life, reflecting changes in tree competition. Growth phase duration distributions summarize the joint trajectory of tree ontogeny and environment without requiring tree growth follow-up data for their estimation.

- *Analyzing fruit tree phenology in various climatic conditions* Yann Guédon, Jean-Michel Legave [AFEF team, AGAP], Gustavo Malagui [Universidade Tecnológica Federal do Paraná]

  The responses of flowering phenology to temperature increases in temperate fruit trees have rarely been investigated in contrasting climatic regions. This is an appropriate framework for highlighting varying responses to diverse warming contexts, which would potentially combine chill accumulation declines and heat accumulation increases. To examine this issue, a data set was constituted in apple tree from flowering dates collected for two phenological stages of three cultivars in seven climate-contrasting temperate regions of Western Europe and in three mild regions, one in Northern Morocco and two in Southern Brazil. Multiple change-point models were applied to flowering date series, as well as to corresponding series of mean temperature during two successive periods, respectively determining for the fulfillment of chill and heat requirements. A new overview in space and time of flowering date changes was provided in apple tree highlighting not only flowering date advances as in previous studies but also stationary flowering date series [18]. At global scale, differentiated flowering time patterns result from varying interactions between contrasting thermal determinisms of flowering dates and contrasting warming contexts. This may explain flowering date advances in most of European regions and in Morocco vs. stationary flowering date series in the Brazilian regions. A notable exception in Europe was found in the French Mediterranean region where the flowering date series was stationary. While the flowering duration series were stationary whatever the region, the flowering durations were far longer in mild regions compared to temperate regions. Our findings suggest a new warming vulnerability in temperate Mediterranean regions, which could shift towards responding more to chill decline and consequently experience late and extended flowering under future warming scenarios.

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

  Sustainable agriculture requires the identification of new, environmentally responsible strategies of crop protection. Modelling of pathosystems can allow a better understanding of the major interactions inside these dynamic systems and lead to innovative protection strategies. In particular, functional–structural plant models (FSPMs) have been identified as a means to optimize the use of architecture-related traits. A current limitation lies in the inherent complexity of this type of modelling, and thus the purpose of this work is to provide a framework to both extend and simplify the modelling of pathosystems using FSPMs. Complex models are disassembled into separate *knowledge sources* originating from different specialist areas of expertise and these can be shared and reassembled into multidisciplinary models. This year, we worked on four application studies that used the framework. In the frame of the PhD of Guillaume Garin, we perform a validation of the wheat septoria model, an analysis of the influence of the wheat architecture on the competition

between septoria and brown rust, and a sensitivity analysis of the response of the severity of septoria to architectural traits. In the frame of the Echap project, we use the wheat-septoria model to indentify optimal date of pesticide application. All these studies allows to populate the framework with consistent example of application, and lead to the development of operational modules that allows the fitting and validation of pathosystem models with experimental data.

6.2. Meristem functioning and development

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

6.2.1. Data acquisition and design of meristem models

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging
  Participants: Léo Guignard, Christophe Godin, Christophe Pradal, Grégoire Malandain [Morpheme, Inria], Gaël Michelin [Morpheme, IPL Morphogenetics, Inria], Guillaume Baty, Sophie Ribes [IBC, UM], Jan Traas [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Yassin Refahi [RDP, ENS-Lyon / Sainsbury Lab, Cambridge, UK].

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

  The MARS-ALT (Multi-Angles Registration and Segmentation - Automatic Lineage Tracking) software pipeline [6] automatically performs a segmentation at cell resolution from 3D or 2D voxel images where the membranes/walls are marked (by a die for example) and makes it possible to follow the lineage of these cells through time.

  This year, the ALT tracking pipeline has been reformulated by using a generic cell modeling approach (enabling for example more than one cell division), and both stability and robustness were improved. The modeling approach is generic and can be used on other kind of data (nuclei, human cells, ...). These trials will be conducted during the year. Moreover, the architecture of the image processing components has been modified (plugin approach) and integrated with the TissueLab platform. Some vizualisation tools have been improved, and the platform includes a module allowing an interaction with data (Alizon Konig, master internship). This point enables an efficient creation of gold standard to validate segmentation results.

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

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

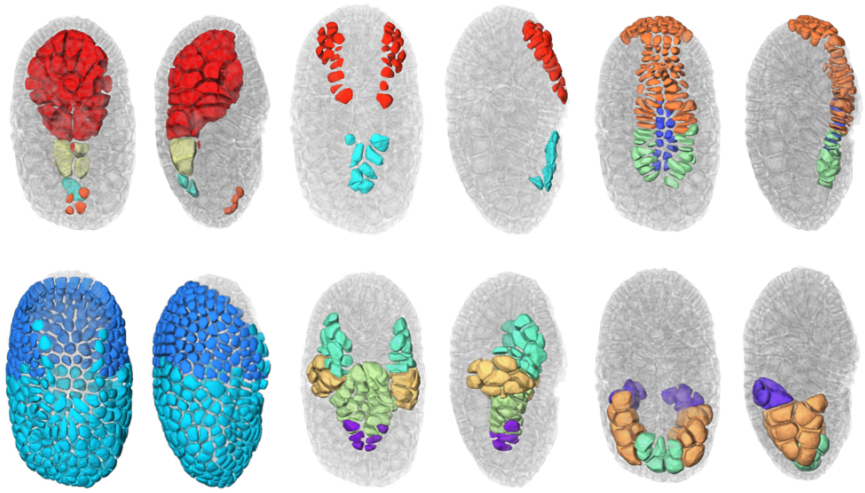


Figure 2. 3D projection of the segmented embryo at the early tailbud stage. The cells are colored by tissue type. The cells are slightly eroded to allow their distinction. The other cells of the embryo are in transparent grey. The doral and lateral sides are shown.

- *Creating mesh representation of cellular structures*
  Participants: Guillaume Cerutti, Sophie Ribes, Christophe Godin, Géraldine Brunoud [RDP, ENS], Carlos Galvan-Ampudia [RDP, ENS], Teva Vernoux [RDP, ENS], Yassin Refahi [RDP, ENS, Sainsbury Lab].

  This research theme is supported the HFSP project Biosensors.

  To produce a more efficient data structure accounting for the geometry of cellular tissues, we studied the problem of reconstructing a mesh representation of cells in a complex, multi-layered tissue structure, based either on membrane/wall images segmented using MARS or on nuclei images of shoot apical meristems. The construction of such mesh structures for plant tissues is currently a missing step in the existing image analysis pipelines.

  We developed tools to reconstruct a 3D cell complex representing the tissue, based on the dual simplicial complex of cell adjacencies. This set of tetrahedra is optimized from a reasonable initial guess to match the adjacencies in the tissue, which proved to produce a very faithful reconstruction[39]. We also developed a set of methods to triangulate such reconstructions, and enhance the quality of triangular mesh representations of plant tissue, simultaneously along several criteria [28].

  These tools can produce light discrete representations of the cell interfaces that enables fast visualization, information projection, and quantitative analysis of the tissue, and have given way to to some of the first biomechanical simulations on real-world data.

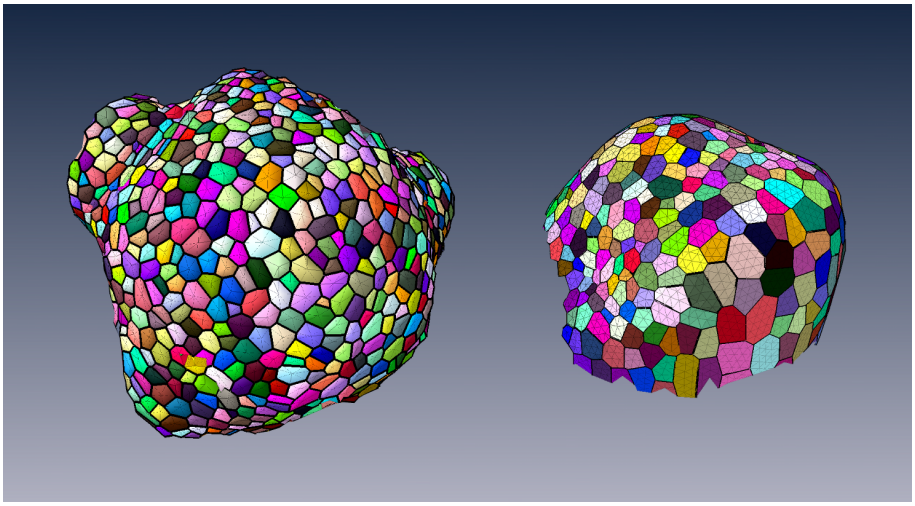


Figure 3. Triangular mesh representations of shoot apical meristem and flower meristem tissues obtained from MARS segmentations

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

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

  To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database [67]. This database makes it possible to store all

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

6.2.2. Shape analysis of meristems

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

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

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

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

6.2.3. Mechanical models of plant tissues

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

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

The rigid cell walls that surround plant cells are the main load-bearing structures in plant tissues. These walls are submitted to stresses due to cell turgor pressure. Above some threshold, these stresses cause deformation in the cell walls and triggers wall **irreversible expansion** (*synthesis*). Shape changes of plant tissues are therefore tightly related to the turgidity of cells and to the mechanical state and the molecular composition of the underlying cell walls. We developped a conceptual and numerical framework to model the mechanical structure of cell walls and their deformation by turgor pressure in 3-dimensions. This framework was used to study the interplay between post-transcriptional regulation, biochemistry, and mechanics within growing plant tissues. This work has been published this year in Plos Computational Biology [13].

In this first step, all mechanical and structural quantities are defined at the tissular scale. This is made possible by abstracting the connection between the actual molecular composition of the walls and the various signalling cascade at play during growth. To extend this approach, we also started to develop a mechanobiological approach relating the irreversible expansion of the walls to molecular mechanisms happening within them, based on the thermodynamical equilibrium of the pectin-based matrix within the wall. We propose that at the molecular scale expansion of this matrix is based on the adsorption of newly synthetized pectin molecules. This adsorption mechanism is regulated by the mechanical stresses applied on the wall. We show that this mechanism belongs to a class of biochemical / biomechanical processes commonly appearing in the dynamics of supra-molecular load-bearing structures: the force-driven polymerization processes. A preliminary version of these ideas (the 1D case) is currently under review in Trends In Plants Sciences.

We also considered to extend the original modeling approach to situations where entire organ dynamics should be modeled over large time lapse (several days) (PhD work of Jean-Philippe Bernard). In our first approach, the mechanical model relies on a finite element method (FEM) to describe the deformation of the tissue. In FEM, the tissue is represented by a mesh. The positions of the vertices at each time step are estimated from a linear system. If the tissue is big or if the mesh is fine, the linear system can be large and thus leads to computational overheads. An alternative way to classical FEM is to use a meshless method where the deformation of the tissue can be characterized by a linear combination of deformations of a finite and small set of frames. Because shape functions are no longer defined on each element but on the whole tissue, they have to be updated at each growth step by estimating a new rest configuration. With meshless method, the discretization of the system can be dynamically updated parsimoniously according to the precision required to model the emergence of shapes. With an uniform distribution of the frames within the volume, our method still leads to computational overheads. However, since the meristem initiates a branching structure at a macroscopic scale, we combined our mechanical model at tissular resolution with classical method used to generate branching structures at macroscopic scales. For this, we use the information of the plant branching structure to distribute the frames along the plant's axes. This allows us to use curvilinear shape functions while describing the branching structure growth using L-systems. This multi-scale framework allows us to define developmental rules which can initiate new organs at the surface of the meristematic dome by softening locally the meristem dome and thus creatin new growing initia. First very encouraging results were obtained this year that demonstrate the feasibility of the approach.

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

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

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

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

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

Finally, given that the cauliflower have different morphologies (i.e. regular and romanesco cauliflower morphologies) we explored the effect of changes in the L-system parameter values over the cauliflower morphology. Interestingly, we discovered by exploring the model that variations in the regulation of some phyllotactic

parameters can produce the different cauliflower morphologies and explain other reported differences among them. Predictions were made using the model and experimental validations of this hypothesis are curren being tested. All our results could provide a comprehensive understanding of how does genes and plant architecture are linked in a dynamical way.

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

Participants: Jonathan Legrand, Yann Guédon, Teva Vernoux [ENS-Lyon].

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

6.2.6. Model integration

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

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

Our approach consists of building a programmable tissue which is able to accept different modeling components. This includes a central data structure representing the tissue in either 2-D or 3-D, which is able to grow in time, models of gene activity and regulation, models of signal exchange (physical and chemical) between cells and models of cell cycle (which includes cell division). An introduction to the modeling of some main components of such integrated system was published as a book chapter in the series of Ecole de Physique des Houches [43]. For each modeling component, one or several approaches are investigated in depth, possibly at different temporal and spatial scales, using the data available from the partners (imaging, gene networks, and expression patterns). Approaches are compared and assessed on the same data. The objective of each sub-model component will be to provide plugin components, corresponding to simplified versions of their models if necessary, that can be injected in the programmable tissue platform. This work is developed in collaboration with the RDP group at ENS-Lyon [70] and the CPIB group in Nottingham, UK [53].

One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development, TissueLab. This platform, based on *OpenAlea*, will be used to carry out integration of the different models developed in this research axis. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform. Currently, robust geometric operations such as division are implemented and tested. Moreover, a redesign of the structure based on more elaborated formalisms such as combinatorial maps is being investigated. A 2D version is being developed in the context of Jean-Louis's Dinh PhD thesis, and will be described in a forthcoming book chapter.

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

6.3.1. Modeling water transport in roots

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

This research theme is supported by the ANR project HydroRoot.

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

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

6.3.2. Mechanical modeling of fruit growth

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

This research theme is supported by the Agropolis project MecaFruit3D.

Fruits and plants in general are large scale hydraulic systems in which growth is closely linked to water fluxes: thanks to osmotic pressure difference, the cells are able to absorb water from their environment and therefore increase their volume; as the cells are bounded by rigid walls, this results in both hydrostatic pressure (the so-called turgor pressure) in the cell and tension in the cell walls; above a threshold, synthesis of new cell wall material occurs and relaxes the tension. This process allows cells to grow, and along with cell division, is responsible for plant growth. In fruits, phloem and xylem vascular networks provide the water fluxes necessary for growth, while the osmotic pressure is mainly regulated by sugar intake from the phloem. The goal of this project is to combine a description of water and sugar fluxes at the fruit scale (see section 4) with a modelling of growth at cell level, as described above.

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

We have developed a numerical method for this coupled system, that allows to simulate in a reasonable amount of time a hundred of connected cells. The non linear system of equations (2) is the bottleneck to reach a higher number of cells; in order to overcome this, we plan to use the framework developed for the mechanical modelling of meristems (see section 6.2.3) and adapt it to this system. This will also allow to address tridimensional tissues.

Numerical simulations exhibit a highly non linear behaviour with respect to the governing parameters. We have identified two clearly distinct growth regimes: one regime that allows large growth heterogeneities by amplifying the effect of differences between cells, and conversely another regime that smoothes differences out and yields a homogeneous growth. On the biological level, the first regime is well adapted to morphogenesis, whereas the second one is well adapted to homothetic growth after the differentiated tissues have been created. A publication of these completely new results is in preparation.

We plan to compare this model to experimental results of the tomato fruit at the tissue level. In the longer term, a continuous version of this multicellular model could be an interesting way to build a model at the fruit scale.

6.3.3. Analyzing root growth and branching

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

This research theme is supported by two PhD programmes.

New 2D and 3D root phenotyping plateforms are emerging with associated image analysis toolbox (e.g. SmartRoot, RhizoScan) and the high-level analysis these complex phenotyping data requires new computational investigation methods.

Here, we aim at developing a pipeline of methods for analyzing root systems at three scales:

1. tissular scale to identify and characterize the division, elongation and mature zones along a root using piecewise heteroscedastic linear models. To this end, we introduced a new slope heuristic for the selection of the number of zones in cell length series [29] [36].
2. individual root scale to analyze the dynamics of lateral root elongation. We investigated the use of semi-Markov switching linear models for classifying roots on the basis of the identification of phases within growth rate profiles,
3. root system scale to analyze the branching structure.

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

6.3.4. Analyzing shoot and leaf elongation

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

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

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

The changes in leaf size and shape during ontogeny associated with the heteroblastic development is a composite trait for which extensive spatio-

temporal data can be acquired using phenotyping platforms such as PHENOPSIS. However, only part of the information contained in such data is exploited and developmental phases are usually defined using a selected organ trait. We introduced new methods for identifying developmental phases in *Arabidopsis* rosette using various traits and minimum a priori assumptions. A first pipeline of analysis was developed, combining image analysis and statistical models to integrate morphological, shape, dimensional and expansion dynamics traits for the successive leaves of the *Arabidopsis* rosette. Dedicated segmentation models called semi-Markov switching models were built for selected genotypes in order to identify rosette developmental phases. Four

successive developmental phases referred to as seedling, juvenile, transition and adult were identified for the different genotypes. We showed that the degree of covering of the leaf abaxial surface with trichomes is not sufficient to define these developmental phases. Using our pipeline of analysis, we were able to identify the supplementary seedling phase and to uncover the structuring role of various leaf traits. This enabled us to compare on a more objective basis the vegetative development of *Arabidopsis* mutants.

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

6.3.5. A stochastic model of phyllotaxis

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

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

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

Recently, our colleagues from ENS-Lyon observed intriguing perturbation in *arabidopsis* mutants. These perturbations were also present, to a lesser extent in the wild type. In a series of works [74], [64], [2], we could show that these perturbations patterns in both wild-type and mutant plants could be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs. After closer inspection, we realized that the mutated gene encodes a protein diffusing from the organs and creating a field around the organs that regulates the plastochron. We could demonstrate that in the mutant, the absence of this field leads to co-initiations and subsequently to the observed permutations.

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

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

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

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

6.4. Generic methodological results

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

6.4.1. Scientific workflows

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

6.4.1.1. OpenAlea scientific workflows

Analyzing biological data may involve very complex and interlinked steps where several tools are combined together. Scientific workflow systems have reached a level of maturity that makes them able to support the design and execution of such in-silico experiments, and thus making them increasingly popular in the bioinformatics community (e.g. to annotate genomes, assemble NGS data, ...) . However, in some emerging application domains such as system biology, developmental biology or ecology, the need for data analysis is combined with the need to model complex multi-scale biological systems, possibly involving multiple simulation steps. This requires the scientific workflow to deal with retro-action to understand and predict the relationships between structure and function of these complex systems. In collaboration with the Zenith EPI, we have proposed a conceptualisation of OpenAlea workflows [34] by introducing the concept of higher-order dataflows as a means to uniformly combine classical data analysis with modeling and simulation. Ongoing work include deploying OpenAlea workflows on a Grid technology using the SciFloware middleware in close collaboration with Zenith within IBC and INRA Phenome projects.

6.4.1.2. Querying Scientific workflows repositories

Several workflow systems have developped scientific workflow repositories (e.g., repositories of Galaxy workflows at IBC, or repositories of OpenAlea workflows). Such repositories have grown to sizes that call for advanced methods to support workflow discovery, in particular for similarity search. Effective similarity search requires both high quality algorithms for the comparison of scientific workflows and efficient strategies for indexing, searching, and ranking of search results. Yet, the graph structure of scientific workflows poses severe challenges at each of these steps. We present a complete system for effective and efficient similarity search in scientific workflow repositories, based on the Layer Decompositon approach to scientific workflow comparison. Layer Decompositon specifically accounts for the directed dataflow underlying scientific workflows and, compared to other state-of-the-art methods, delivers best results for similarity search at comparably low runtimes. Stacking Layer Decomposition with even faster, structure-agnostic approaches allows us to use proven, off-the-shelf tools for workflow indexing to further reduce runtimes and scale similarity search to sizes of current repositories [25]. Very efficient and powerful ranking methods have been used in this work. We based our choice on the large scale study of algorithms for rank aggregation with ties we performed [56].

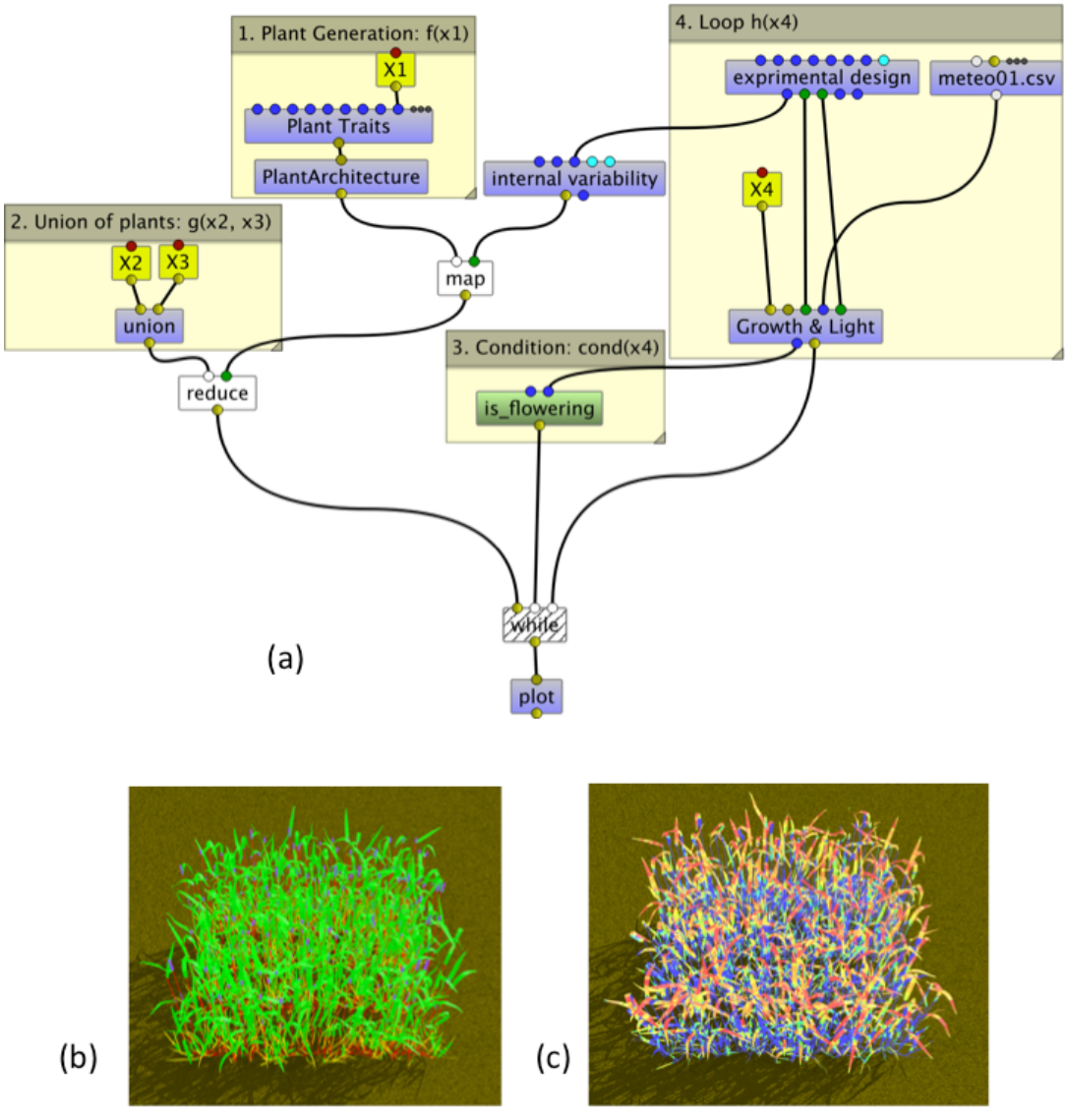


Figure 4. (a) OpenAlea workflow [34] for simulating Maize and Wheat crop performance based on phenotypic and environment data, and two image outputs (b and c). Colors represent the organ's type in (b) and the amount of intercepted light in (c).

6.4.2. Statistical modeling

Participants: Yann Guédon, Jean Peyhardi.

We develop statistical models and methods for identifying and characterizing developmental patterns in plant phenotyping data. Phenotyping data are very diverse ranging from the tissular to the whole plant scale but are often highly structured in space, time and scale. Problems of interest deal with the definition of new family of models specifically adapted to plant phenotyping data and the design of new methods of inference concerning both model structure, model parameters and latent structure. This is illustrated this year by [17] and [22].

6.4.3. Lossy compression of tree structures

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

the degree of self-nestedness of a tree as the edit-distance between the considered tree structure and its nearest embedded self-nested version. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We thus introduced a lossy compression method that consists in computing in polynomial time for trees with bounded outdegree the reduction of a self-nested tree that closely approximates the initial tree. This approximation relies on an indel edit distance that allows (recursive) insertion and deletion of leaf vertices only. We showed in a conference paper accepted at DCC'2016 [46] with a simulated dataset that the error rate of this lossy compression method is always better than the loss based on the nearest embedded self-nestedness tree [7] while the compression rates are equivalent. This procedure is also a keystone in our new topological clustering algorithm for trees. In addition, we obtained new theoretical results on the combinatorics of self- nested structures. The redaction of an article is currently in progress.

ARAMIS Project-Team

7. New Results

7.1. Learning spatiotemporal trajectories from manifold-valued longitudinal data

Participants: Jean-Baptiste Schiratti [Correspondant], Stéphanie Allassonniere, Olivier Colliot, Stanley Durrleman.

We propose a Bayesian mixed-effects model to learn typical scenarios of changes from longitudinal manifold-valued data, namely repeated measurements of the same objects or individuals at several points in time. The model allows the estimation of a group-average trajectory in the space of measurements. Random variations of this trajectory result from spatiotemporal transformations, which allow changes in the direction of the trajectory and in the pace at which trajectories are followed. The use of the tools of Riemannian geometry allows to derive a generic algorithm for any kind of data with smooth constraints, which lie therefore on a Riemannian manifold. Stochastic approximations of the Expectation-Maximization algorithm is used to estimate the model parameters in this highly non-linear setting.

The method is used to estimate a data-driven model of the progressive impairments of cognitive functions during the onset of Alzheimer's disease. Experimental results show that the model correctly put into correspondence the age at which each individual was diagnosed with the disease, thus validating the fact that it effectively estimated a normative scenario of disease progression. Random effects provide unique insights into the variations in the ordering and timing of the succession of cognitive impairments across different individuals.

More details in [30] and [31].

7.2. Joint Morphometry of Fiber Tracts and Gray Matter structures using Double Diffeomorphisms

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

This work proposes an atlas construction method to jointly analyse the relative position and shape of fiber tracts and gray matter structures. It is based on a double diffeomorphism which is a cascade of two diffeomorphisms. The first deformation acts only on the white matter keeping fixed the gray matter of the atlas. The resulting white matter, together with the gray matter, are then deformed by the second diffeomorphism which puts into correspondence the homologous anatomical structures across subjects. The first diffeomorphism makes the fiber bundles slide on the fixed gray matter revealing the variability in structural connectivity within the population, namely both the changes in the connected areas and in the geometry of the pathway of the tracts. Fiber bundles are approximated with weighted prototypes using the metric of weighted currents. The algorithm is based on a Bayesian framework which allows the automatic estimation of the covariance matrix of deformation parameters and of the noise variance of each structure. This approach is applied to patients with Tourette syndrome and controls showing a variability in the structural connectivity of the left cortico-putamen circuit.

More details in [26].

7.3. Bayesian Mixed Effect Atlas Estimation with a Diffeomorphic Deformation Model

Participants: Stanley Durrleman [Correspondant], Stéphanie Allassonniere, Estelle Kuhn.

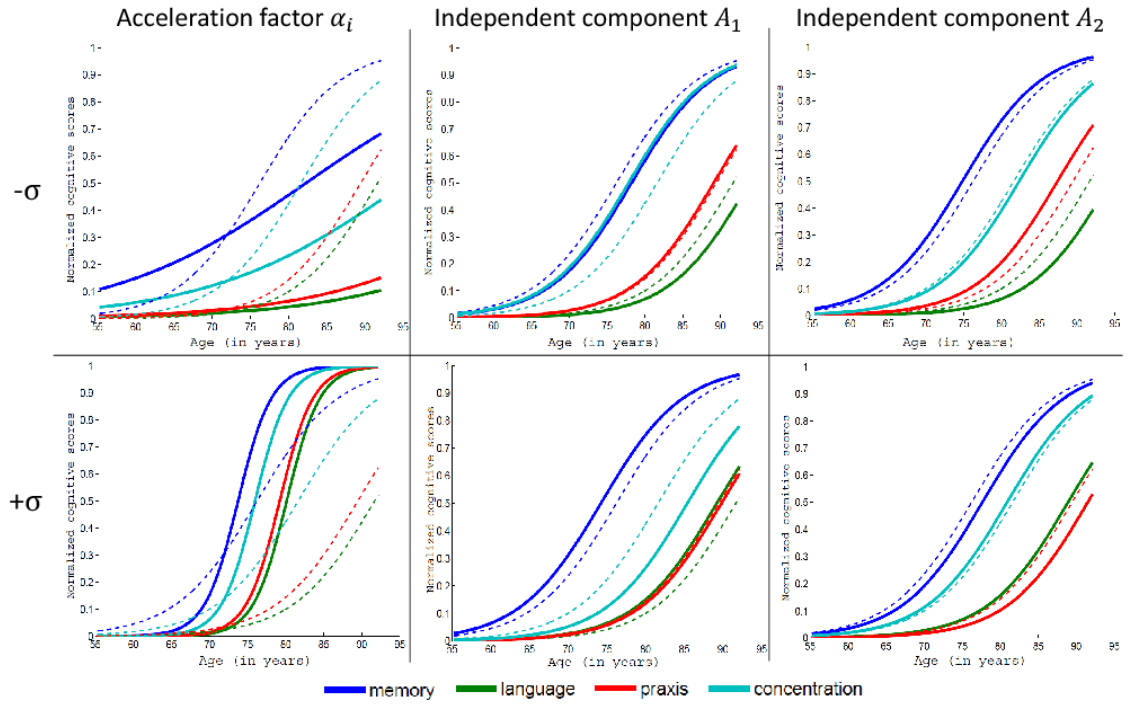


Figure 1. Disease progression model obtained from neuropsychological assessments of 248 patients observed at multiple times (from 3 to 11 times) who converted from Mild Cognitive Impairment stage to Alzheimer's disease during the observation period. Dashed lines represent the average scenario of disease progression (same in all plots). Solid lines represent the variability of this scenario within the observed population in terms of pace of disease progression (left) and relative timing and ordering of the decline of cognitive functions (middle and right).

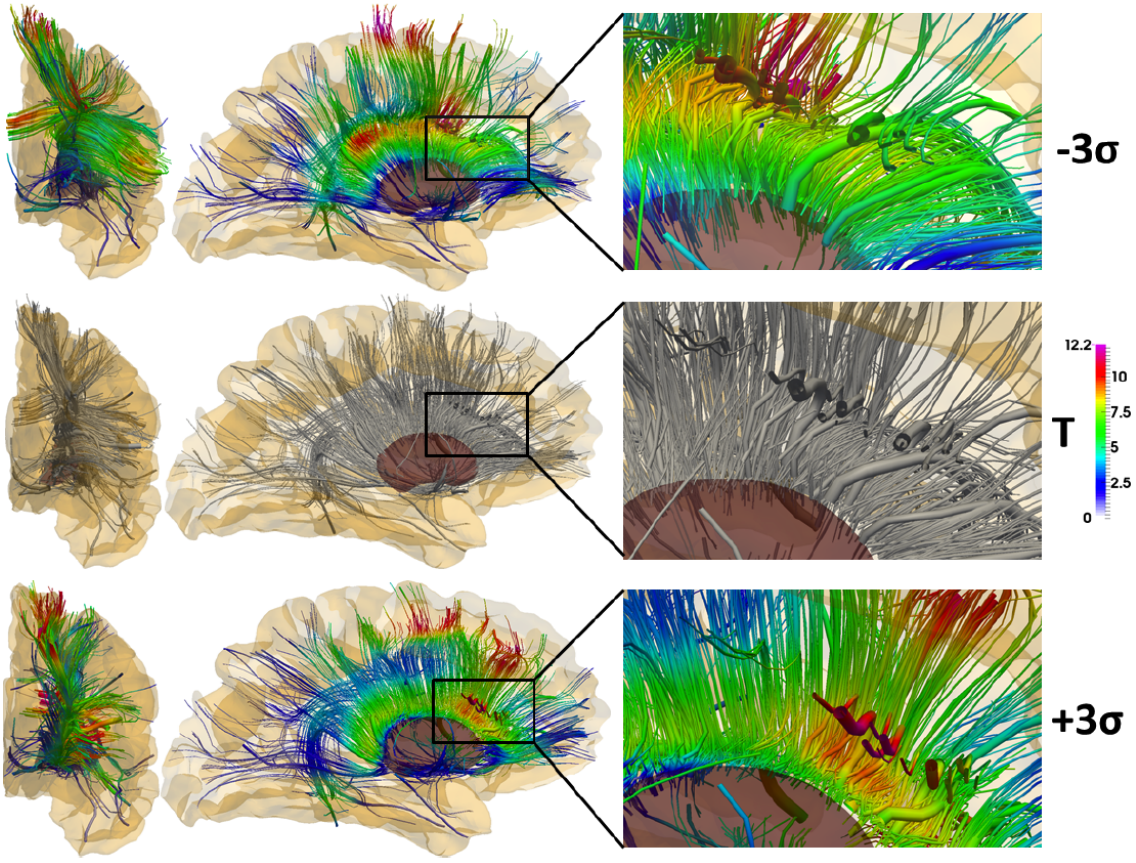


Figure 2. Estimation of a virtual representation of brain structure from anatomical and diffusion images of 3 patients with Gilles de la Tourette syndrome and 2 control subjects. Deformation of the white matter fiber bundle along the first mode of variability is shown while the estimated grey matter frame is kept fixed. Colors refer to the magnitude of displacement during deformation.

In this work, we introduced a diffeomorphic constraint on the deformations considered in the deformable Bayesian Mixed Effect (BME) Template model. Our approach is built on a generic group of diffeomorphisms, which is parameterized by an arbitrary set of control point positions and momentum vectors. This enables us to estimate the optimal positions of control points together with a template image and parameters of the deformation distribution which compose the atlas. We propose to use a stochastic version of the Expectation-Maximization (EM) algorithm where the simulation is performed using the Anisotropic Metropolis Adjusted Langevin Algorithm (AMALA). We propose also an extension of the model including a sparsity constraint to select an optimal number of control points with relevant positions. Experiments are carried out on the USPS database, on mandibles of mice, and on 3D murine dendrite spine images.

More details in [2].

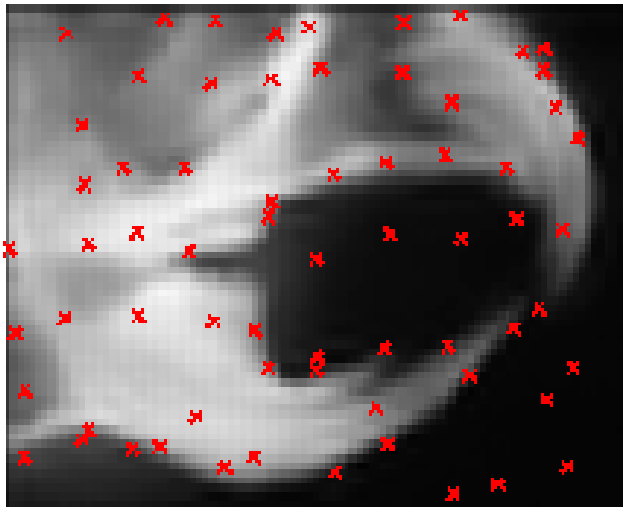


Figure 3. Template image of mouse mandible obtained from 36 X-ray image using 70 control points.

7.4. A sub-Riemannian modular approach for diffeomorphic deformations

Participants: Barbara Gris [Correspondant], Stanley Durrleman, Alain Trouvé.

We develop a generic framework to build large deformations from a combination of base modules. These modules constitute a dynamical dictionary to describe transformations. The method, built on a coherent sub-Riemannian framework, defines a metric on modular deformations and characterises optimal deformations as geodesics for this metric. We present a generic way to build local affine transformations as deformation modules, and display examples.

More details in [27].

7.5. Results of a multicenter randomized placebo-controlled clinical trial in prodromal Alzheimer's disease

Participants: Bruno Dubois, Marie Chupin, Harald Hampel, Simone Lista, Enrica Cavedo, Bernard Croisille, Guy Louis Tisserand, Jacques Touchon, Alain Bonafé, Pierre-Jean Ousset, Amir Ait Ameur, Olivier Rouaud,

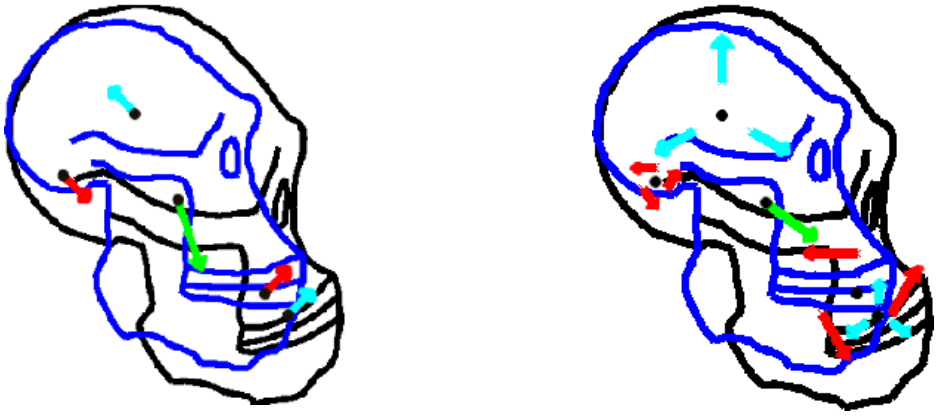


Figure 4. Initial position of deformation modules and their control parameters (left) leads to the construction of local scaling (cyan), rotation (red) and translation (green) (right), which combine together to deform the blue shape into the black one.

Frédéric Ricolfi, Alain Viguetto, Florence Pasquier, Christine Delmaire, Mathieu Ceccaldi, Nadine Girard, Carole Dufouil, Stéphane Lehéricy, Isabelle Tonelli, Françoise Duveau, Olivier Colliot, Line Garnero, Marie Sarazin, Didier Dormont [Correspondant].

Our team coordinated neuroimage acquisition and analysis of a multicenter randomized placebo-controlled clinical trial aiming to assess the efficacy of donepezil in prodromal Alzheimer's disease. Subjects underwent two brain magnetic resonance imaging scans (baseline and final visit). The primary efficacy outcome was the annualized percentage change (APC) of total hippocampal volume (left + right) measured by the software (see Section SACHA 6.3) developped by our team. Two-hundred and sixteen only subjects were randomized across 28 French expert clinical sites. In the per protocol population (placebo = 92 and donepezil = 82), the donepezil group exhibited a significant reduced rate of hippocampal atrophy (APC= -1.89%) compared with the placebo group (APC= -3.47%), P <.001. There was no significant difference in neuropsychological performance between treatment groups. A 45% reduction of rate of hippocampal atrophy was observed in prodromal AD following 1 year of treatment with donepezil compared with placebo.

This new approach opens interesting perspectives for the evaluation of treatments in neurodegenerative diseases.

More details in [12].

7.6. Sulcal morphology as a new imaging marker for the diagnosis of early onset Alzheimer's disease

Participants: Lorraine Hamelin, Bruno Dubois, Marie Chupin, Olivier Colliot [Correspondant], Marie Sarazin.

We investigated the utility of sulcal width measures in the diagnosis of Alzheimer's disease (AD). Sixty-six biologically confirmed AD patients (positive amyloid positron emission tomography [PET] and/or AD cerebrospinal fluid profile) were contrasted to 35 controls with negative amyloid PET. Patients were classified into prodromal or dementia stages as well as into late onset (LOAD, n = 31) or early onset (EOAD, n = 35) subgroups according to their age of onset. An automated method was used to calculate sulcal widths and hippocampal volumes (HV). In EOAD, the greatest ability to differentiate patients from age-matched controls, regardless of severity, was displayed by sulcal width of the temporoparietal cortex. In this region, diagnosis

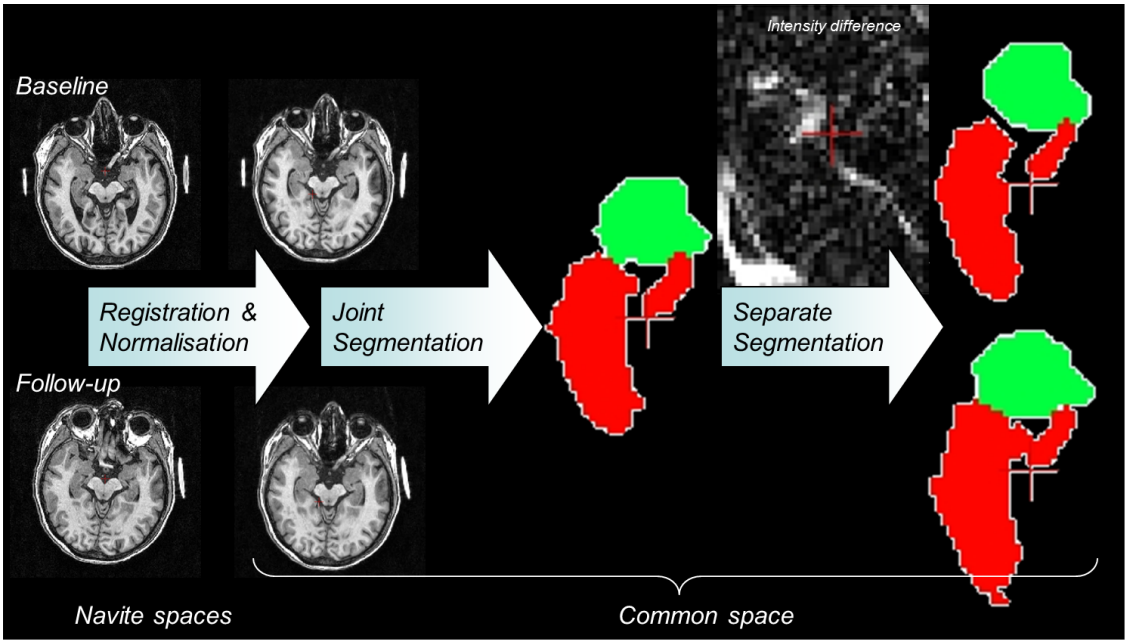


Figure 5. Hippocampus longitudinal segmentation method illustrating preliminary registration of the baseline and final visit magnetic resonance imaging (MRI) scans in a common space followed by normalization of the intensities of both scans. The baseline and final visit MRI scans were then segmented jointly. The resulting segmentation was then used as an initialization of separate segmentations while keeping the two segmentations consistent between the two time-points.

accuracy was better than the HV, especially at prodromal stage. In LOAD, HV provided the best discrimination power from age-matched controls. In conclusion, sulcal width measures are better markers than the HV for identifying prodromal AD in patients aged <65 years. In contrast, in older patients, the risk of over-diagnosis from using only sulcal enlargement is important.

More details in [14].

7.7. Imaging Markers of the Presymptomatic GRN Disease

Participants: Paola Caroppo, Stanley Durrleman, Alexandre Routier, Olivier Colliot [Correspondant], Alexis Brice, Isabelle Le Ber.

The preclinical stage of frontotemporal lobar degeneration (FTLD) is not well characterized. We conducted a brain metabolism (FDG-PET) and structural (cortical thickness) study to detect early changes in asymptomatic GRN mutation carriers (aGRN+) that were evaluated longitudinally over a 20-month period. At baseline, a left lateral temporal lobe hypometabolism was present in aGRN+ without any structural changes. Importantly, this is the first longitudinal study and, across time, the metabolism more rapidly decreased in aGRN+ in lateral temporal and frontal regions. The main structural change observed in the longitudinal study was a reduction of cortical thickness in the left lateral temporal lobe in carriers (Figure 6). A limit of this study is the relatively small sample (n=16); nevertheless, it provides important results. First, it evidences that the pathological processes develop a long time before clinical onset, and that early neuroimaging changes might be detected approximately 20 years before the clinical onset of disease. Second, it suggests that metabolic changes are detectable before structural modifications and cognitive deficits. Third, both the baseline and longitudinal studies provide converging results implicating lateral temporal lobe as early involved in GRN disease. Finally, our study demonstrates that structural and metabolic changes could represent possible biomarkers to monitor the progression of disease in the presymptomatic stage toward clinical onset.

More details in [6].

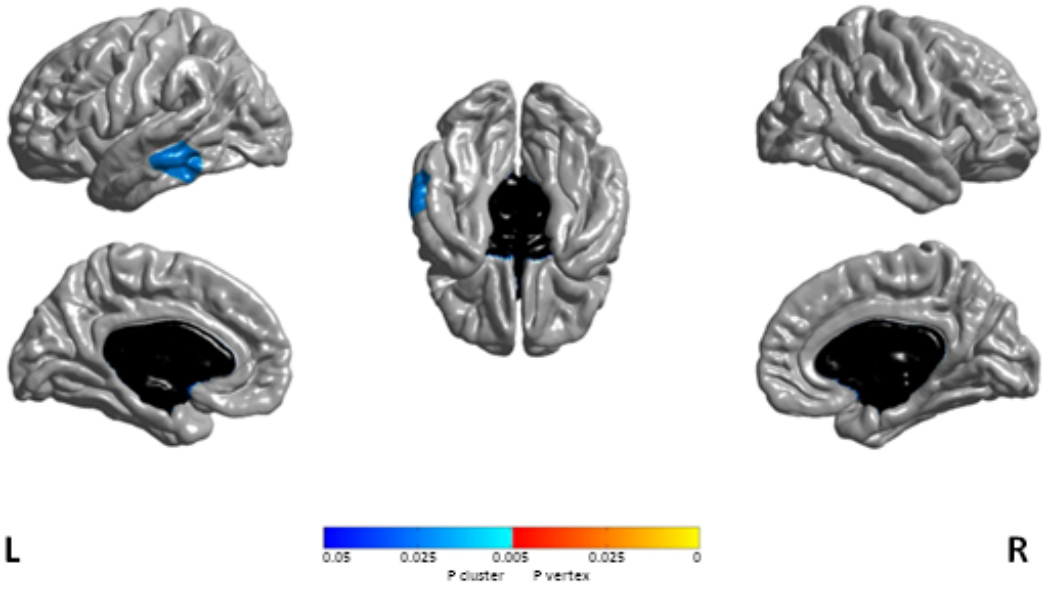


Figure 6. Cluster with significant cortical thickness changes in aGRN+ between the two time-points ($p < 0.05$ corrected). L, left; R, right.

7.8. Incomplete Hippocampal Inversions in healthy subjects: a comprehensive study of over 2000 participants

Participants: Claire Cury [Correspondant], Joan Glaunès, Dominique Hasboun, Fanny Cohen, Jorge Samper-González, Roberto Toro, Vincent Frouin, Gunter Schumann, Olivier Colliot.

The incomplete-hippocampal-inversion (IHI), also known as malrotation, is an atypical anatomical pattern of the hippocampus, which has been reported in healthy subjects in different studies. However, extensive characterization of IHI in a large sample has not yet been performed. Furthermore, it is unclear whether IHI are restricted to the medial-temporal lobe or are associated with more extensive anatomical changes. Here, we studied the characteristics of IHI in a community-based sample of 2008 subjects of the IMAGEN database and their association with extra-hippocampal anatomical variations. The presence of IHI was assessed on T1-weighted anatomical magnetic resonance imaging (MRI) using visual criteria. We assessed the association of IHI with other anatomical changes throughout the brain using automatic morphometry of cortical sulci. We found that IHI were much more frequent in the left hippocampus (left: 17%, right: 6%, χ^2-test, $p < 10^{-28}$). Compared to subjects without IHI, subjects with IHI displayed morphological changes in several sulci located mainly in the limbic lobe. Our results demonstrate that IHI are a common left-sided phenomenon in normal subjects and that they are associated with morphological changes outside the medial temporal lobe.

More details in [9].

7.9. Analysis of anatomical variability using diffeomorphic iterative centroid in patients with Alzheimer's disease

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

We proposed a new approach for template-based analysis of anatomical variability in populations, in the framework of Large Deformation Diffeomorphic Metric Mappings and mathematical currents. We proposef a fast approach in which the template is computed using an diffeomorphic iterative centroid method. Statistical analysis is then performed on the initial momenta that define the deformations between the centroid and each individual subject. We applied the approach to study the variability of the hippocampus in 134 patients with Alzheimer's disease (AD) and 160 elderly control subjects. We show that this approach can describe the main modes of variability of the two populations and can predict the performance to a memory test in AD patients.

More details in [8].

7.10. Innovation-based sparse estimation of functional connectivity from multivariate autoregressive models

Participants: Fabrizio de Vico Fallani [Correspondant], Stéphanie Allassonniere [Correspondant], Francois Deloche.

One of the main limitations of functional connectivity estimators of brain networks is that they can suffer from statistical reliability when the number of areas is large and the available time series are short. To estimate directed functional connectivity with multivariate autoregressive (MVAR) model on sparse connectivity assumption, we propose a modified Group Lasso procedure with an adapted penalty. Our procedure includes the innovation estimates as explaining variables. This approach is inspired by two criteria that are used to interpret the coefficients of the MVAR model, the Directed Transfer Function (DTF) and the Partial Directed Coherence (PDC). A causality measure can be deduced from the output coefficients which can be understood as a synthesis of PDC and DTF. We demonstrate the potential of our method and compare our results with the standard Group Lasso on simulated data.

More details in [25]

7.11. Lucid Dreaming in Narcolepsy

Participants: Pauline Daudet, Mario Chavez [Correspondant], Smaranda Leu-Semenescu, Jean-Louis Golmard, Isabelle Arnulf.

Lucid dreaming is the experience of being aware of dreaming while asleep and continuing to dream. Lucid dreams generally arise in REM sleep. Compared to non-lucid REM sleep, lucid REM sleep is associated with local frontal lobe EEG changes in the 40 Hz band, increased brain coherence, and increased activity on functional MRI in the bilateral precuneus, cuneus, parietal lobules, and prefrontal and occipito-temporal cortices, which may correspond to restored reflective consciousness. We decided to study the frequency and determinants of lucid dreaming in narcolepsy and to challenge patients' alleged ability to achieve lucid dreaming using sleep monitoring during nighttime and daytime sleep. Compared to 53 healthy controls, the 53 narcolepsy patients reported more frequent dream recall, nightmares and recurrent dreams. The frequency of cataplexy, hallucinations, sleep paralysis, dyssomnia, positivity, and the severity of sleepiness were similar in narcolepsy with and without lucid dreaming. The delta power in the electrode average, in delta, theta, and alpha powers in C4, and coherences between frontal electrodes were lower in lucid than non-lucid REM sleep in spectral EEG analysis. The duration of REM sleep was longer, the REM sleep onset latency tended to be shorter, and the percentage of atonia tended to be higher in lucid vs. non-lucid REM sleep; the arousal index and REM density and amplitude were unchanged. Our results suggest that narcoleptics have a high propensity for lucid dreaming without differing in REM sleep characteristics from people without narcolepsy. This also suggests that narcolepsy patients may provide useful information in future studies on the nature of lucid dreaming.

More details in [11]

7.12. An Algebraic Topological Method for Multimodal Brain Networks Comparisons

Participants: Tiago Simas, Mario Chavez [Correspondant], Pablo Rodriguez, Albert Diaz-Guilera.

Understanding brain connectivity is one of the most important issues in neuroscience. Nonetheless, connectivity data can reflect either functional relationships of brain activities or anatomical connections between brain areas. Although both representations should be related, this relationship is not straightforward. We have devised a powerful method that allows different operations between networks that share the same set of nodes, by embedding them in a common metric space, enforcing transitivity to the graph topology. Here, we apply this method to construct an aggregated network from a set of functional graphs, each one from a different subject. Once this aggregated functional network is constructed, we use again our method to compare it with the structural connectivity to identify particular brain regions that differ in both modalities (anatomical and functional). Remarkably, these brain regions include functional areas that form part of the classical resting state networks. We conclude that our method -based on the comparison of the aggregated functional network- reveals some emerging features that could not be observed when the comparison is performed with the classical averaged functional network.

More details in [23]

7.13. Steady state visual evoked potentials-based patient interface under breathing constraints

Participants: Xavier Navarro [Correspondant], Sebastien Campion, Fabrizio de Vico Fallani [Correspondant], Pierre Pouget, Thomas Similowski, Mathieu Raux, Mario Chavez.

Steady state visual evoked potentials (SSVEP) have been widely utilized in brain computer interfacing (BCI) in last years. In this paper, we present a study exploring the possibilities of SSVEP to manage the communication between patients suffering respiratory disorders and health care providers. By imposing different breathing constraints, five healthy subjects communicated their breathing sensations (breathing well/breathing bad) using a visual frequency tagging paradigm: two visual stimuli with different flickering frequencies (15 and 20 Hz) were simultaneously presented on a screen. Using electroencephalographic (EEG) signals from only three EEG electrodes, two spectral features were extracted by a spatial filter in a sliding window, then classified by an unsupervised algorithm based on k-medians. Average detection success rates were of 70% during breathing

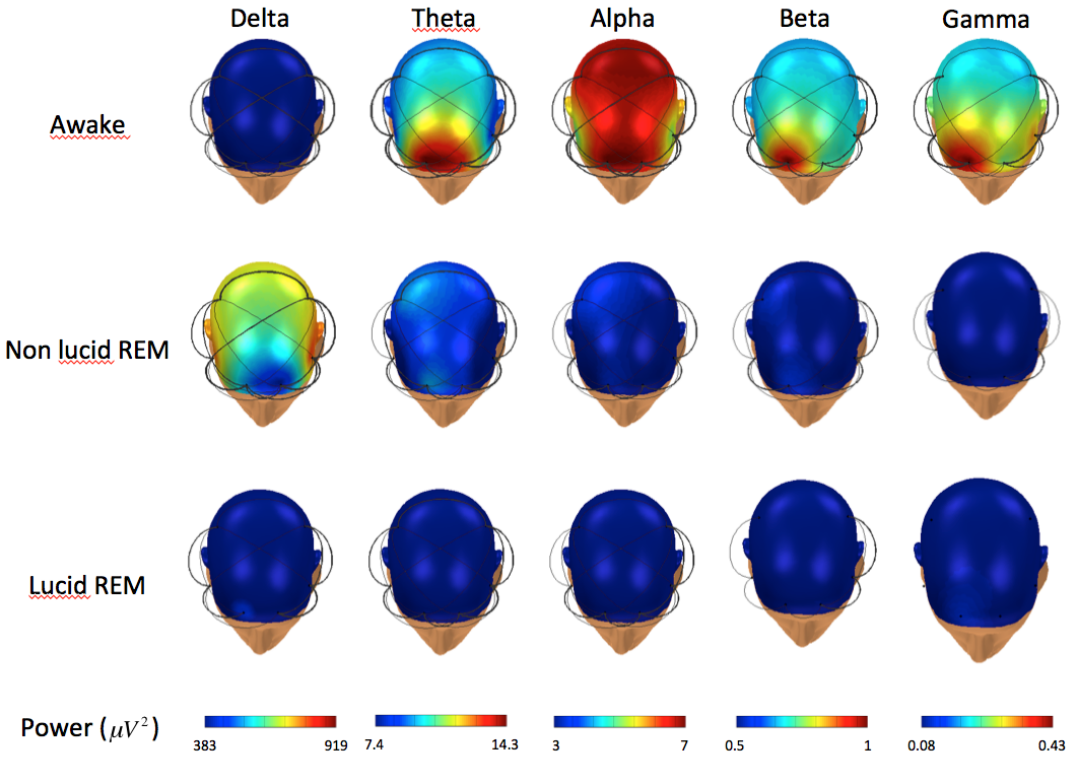


Figure 7. Topographical distribution (obtained by a spherical spline interpolation) of EEG spectral power during wakefulness (top row), non-lucid (middle row) and lucid (bottom row) REM sleep for different frequency bands. Signi cant couplings between the electrodes are indicated by the black links (the thickness is proportional to the coherence value). Colors from dark blue (lower EEG power) to dark red (higher EEG power) indicated for each EEG band in the Power line (bottom row).

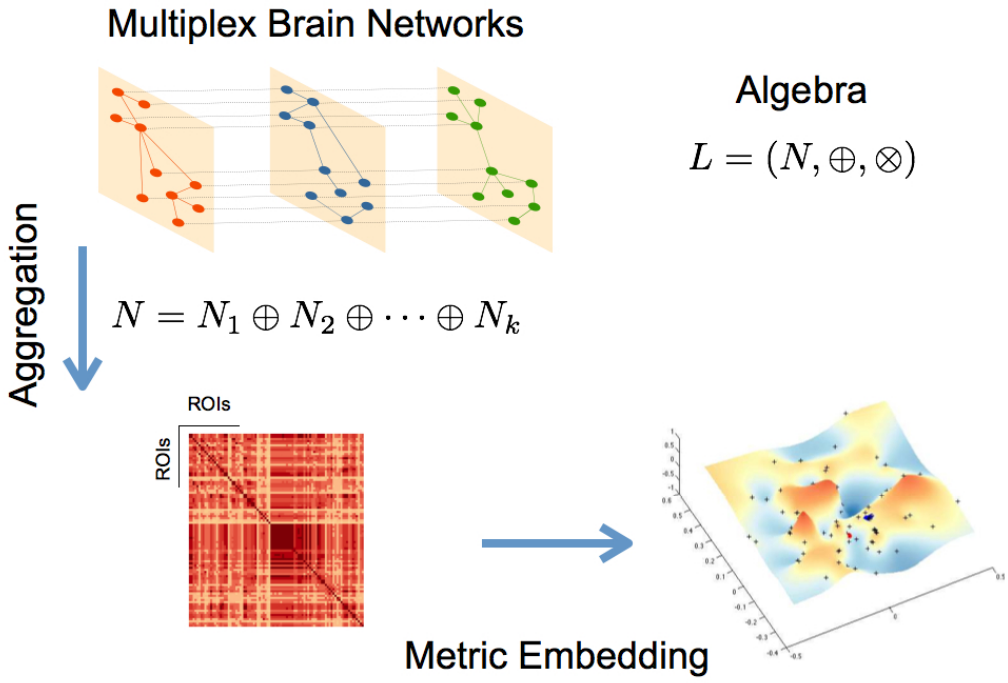


Figure 8. Schematic representation of the main steps for the described networks aggregation and metric embedding (defined here for the algebra L)

discomfort, and of 83% when subjects breathed comfortably. Results suggest that SSVEP-based BCI may be a promising choice to improve patient-caregiver communication in situations of breathing discomfort when verbal communication is difficult.

More details in [29]

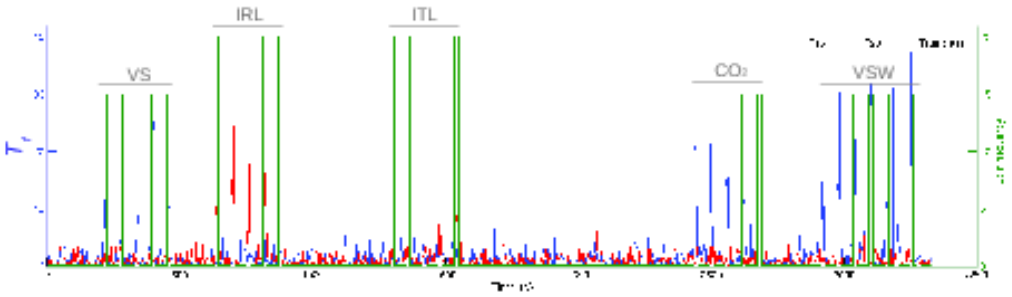


Figure 9. An example of T_f obtained after applying the spatial filters on 15 Hz (blue curve) and 20 Hz (red curve) during the experiment in subject 3. The statistics T_f reflects the signal-to-noise ratio at frequency f with respect to the no-stimulus power.

ASCLEPIOS Project-Team

6. New Results

6.1. Medical Image Analysis

6.1.1. Longitudinal Analysis and Modeling of Brain Development

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

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

Processing pipeline - brain development - adolescence - longitudinal analysis - non-rigid registration algorithm - extrapolation

1. We proposed and detailed a deformation-based morphometry computational framework, called Longitudinal Log-Demons Framework (LLDF), which estimates the longitudinal brain deformations from image data series, transports them in a common space and performs statistical group-wise analyses (see Fig. 1). This processing pipeline is based on freely available softwares and relies on the LCC log-Demons non-linear diffeomorphic registration algorithm with an additional modulation of the similarity term using a confidence mask to increase robustness with respect to brain boundary intensity artifacts.
2. The LLDF framework is applied to the study of longitudinal trajectories during adolescence, for which little is known. The aim of this project is to provide models of brain development during adolescence based on diffeomorphic registration parametrised by SVFs. Our study focused particularly on the link between sexual dimorphism and the longitudinal evolution of the brain. This work was done in collaboration with J.L. Martinot et H. Lemaître (Inserm U1000).

6.1.2. Inter-Operative Relocalization in Flexible Endoscopy

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

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

Computer Assisted Intervention, Barrett's Esophagus, Biopsy Relocalization, Electromagnetic tracking

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

In an earlier paper, we introduced the first approach to inter-operative relocalization using electromagnetic tracking system. In [21], we propose three incremental experiments to our approach. First, we analyse the error bounds of our system on synthetic data with a realistic noise model. Second, we provide a pseudo ground-truth on *in-vivo* pig data using an optical tracking system. Accuracy results obtained were consistent with the synthetic experiments despite uncertainty introduced due to breathing motion, and remain inside acceptable error margins according to medical experts. Finally, a third experiment was designed using data from pigs to simulate a real task of biopsy site relocalization, and evaluated by ten experts. It clearly demonstrated the benefit of our system towards assisted guidance by improving the biopsy site retrieval rate from 47.5% to 94%.

This inter-operative relocalization framework was then further extended in [53] to provide a constrained image based search as shown in Fig. 2 to obtain the best view point match to the live view. Within this context, we investigate the effect of (a) the choice of feature descriptors and colour-space, (b) filtering of uninformative frames and (c) endoscopic modality, for view point localization. Our experiments indicate an improvement in the best view-point retrieval rate to [92%, 87%] from [73%, 76%] (in our previous approach) for Narrow band imaging and white-light endoscopic image modalities.

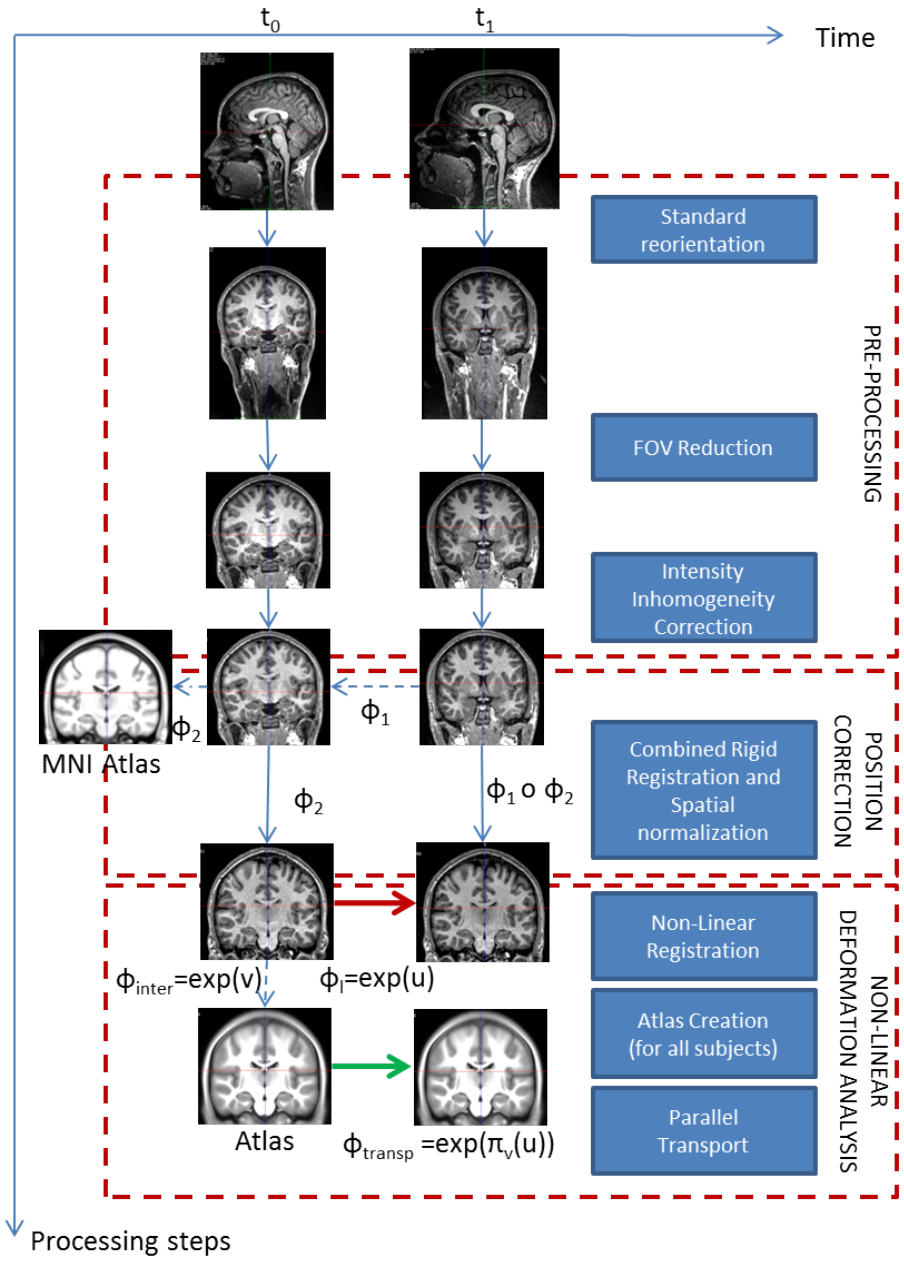


Figure 1. Proposed processing pipeline for longitudinal analysis: the pipeline is composed of three major steps. Starting with raw images, we first pre-process them, then correct the spatial position differences to end up with the longitudinal deformations for each subject in the atlas space.

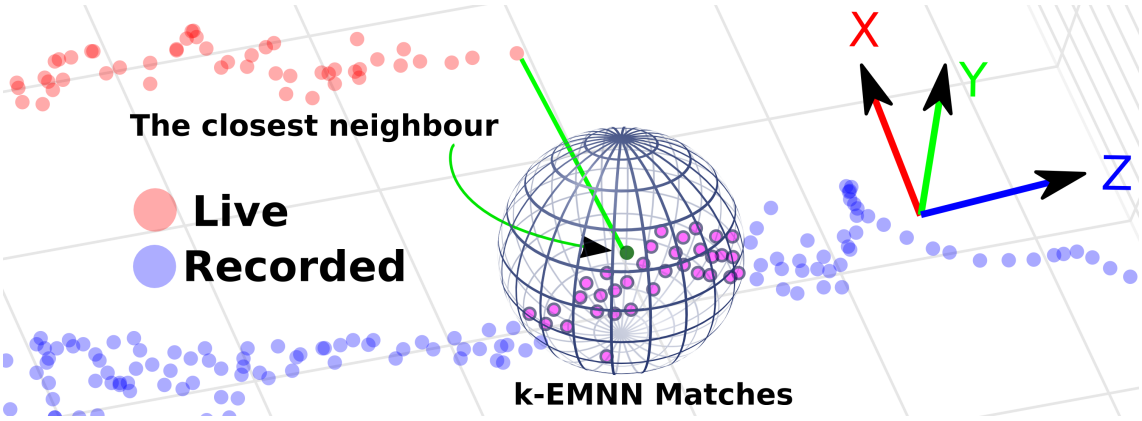


Figure 2. ElectroMagnetic tracker Nearest Neighbour (EMNN) matches. Firstly, the EMNN is obtained using the 3D position match. Then, in a radius r around this match, all the points on the trajectory are considered as the k-EMNN matches. For the images in the k-EMNN matches, scene matching is performed.

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

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

This work is supported by the National Association for Research in Technology (ANRT) through the CIFRE Grant 2013-1165 and Oticon Medical (Vallauris). Part of this work is also funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images). This work is done in collaboration with the Department of Ear Nose Throat Surgery (IUFC, Nice) and the Nice University Hospital (CHU).

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

- We designed a parametric shape model of the intracochlear anatomy with anatomical prior learned from temporal bones high-resolution images, see Fig. 3 .
- We evaluated the cochleostomy location regarding two surgical approaches (endaural compared to conventional posterior tympanotomy) [20].

6.1.4. Structured sparse Bayesian modelling for non-rigid registration and cardiac motion tracking

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

This work has been partly supported by the Inria – Microsoft Research Joint Center and by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Non-rigid Registration - Structured Sparse Bayesian Learning - Automatic Relevance Determination - Reversible-jump Markov Chain Monte Carlo - Cardiac Motion Tracking - Uncertainty Quantification

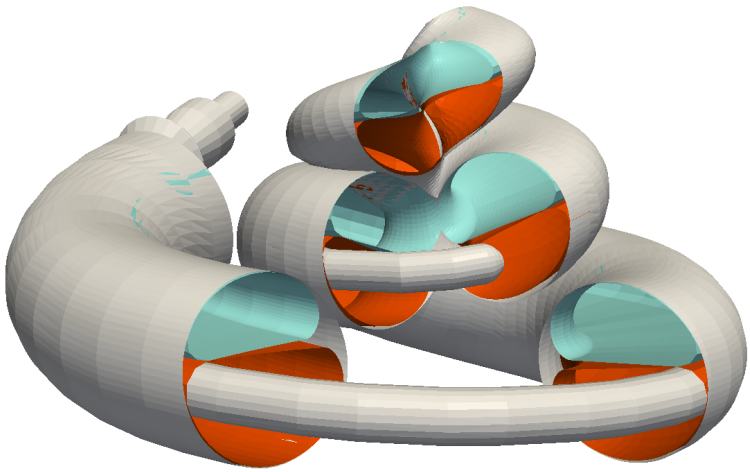


Figure 3. Cochlear implant electrode-array with respect to scala tympani (red) and scala vestibuli (blue).

We developed a generic structured sparse Bayesian model of image registration with three main contributions: an extended image similarity term, the automated tuning of registration parameters and uncertainty quantification. We proposed an approximate inference scheme that is tractable on 4D clinical data. We demonstrated the performance of our approach on cine MR, tagged MR and 3D Ultra Sound cardiac images, and showed state-of-the-art results on benchmark datasets evaluating accuracy of motion and strain.

Moreover, we evaluated the quality of uncertainty estimates returned by the approximate inference scheme. We compare the predictions of the approximate scheme with those of an inference scheme developed on the grounds of reversible jump Markov Chain Monte-Carlo [94](see Fig. 4). We provided more insight into the theoretical properties of the sparse structured Bayesian model and into the empirical behaviour of both inference schemes.

This work is described in the PhD manuscript of Loïc Le Folgoc, defended at Université Nice Sophia Antipolis, 2015 [6].

6.1.5. *Image Segmentation and Synthesis of brain tumor MR images*

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

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

Brain, MRI, Glioma, Patch-based Segmentation, Image synthesis

The segmentation of glioblastoma, the most severe case of brain tumors, is a crucial step for diagnostic assessment and therapy planning. In order to perform the manual delineation of the tumor compartments, the clinicians have to concurrently screen multi-channel 3D MRI, which makes the process both time-consuming and subject to inter-expert delineation variability.

We have developed 2 contributions for the analysis of MR brain tumor images:

- A patch-based multi-atlas automatic glioma segmentation algorithm[13]. Unlike prior work on patch-based multi-atlas segmentation, our approach does not assume any prior knowledge about the location of pathological structures (no local search window).

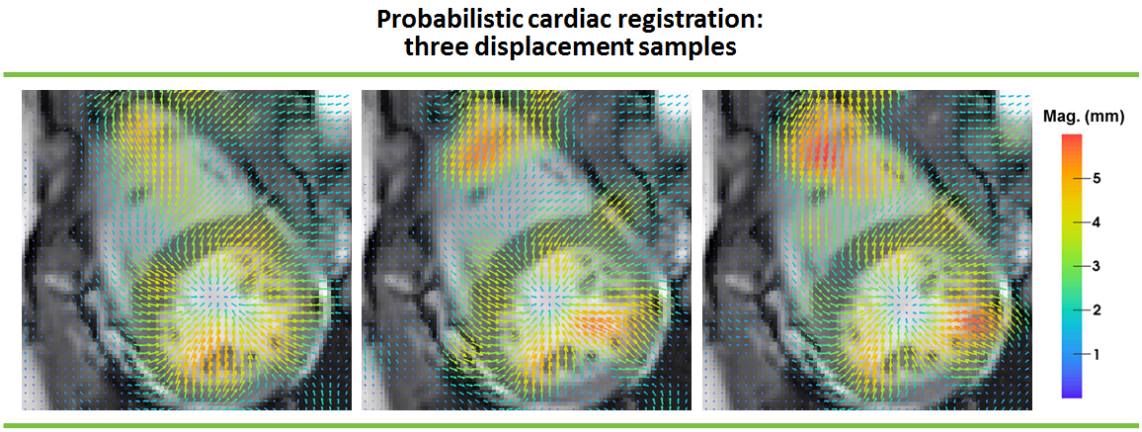


Figure 4. Three displacement fields sampled from the same posterior distribution of coefficients associated with the registration of two cardiac images.

- A patch-based image synthesis algorithm (see Fig.5) [4], which generates multi-sequence MR images of the brain with glioma from a single label image. The synthesis of images may be useful to benchmark segmentation algorithms or to increase the size of annotated medical image databases.

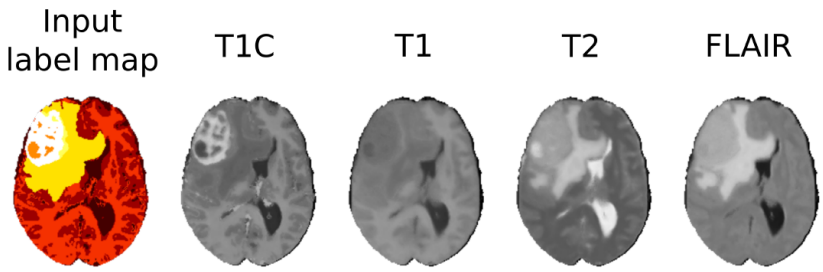


Figure 5. Synthesis of high-grade glioma MR image from a single label map.

6.1.6. Infarct localization from myocardial deformation

Participants: Nicolas Duchateau [Correspondent], Maxime Sermesant.

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

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

- We investigate new methods for predicting the location of myocardial infarcts from local wall deformation [31], which is useful for risk stratification from routine examinations such as 3D echocardiography. In a broader perspective, this project also aims at determining relevant biomarkers to study cardiac function [54], and eventually at combining several of those markers in an efficient manner [59].

- Non-linear dimensionality reduction aims at estimating the Euclidean space of coordinates encoding deformation patterns, and is combined with multi-scale kernel regressions to infer the low-dimensional coordinates and the infarct location of new cases.
- These concepts were tested on 500 synthetic cases with infarcts of random extent, shape, and location, generated from a realistic electromechanical model. Our prediction goes beyond the current diagnosis of infarct either achieved at the global or segmental level, and significantly outperforms the clinically-used thresholding of the deformation patterns.

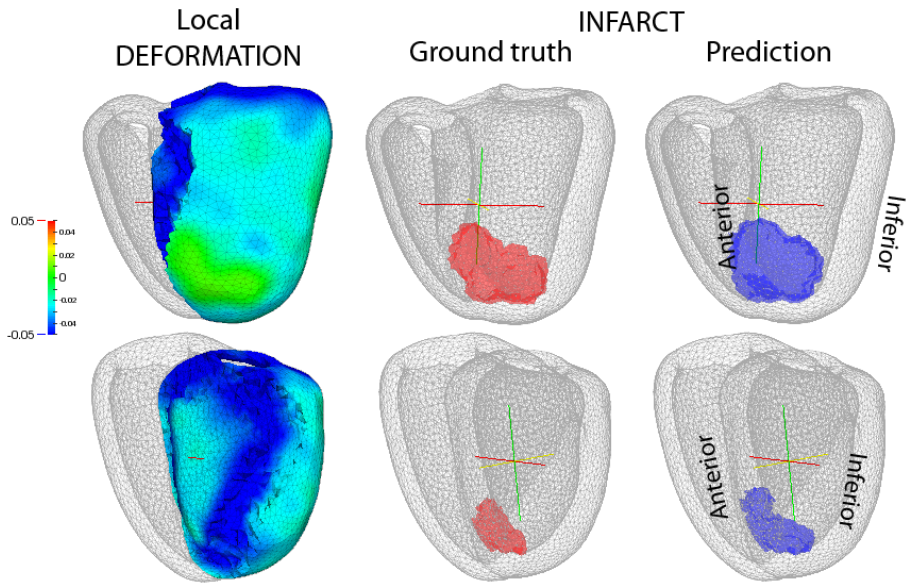


Figure 6. Examples of myocardial deformation patterns, ground truth infarct location, and estimated infarct location.

6.2. Computational Anatomy

6.2.1. Geometric generative model of organ shapes: statistical properties of template shape estimation

Participants: Nina Miolane [Correspondent, Inria - Stanford], Xavier Pennec [Inria], Susan Holmes [Stanford].

This work is conducted jointly with the Department of Statistics of Stanford, in the context of the associated team GeomStats and the FSCIS (France-Stanford Center for Interdisciplinary Studies) fellowship of Nina Miolane.

template, atlas, consistency, estimation theory, Expectation-Maximization algorithm, shapes, quotient space, lie group, sub-Riemannian, in-painting, neuro-geometry, visual cortex, diffusion

This work focuses on the interaction between statistics and geometry, for applications in Medical Imaging. The first part deals with a generative model of (organ) shapes and, more precisely, on the estimation of the mean shape or template. The second part of this work surveys and unveils the mathematical framework needed to extend Neurogeometry, used in 2D Computer Vision, to applications in 3D imaging.

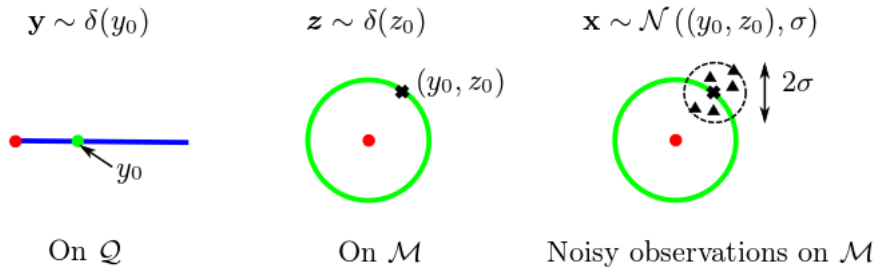


Figure 7. A schematic representation of the generative model of organ shapes.

- In the first part, we define a geometric statistical framework of an organ shapes generative model (see Figure 7). This is done through the differential geometry of quotient spaces.
- Then, we interpret the computation of the mean organ shape (or template) through the max-max algorithm, as an approximate maximum likelihood estimation in this framework.
- Finally, we study the statistical properties of the template computed with this procedure. More precisely, we show that the estimation is inconsistent and that the inconsistency cannot be neglected when the real template is close to the singularity of the quotient space at the scale of the ambient noise on the images [44].
- In the second part, the particularities of a 3D neurogeometry are highlighted with respect to the 2D case. They rely on the fact that 2D neurogeometry is inspired by the primary visual cortex, which codes for our 2D visual field (our retina is 2D). Imagining a 3D visual field or a 3D retina would give rise to a 3D neurogeometry.
- The conceptual framework of a 3D neurogeometry is more subtle, and a new level of mathematical structures arises (see Figure 8). Thus, inpainting (sub-Riemannian diffusion) have to be generalized.
- Applications for in-painting or super-resolution in 3D medical images are described [43].

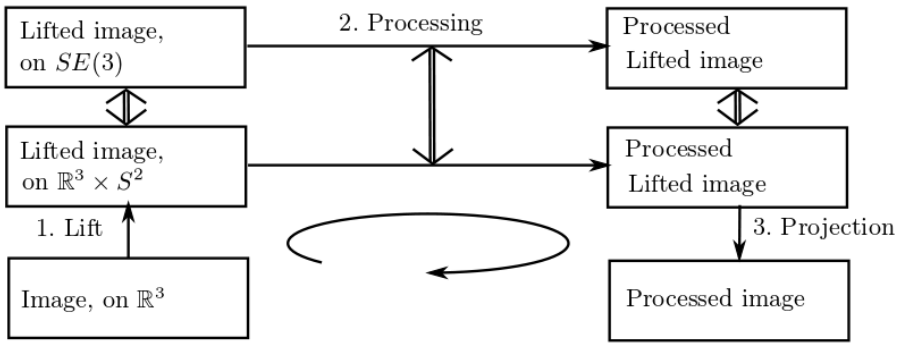


Figure 8. The 3 steps of an image processing pipeline in 3D neurogeometry.

6.2.2. Compact representation of longitudinal deformations

Participants: Raphaël Sivera [Correspondent], Hervé Delingette, Nicholas Ayache.

This work is supported by a PhD fellowship from the University Nice Sophia Antipolis and by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Longitudinal modeling, Learning in manifolds, Structured sparsity.

The analysis of dynamic or longitudinal series of medical images is important to better understand the observed evolutions of the organs but also to provide robust computer aided diagnosis tools. This analysis can be performed through a reduced representation of geometric transformations capturing the deformation between 2 time points.

In the context of cardiac motion analysis, we proposed a framework to represent arbitrary diffeomorphisms described as Stationary Velocity Fields (SVF) in a low dimensional linear space (see fig. 9).

To this end, we first improved the Inverse Scaling and Squaring (ISS) algorithm from [83] to transform displacement fields into SVFs. Second, through a structured sparse decomposition of these deformations over the cardiac cycle, we provided a preliminary approach for comparing trajectories of cine-MR images between two patients.

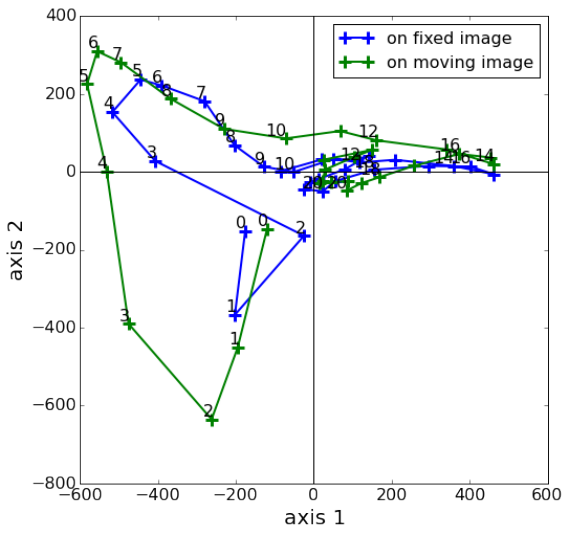


Figure 9. Trajectories of two registered cardiac cycles projected on a 2D space using dimensionality reduction tools.

6.2.3. Statistical analysis of heart motion

Participants: Marc-Michel Rohé [Correspondent], Nicolas Duchateau, Maxime Sermesant, Xavier Pennec.

This work is partly supported by the FP7 European project MD-Paedigree and by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Statistical analysis, Registration, Reduced order models, Machine learning

This work aims at developing statistical tools to analyze cardiac motion. In particular, we are interested in approximating complex motion models with few parameters or modes that are clinically relevant (reduced models). To this end, we have introduced a polyaffine cardiac motion model that reduces the deformation parameters to a few interpretable parameters, and the most important modes to represent the variability seen in a population are automatically selected. We then performed a group-wise statistical analysis, which relates the model parameters to clinical indices specific to a given pathology. This method was used to classify a population of healthy/infarcted hearts [48] (see Fig. 10), as well as to study cardiac motion of adolescents with cardiomyopathies within the European project "MD-Paedigree".

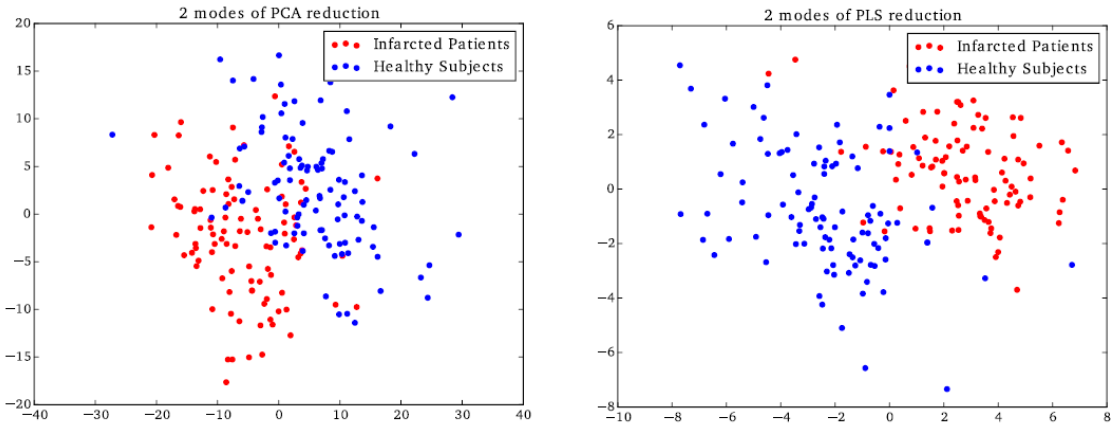


Figure 10. Projection of healthy/infarcted patients cardiac motions on two modes extracted from PCA (left) and PLS (right) methods.

6.2.4. Statistical Learning via Synthesis of Medical Images

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

This work is partly supported by a grant from Microsoft Research-Inria Joint Centre and by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

statistical learning, synthesis

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

We tackled the problem of better exploiting existing training sets, via a smart modeling of the image space, and applying conventional random forests using guided bagging [99]. Synthesis of complex data, such as cardiac diffusion images (DTI), was also done, with a refined version of [98].

Then, we tackled the problem of exploiting *Geometry in Data*, via intrinsic representations of shapes and data [27]. Spectral decomposition (Fig. 11) of shapes provides a new intrinsic framework for synthesizing complex shapes such as cerebral surfaces [35], and describing functions efficiently on these complex surfaces. This framework establishes the basics for machine learning of surface data [36]. An early application was conducted on retinotopy [57] (the study of functions in the visual cortex).

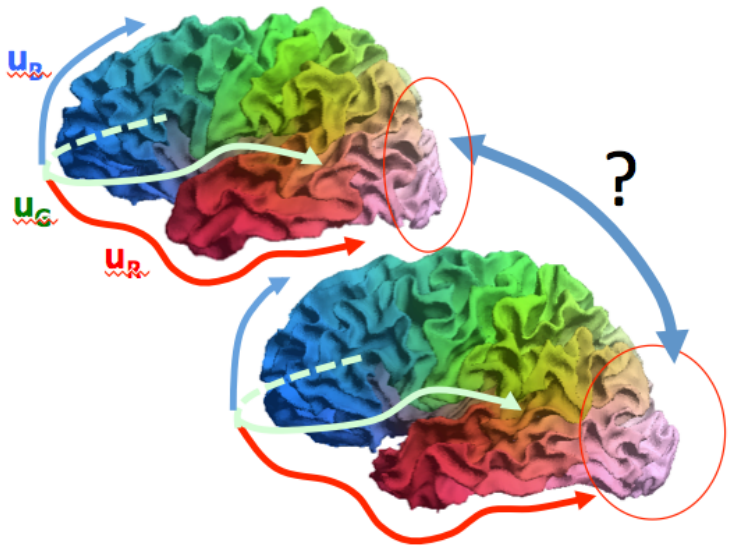


Figure 11. Spectral Representation – How to find an intrinsic representation of shapes for finding fast point correspondences and for learning surface data.

6.2.5. Consistency of the estimation of the template in quotient spaces

Participants: Loïc Devilliers [Correspondent], Stéphanie Allassonnière [Ecole Polytechnique], Xavier Pennec.

Template estimation, Fréchet mean, quotient spaces

In [24], we studied the estimation of the template (the mean shape of our data) when the data is transformed by unknown group elements. In the case of a finite group acting isometrically on a linear space, we proved that the estimation of the template using the Fréchet mean in the quotient space is not always consistent.

6.3. Computational Physiology

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

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

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

Radiofrequency Ablation Modeling, Patient-Specific Simulation, Lattice Boltzmann Method, Computer Model, Computational Fluid Dynamics, Heat Transfer, Cellular Necrosis, Parameter Estimation, Therapy Planning, Liver, Pre-clinical Study, Medical Imaging

Radio Frequency Ablation (RFA) is a minimally invasive therapy suited for liver tumor ablation. However, a patient-specific predictive tool is needed to plan and guide the treatment.

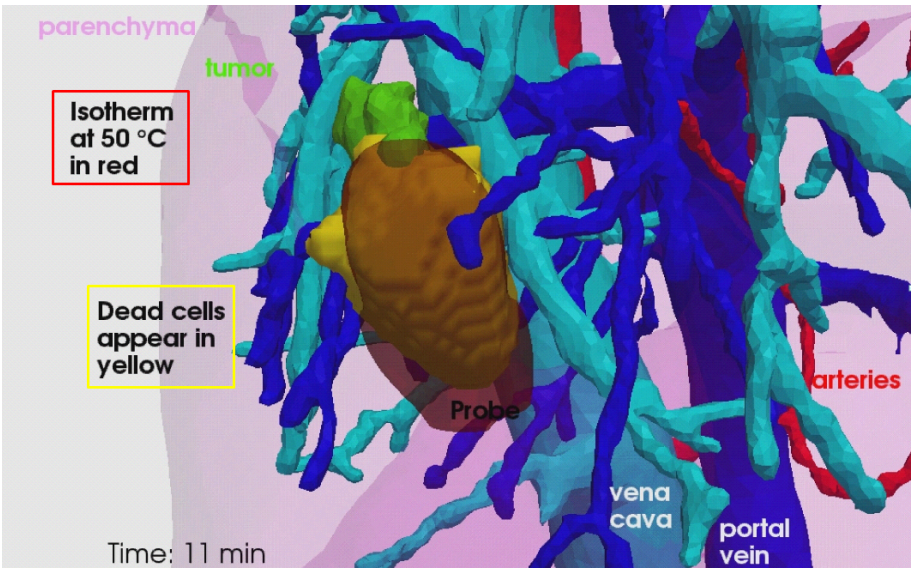


Figure 12. Computed isotherm at 50°C and computed necrosis in a subject-specific geometry.

We developed a computational framework for patient-specific planning of RFA, which includes the following contributions:

- A detailed computational model of the biophysical mechanisms (heat transfer, cellular necrosis, hepatic blood flow) involved in RFA of abdominal tumors based on patient images.
- A new implementation of the bio-heat equations coupled with a cellular necrosis model using the Lattice Boltzmann Method (LBM) on Graphics Processing Units (GPU), which allows near real-time computation.
- A Computational Fluid Dynamics (CFD) and porous media solver using LBM algorithm to compute the patient-specific blood flow in the hepatic circulatory system and the blood flow distribution inside the parenchyma.
- A complete patient-specific geometry including hepatic venous and arterial circulation system.
- The automatic estimation of the main parameters of the model. Two personalization strategies tested and evaluated on clinical and pre-clinical data.
- The evaluation of the proposed model on a clinical dataset of ten patients (see Fig. 12).
- The evaluation on a preclinical dataset of five swines from a comprehensive experimental set-up specially designed for RFA model validation.

The proposed RFA model and its evaluation on clinical data are presented in [10], and the evaluation of the RFA model on pre-clinical data is presented in [25]. The proposed model, its personalisation and its evalutation against clinical and preclinical data are presented in Chloé Audigier's PhD thesis [1].

6.3.2. Learning Cardiac Ablation Targets from Image Data and Simulation

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

This work was supported by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Cardiac electrophysiology modeling, Intracardiac electrogram modeling, Machine learning, Radiofrequency ablation planning, electroanatomical mapping, local abnormal ventricular activities (LAVA)

Ventricular radiofrequency ablation can have a critical impact on preventing sudden cardiac arrest but it is challenging due to a highly complex arrhythmogenic substrate. We used advanced delayed enhanced-MR image characteristics in a machine learning framework to predict the presence of local abnormal ventricular activities (LAVA). Furthermore, we enriched these predictions through MR image-based patient-specific electrophysiology simulations and the modeling of normal and LAVA-like intracardiac electrograms using the dipole approach and their incorporation in the learning framework (see Fig. 13). Confidence maps can then be generated and analyzed prior to RFA to guide the intervention.

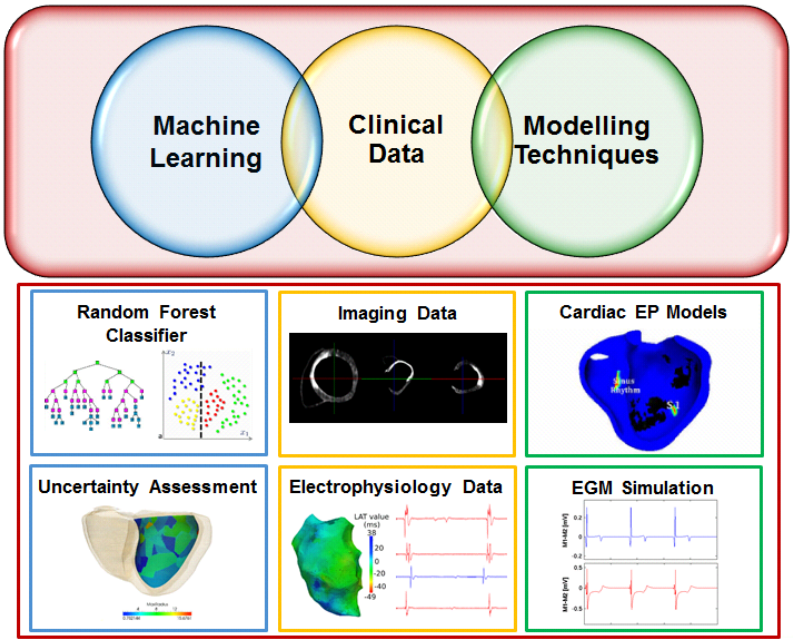


Figure 13. Coupled learning and simulation framework for LAVA identification.

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

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

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

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

- We developed a framework to generate patient specific multiple time-point images based on our biophysical model of brain deformation due to atrophy in Alzheimer's Disease (AD)[34]. From two time-point brain MRIs of a patient, we used the framework to simulate a new time-point brain MRI with the personalized atrophy for the patient (see Fig. 14).
- The framework can be used to evaluate methods that study the temporal relationships, ordering and co-evolution of atrophy in different structures of the brain.

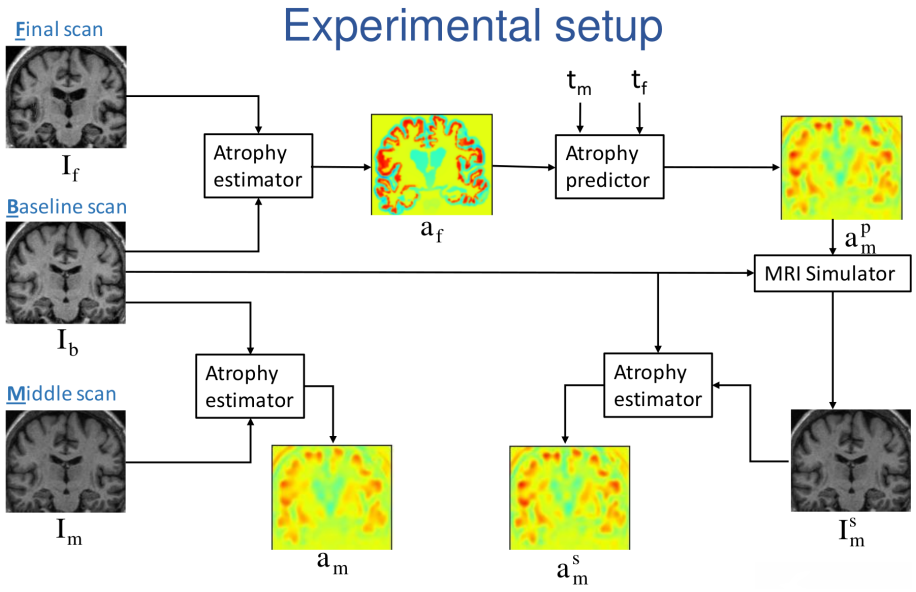


Figure 14. From a given baseline scan and a final scan at time t_f, a middle scan at time t_m is predicted and simulated. An experimenatal setup that allows comparing the atrophy (a^s_m) estimated from a simulated middle-scan against the atrophy (a_m) estimated from the real middle scan is shown.

6.3.4. Brain Tumor Growth Personalization and Segmentation Uncertainty

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

This work is carried out between Asclepios research group, Inria Sophia Antipolis, France and the Department of Radiation Oncology of the Massachusetts General Hospital, Boston, USA. It is supported by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Tumor growth, radiotherapy, modeling, personalization, segmentation, uncertainty, Bayesian

- We developed a method for the Bayesian personalization of a brain tumor growth model based on clinical MRIs [37] (see Fig. 15 Left).
- We proposed an algorithm for the sampling of several plausible segmentations, based on a single clinical segmentation (see Fig. 15 Right). This allows the uncertainty quantification of the radiotherapy plan based on several sample clinical target volumes [38]. This paper received the Young Scientist Award at the 2015 MICCAI conference in Munich, Germany.

6.3.5. Uncertainty quantification in personalised Cardiac models. Application to myocardial fiber uncertainty.

Participants: Roch-Philippe Molléro [Correspondent], Dominik Neumann [Siemens], Marc-Michel Rohé, Hervé Delingette, Maxime Sermesant, Xavier Pennec, Nicholas Ayache, Tommaso Mansi [Siemens].

This work was partly supported by the FP7 European project MD-Paedigree and was done in collaboration with Siemens Corporate Technology, Erlangen, Germany and Siemens Corporate Research, Princeton, New Jersey.

Heart Modeling - Myocardial Fibers - Biophysical Simulation - Uncertainty Quantification

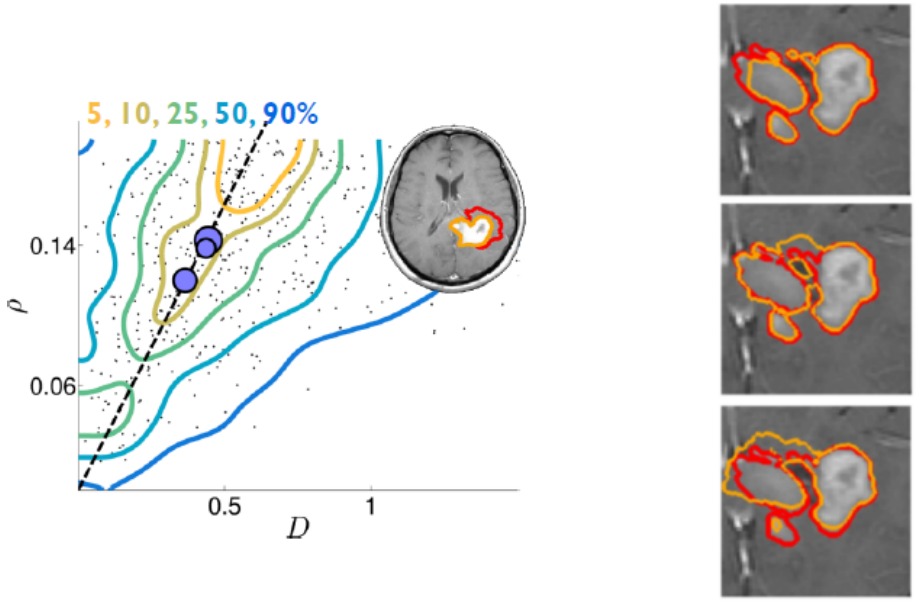


Figure 15. (Left) Bayesian personalization of a glioblastoma patient. Isocontours of the posterior probability of the diffusion parameter D and the proliferation parameter ρ; (Right) Different sampled plausible segmentations in orange based on the clinical segmentation in red.

Computational models of the heart are of increasing interest for clinical applications due to their discriminative and predictive power. However, the personalisation step to go from a generic model to a patient-specific one is still challenging. In particular, it is still difficult to quantify the uncertainty on the estimated parameters and predicted values.

We developped a pipeline (see Fig. 16) to evaluate the impact of myocardial fibre uncertainty on the personalisation of an electromechanical model of the heart from ECG and medical images:

- We studied how to estimate the variability of the fibre architecture among a given population (from a myocardial fiber atlas).
- Then, we showed the variability of the personalised simulations, in electrophysiology (EP) and in biomechanics, with respect to the principal variations of the fibres.
- Finally, we discussed how the variations in this population of fibres impact the parameters of the personalised simulations.

This work led to a paper at FIMH 2015 conference in Maastricht, The Netherlands [45].

6.3.6. Non-invasive personalisation of the electrical heart model

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

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

Cardiac Modelling, Personalised Simulation, Electrical Simulation

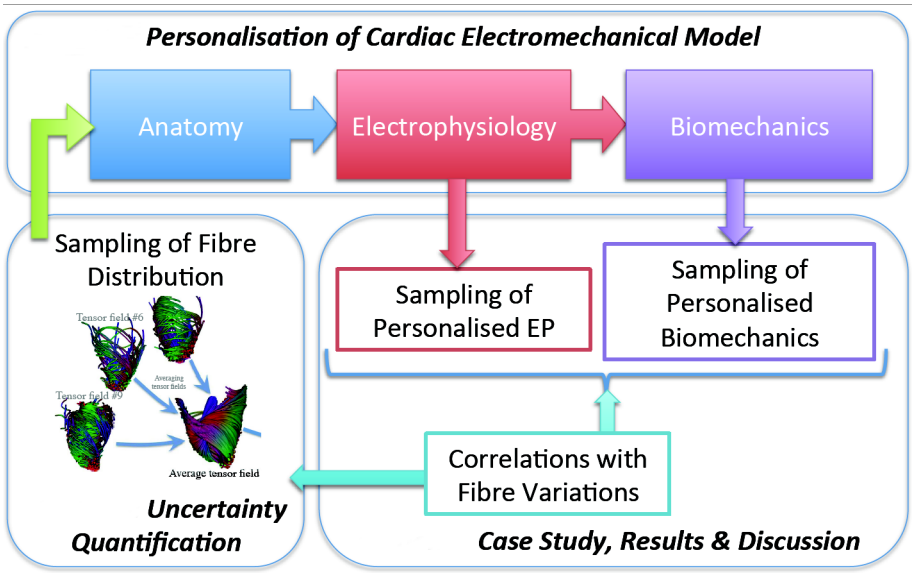


Figure 16. Global scheme of fibre variability propagation along the personalisation pipeline.

Non-invasive cardiac electrical data has been acquired at St Thomas' Hospital, London. It consists in Body Surface Potential Mapping (BSPM), which are recordings of the electrical potential on several locations on the surface of the torso (see Fig. 17). From BSPMs and MRI data of the heart, we aim at personalizing the electrical propagation model of the heart previously developed within the Asclepios team.

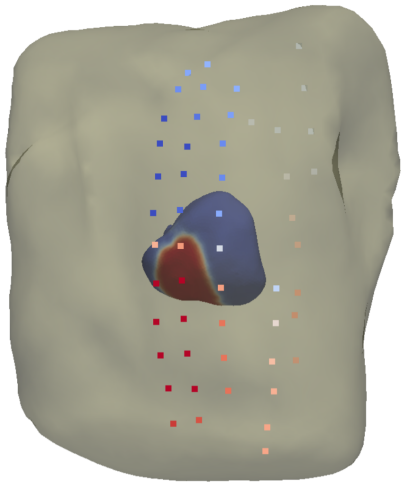


Figure 17. Torso representation used for personalizing cardiac electrical parameters from non-invasive observations.

ATHENA Project-Team

6. New Results

6.1. Modeling in Diffusion MRI

6.1.1. Improving fiber alignment in HARDI by combining contextual PDE flow with constrained spherical deconvolution

Participants: Jorg Portegies [Department of Mathematics and Computer Science, Eindhoven University of Technology], Rutger Fick, Gonzalo Sanguinetti [Department of Mathematics and Computer Science, Eindhoven University of Technology], Shephan Meesters [Department of Mathematics and Computer Science, Eindhoven University of Technology], Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Remco Duits [Department of Mathematics and Computer Science, Eindhoven University of Technology].

We propose two strategies to improve the quality of tractography results computed from diffusion weighted magnetic resonance imaging (DW-MRI) data. Both methods are based on the same PDE framework, defined in the coupled space of positions and orientations, associated with a stochastic process describing the enhancement of elongated structures while preserving crossing structures. In the first method we use the enhancement PDE for contextual regularization of a fiber orientation distribution (FOD) that is obtained on individual voxels from high angular resolution diffusion imaging (HARDI) data via constrained spherical deconvolution (CSD). Thereby we improve the FOD as input for subsequent tractography. Secondly, we introduce the fiber to bundle coherence (FBC), a measure for quantification of fiber alignment. The FBC is computed from a tractography result using the same PDE framework and provides a criterion for removing the spurious fibers. We validate the proposed combination of CSD and enhancement on phantom data and on human data, acquired with different scanning protocols. On the phantom data we find that PDE enhancements improve both local metrics and global metrics of tractography results, compared to CSD without enhancements. On the human data we show that the enhancements allow for a better reconstruction of crossing fiber bundles and they reduce the variability of the tractography output with respect to the acquisition parameters. Finally, we show that both the enhancement of the FODs and the use of the FBC measure on the tractography improve the stability with respect to different stochastic realizations of probabilistic tractography. This is shown in a clinical application: the reconstruction of the optic radiation for epilepsy surgery planning.

This work has been published in [19]

6.1.2. Sparse reconstruction challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use?

Participants: Lipeng Ning [Brigham and Women's Hospital, Harvard Medical School, Boston], Frederik Laun [German Cancer Research Institute], Yogesh Rathi [Brigham and Women's Hospital, Harvard Medical School, Boston], Thinhinane Megherbi [ParIMed Team, LRPE, USTHB, Algiers], Mario Zuccheli [Dpt of Computer Science, University of Verona], Gloria Menegaz [Dpt of Computer Science, University of Verona], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Aurobrata Ghosh, Rutger Fick, Rachid Deriche.

Diffusion magnetic resonance imaging (dMRI) is the modality of choice for investigating in-vivo white matter connectivity and neural tissue architecture of the brain. The diffusion-weighted signal in dMRI reflects the diffusivity of water molecules in brain tissue and can be utilized to produce image-based biomarkers for clinical research. Due to the constraints on scanning time, a limited number of measurements can be acquired within a clinically feasible scan time. In order to reconstruct the dMRI signal from a discrete set of measurements, a large number of algorithms have been proposed in recent years in conjunction with varying sampling schemes, i.e., with varying b-values and gradient directions. Thus, it is imperative to compare the performance of these reconstruction methods on a single data set to provide appropriate guidelines to

neuroscientists on making an informed decision while designing their acquisition protocols. For this purpose, the SPArse Reconstruction Challenge (SPARC) was held along with the workshop on Computational Diffusion MRI (at MICCAI 2014) to validate the performance of multiple reconstruction methods using data acquired from a physical phantom. A total of 16 reconstruction algorithms (9 teams) participated in this community challenge. The goal was to reconstruct single b-value and/or multiple b-value data from a sparse set of measurements. In particular, the aim was to determine an appropriate acquisition protocol (in terms of the number of measurements, b-values) and the analysis method to use for a neuroimaging study. The challenge did not delve on the accuracy of these methods in estimating model specific measures such as fractional anisotropy (FA) or mean diffusivity, but on the accuracy of these methods to fit the data. This work presents several quantitative results pertaining to each reconstruction algorithm. The conclusions in this work provide a valuable guideline for choosing a suitable algorithm and the corresponding data-sampling scheme for clinical neuroscience applications.

This work has been published in [18].

6.1.3. A Unifying framework for spatial and temporal diffusion in dMRI

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

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

This work has been published in [26]

6.1.4. Exploiting the phase in dMRI for microstructure recovery: Towards axonal tortuosity via asymmetric diffusion processes

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

Microstructure recovery procedures via Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) usually discard the signal's phase, assuming symmetry in the underlying diffusion process. In this work, we propose to recover the Ensemble Average Propagator (EAP) directly from the complex DW signal in order to describe also eventual diffusional asymmetry, thus obtaining an asymmetric EAP. The asymmetry of the EAP is then related to tortuosity of undulated white matter axons, which are found in pathological scenarios associated with axonal elongation or compression. We derive a model of the EAP for this geometry and quantify its asymmetry. Results show that the EAP obtained when accounting for the DW signal's phase provides useful microstructural information in such pathological scenarios. Furthermore, we validate these results in-silico through 3D Monte-Carlo simulations of white matter tissue that has experienced different degrees of elongation/compression.

This work has been published in [35]

6.1.5. A temperature phantom to probe the Ensemble Average Propagator asymmetry: an in-silico study

Participants: Marco Pizzolato, Demian Wassermann, Tanguy Duval [Institute of Biomedical Engineering, Polytechnique Montréal, Montréal], Jennifer Campbell [Montreal Neurological Institute, MCGill University], Timothé Boutelier [Olea Medical, La Ciotat], Julien Cohen-Adad [Institute of Biomedical Engineering, Polytechnique Montréal, Montréal], Rachid Deriche.

The detection and quantification of asymmetry in the Ensemble Average Propagator (EAP) obtained from the Diffusion-Weighted (DW) signal has been shown only for theoretical models. EAP asymmetry appears for instance when diffusion occurs within fibers with particular geometries. However the quantification of EAP asymmetry corresponding to such geometries in controlled experimental conditions is limited by the difficulty of designing fiber geometries on a micrometer scale. To overcome this limitation we propose to adopt an alternative paradigm to induce asymmetry in the EAP. We apply a temperature gradient to a spinal cord tract to induce a corresponding diffusivity profile that alters locally the diffusion process to be asymmetric. We simulate the EAP and the corresponding complex DW signal in such a scenario. We quantify EAP asymmetry and investigate its relationship with the applied experimental conditions and with the acquisition parameters of a Pulsed Gradient Spin-Echo sequence. Results show that EAP asymmetry is sensible to the applied temperature-induced diffusivity gradient and that its quantification is influenced by the selected acquisition parameters.

This work has been published in [36]

6.1.6. How to get more out of a clinically feasable 64 gradient dMRI acquisition: multi-shell versus single-shell

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

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

This work has been published in [29].

6.2. Tissue Microstructures features recovery & applications

6.2.1. Laplacian-regularized MAP-MRI : Improving axonal caliber estimation

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

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

This work has been published in [27]

6.2.2. Using 3D-SHORE and MAP-MRI to obtain both tractography and microstructural contrasts from a clinical dMRI acquisition

Participants: Rutger Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

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

This work has been published in [28]

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

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

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

This work has been accepted at the conference ISBI 2016.

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

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

The recovery of microstructure-related features of the brain's white matter is a current challenge in diffusion MRI. To robustly estimate these important features from diffusion MRI data, we propose to analytically regularize MAP-MRI's coefficient estimation using the norm of the Laplacian of the reconstructed signal. We first compare our approach, which we call MAPL, with competing state-of-the-art functional basis approaches. We show that it outperforms the original MAP-MRI implementation and the recently proposed modified Spherical Polar Fourier (mSPF) basis with respect to signal fitting, EAP and ODF reconstruction in noisy, sparsely sampled data of a physical phantom with reference gold standard data. Then, to reduce the variance of parameter estimation using multi-compartment tissue models, we propose to use MAPL's signal

fitting and extrapolation as a preprocessing step. We study the effect of MAPL on the estimation of axon diameter using a simplified Axcaliber model and axonal dispersion using the Neurite Orientation Dispersion and Density Imaging (NODDI) model. We show the positive effect of using it as a preprocessing step in estimating and reducing the variances of these parameters in the Corpus Callosum of six different subjects of the MGH Human Connectome Project. Finally we correlate the estimated axon diameter, dispersion and restricted volume fractions with Fractional Anisotropy (FA) and clearly show that changes in FA significantly correlate with changes with all estimated parameters. Overall, we illustrate the potential of using a well-regularized functional basis together with multi-compartment approaches to recover important microstructure tissue parameters with much less variability, thus contributing to the challenge of better understanding microstructure-related features of the brain's white matter.

This work has been submitted to the journal NeuroImage.

6.3. Towards microstructural based tractography

6.3.1. AxTract: Microstructure-driven tractography based on the Ensemble Average Propagator

Participants: Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Rutger Fick, Maxime Descoteaux [SCIL Lab., Sherbrooke University], Demian Wassermann, Rachid Deriche.

In this work, we propose a novel method to simultaneously trace brain white matter (WM) fascicles and estimate WM microstructure characteristics. Recent advancements in diffusion-weighted imaging (DWI) allow multi-shell acquisitions with b-values of up to 10,000 s/mm2 in human subjects, enabling the measurement of the ensemble average propagator (EAP) at distances as short as 10 micro-meters. Coupled with continuous models of the full 3D DWI signal and the EAP such as Mean Apparent Propagator (MAP) MRI, these acquisition schemes provide unparalleled means to probe the WM tissue in vivo. Presently, there are two complementary limitations in tractography and microstructure measurement techniques. Tractography techniques are based on models of the DWI signal geometry without taking specific hypotheses of the WM structure. This hinders the tracing of fascicles through certain WM areas with complex organization such as branching, crossing, merging, and bottlenecks that are indistinguishable using the orientation-only part of the DWI signal. Microstructure measuring techniques, such as AxCaliber, require the direction of the axons within the probed tissue before the acquisition as well as the tissue to be highly organized. Our contributions are twofold. First, we extend the theoretical DWI models proposed by Callaghan et al. to characterize the distribution of axonal calibers within the probed tissue taking advantage of the MAP-MRI model. Second, we develop a simultaneous tractography and axonal caliber distribution algorithm based on the hypothesis that axonal caliber distribution varies smoothly along a WM fascicle. To validate our model we test it on in-silico phantoms and on the HCP dataset

This work has been published in [23]

6.3.2. Studying white matter tractography reproducibility through connectivity matrices

Participants: Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion-weighted imaging is often used as a starting point for in vivo white matter (WM) connectivity to reconstruct potential WM pathways between brain areas. In this study, we investigate the reproducibility of the connectivity matrix, resulting from different tractography parameters. We vary the number of streamlines used to construct the matrix in cortical to cortical connectivity and analyze its effects. We also compare the effect of probabilistic and deterministic local streamline tractography algorithms, seeding both from the WM and from WM-grey matter interface.

This work has been published in [31]

6.3.3. Structural connectivity reproducibility through multiple acquisitions

Participants: Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

dMRI is often used to reconstruct white matter pathways between brain areas for in vivo brain connectivity. In this study, we investigate the reproducibility and the specificity of connectivity matrices in cortico-cortical connectivity using probabilistic and deterministic streamline tractography, seeding from both the white matter and the white matter-grey matter interface.

This work has been published in [30]

6.4. Computational Diffusion MRI

6.4.1. Robust and efficient linear registration of white-matter fascicles in the space of streamlines

Participants: Eleftherios Garyfallidis [SCIL Lab., Sherbrooke University], Omar Cepeda [SCIL Lab., Sherbrooke University], Demian Wassermann, Maxime Descoteaux [SCIL Lab., Sherbrooke University].

The neuroscientific community is very much interested in analyzing specific white matter bundles like the arcuate fasciculus, the corticospinal tract, or the recently discovered Aslant tract to study sex differences, lateralization and many other connectivity applications. For this reason, experts spend time manually segmenting these fascicles and bundles using streamlines obtained from diffusion MRI tractography. However, to date, there are very few computational tools available to register these fascicles directly so that they can be analyzed and their differences quantified across populations. In this work, we introduce a novel, robust and efficient framework to align bundles of streamlines directly in the space of streamlines. We call this framework Streamline-based Linear Registration. We first show that this method can be used successfully to align individual bundles as well as whole brain streamlines. Additionally, if used as a piecewise linear registration across many bundles, we show that our novel method systematically provides higher overlap (Jaccard indices) than state-of-the-art nonlinear image-based registration in the white matter. We also show how our novel method can be used to create bundle-specific atlases in a straightforward manner and we give an example of a probabilistic atlas construction of the optic radiation. In summary, Streamline-based Linear Registration provides a solid registration framework for creating new methods to study the white matter and perform group-level tractometry analysis.

This work has been published in [14]

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

Participants: Brahim Belaoucha, Maureen Clerc, Théodore Papadopoulo.

In this work, we present a method that aims at parcellating the cortical surface from individual anatomy. The parcellation is obtained using the mutual nearest neighbor criteria to obtain regions that have similar fiber distribution. The later is obtained by applying a probabilistic tractography on the diffusion MRI (dMRI), a non-invasive modality allowing the access to the structural information of the cortical surface. The proposed algorithm is compared to some of the atlases that can be found in the literature. We show that these atlases have lower similarity of fibers distributions than the proposed algorithm.

This work has been accepted at the conference ISBI 2016.

6.5. Clinical and Neurocognitive Applications of Diffusion MRI

6.5.1. Plasticity of left perisylvian white-matter tracts is associated with individual differences in math learning brain structure and function

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

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

This work has been published in [16]

6.5.2. Prefrontal cortex white matter tracts in prodromal Huntington disease

Participants: Joy T. Matsui [Iowa University], Jatin G. Vaidya [Iowa University], Demian Wassermann [Iowa University], Regina Eunyoung Kim [Iowa University], Vincent A. Magnotta [Iowa University], Hans J. Johnson [Iowa University], Jane S. Paulsen [Iowa University], Predict-Hd Investigators And Coordinators Of The Huntington Study Group [NIH].

Huntington disease (HD) is most widely known for its selective degeneration of striatal neurons but there is also growing evidence for white matter (WM) deterioration. The primary objective of this research was to conduct a large-scale analysis using multi-site diffusion-weighted imaging (DWI) tractography data to quantify diffusivity properties along major prefrontal cortex WM tracts in prodromal HD. Fifteen international sites participating in the PREDICT-HD study collected imaging and neuropsychological data on gene-positive HD subjects without a clinical diagnosis (i.e. prodromal) and gene-negative control subjects. The anatomical prefrontal WM tracts of the corpus callosum (PFCC), anterior thalamic radiations (ATR), inferior fronto-occipital fasciculi (IFO), and uncinate fasciculi (UNC) were identified using streamline tractography of DWI. Within each of these tracts, tensor scalars for fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity coefficients were calculated. We divided prodromal HD subjects into three CAG-age product (CAP) groups having Low, Medium, or High probabilities of onset indexed by genetic exposure. We observed significant differences in WM properties for each of the four anatomical tracts for the High CAP group in comparison to controls. Additionally, the Medium CAP group presented differences in the ATR and IFO in comparison to controls. Furthermore, WM alterations in the PFCC, ATR, and IFO showed robust associations with neuropsychological measures of executive functioning. These results suggest that long-range tracts essential for cross-region information transfer show early vulnerability in HD and may explain cognitive problems often present in the prodromal stage.

This work has been published in [17]

6.6. Perfusion MRI

6.6.1. Perfusion MRI deconvolution with delay estimation and non-negativity constraints

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

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

This work has been published in [34]

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

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

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

This work has been accepted at the conference ISBI 2016.

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

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

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

This work has been submitted to the conference ISMRM 2016.

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

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

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

This work has been submitted to the conference ISMRM 2016.

6.7. MEG, EEG and cochlear modeling

6.7.1. MEM-diffusion MRI framework to solve MEEG inverse problem

Participants: Brahim Belaoucha, Jean-Marc Lina, Maureen Clerc, Théodore Papadopoulo.

In this work, we present a framework to fuse information coming from diffusion magnetic resonance imaging (dMRI) with Magnetoencephalography (MEG)/ Electroencephalography (EEG) measurements to reconstruct the activation on the cortical surface. The MEG/EEG inverse-problem is solved by the Maximum Entropy on the Mean (MEM) principle and by assuming that the sources inside each cortical region follow Normal distribution. These regions are obtained using dMRI and assumed to be functionally independent. The source reconstruction framework presented in this work is tested using synthetic and real data. The activated regions for the real data is consistent with the literature about the face recognition and processing network.

This work was published in the proceedings of the conference EUSIPCO 2015 [22].

6.7.2. MEG/EEG reconstruction in the reduced source space

Participants: Brahim Belaoucha, Théodore Papadopoulo.

Obtaining the brain activity with the distributed source model from MEG or EEG measurements is ill-posed problem due to the high number of unknowns compared to the number of measurements. The idea of ths work is to reduce the solution space size from the number of sources to a smaller space. Assuming that sources inside each functional region have equal activation allows us to reduce the number of columns in the leadfield matrix from the number of nodes S required to model the cortex to a number of regions K, which is much smaller. These regions are obtained from a dMRI parcellation-based region growing algorithm. A region is assumed to contain sources that have similar fibers distribution. To obtain a sparse solution, we assume that only a few regions are active simultaneously. BIC1 is used to obtain the optimal number of regions (Kp) that explains the MEG/EEG data.

We compared the results of the proposed method to the ones from Minimum Norm Estimate (MNE) and LASSO. The first gives a smooth solution and the second gives a sparse solution. To test the accuracy of the reconstruction, we activated simultaneously from two to five regions in both hemispheres with synthetic low SNR signals (10 dB). Our approach could detect the right number of activated regions and provided more accurate reconstructions compared to MNE and LASSO.

Our approach assumes that few regions are active simultaneously which allows us to reduce the space to a few unkowns. It can be seen as an approximation to the l0 norm. Even though assuming a constant activation in each functional region is a hard constraint, it allows us to reduce the space size from S to K. The obtained solution can be used to detect extended sources (e.g epileptic activity) or as an initialization step to other approaches to obtain more detailed solutions in the active regions.

This work was presented at the conference BaCI 2015 [24].

6.7.3. Realistic simulation of electric potential distributions of different stimulation modes in an implanted cochlea

Participants: Kai Dang, Maureen Clerc, Clair Vandersteen [Institut Universitaire de la Face et du Cou, Nice], Nicolas Guevara [Institut Universitaire de la Face et du Cou, Nice], Dan Gnansia [Oticon Medical/Neurelec].

Simulation of the intracochlear potentials is an important approach to study the activation of auditory nerve fibers under electrical stimulations. However, it is still unclear to which extent the simulation results are affected by precision in reproducing the exact cochlear geometry. In this study, we address to this question by comparing the actual electric potential measured from implanted human specimen with the simulation outputs from two different parametric 3D cochlear models. One of the models is created from the default values while the other is adapted to the micro-CT scan data of the implanted cochlea.

This work was presented at the Association of Research in Otolaryngology 38th MidWinter Meeting, Feb 2015, Baltimore, United States [38].

We also made an in situ validation of electrical models: Cochlear implants have been proved to be an effective treatment for patients with sensorineural hearing loss. Among all the approaches that have been developed to design better cochlear implants, 3D model-based simulation stands out due to its detailed description of the electric field which helps reveal the electrophysiological phenomena inside the cochlea. With the advances in the cochlear implant manufacturing technology, the requirement on simulation accuracy increases. Improving the simulation accuracy relies on two aspects: 1) a better geometrical description of the cochlea that is able to distinguish the subtle differences across patients; 2) a comprehensive and reliable validation of the created 3D model. In this paper, targeting at high precision simulation, we propose a parametric cochlea model which uses micro-CT images to adapt to different cochlea geometries, then demonstrate its validation process with multi-channel stimulation data measured from a implanted cochlea. Comparisons between the simulation and validation data show a good match under a variety of stimulation configurations. The results suggest that the electric field distribution is affected by the geometric characteristics of each individual cochlea. These differences can be correctly reflected by simulations based on a 3D model tuned with personalized data.

This work was presented at the 7th International IEEE EMBS Conference on Neural Engineering, Apr 2015, Montpellier, France [25].

6.7.4. *Influence of skull modelling on conductivity estimation for EEG source analysis*

Participants: Christos Papageorgakis, Maureen Clerc, Benjamin Lanfer [BESA GmbH].

The skull conductivity strongly influences the accuracy of EEG source localization methods. As the conductivity of the skull has strong inter-individual variability, conductivity estimation techniques are required. Typically, conductivity estimation is performed on data from a single event-related stimulation paradigm, which can be explained by one dipole source. A conductivity value for the skull can be estimated as the value for which the single dipole source provides the best goodness of fit to the data. This conductivity value is then used to analyse the actual data of interest. It is known that the optimal local skull conductivity when modelling the skull as one compartment depends on the amount of spongiosa present locally. The research question arising is: Is conductivity estimation based on data from a single paradigm meaningful without accounting for the internal skull structure ?

This work was presented at the conference BaCI 2015 [33], and is submitted for journal publication.

6.7.5. *Dictionary learning for M/EEG multidimensional data*

Participants: Christos Papageorgakis, Sebastian Hitziger, Théodore Papadopoulo.

Signals obtained from magneto- or electroencephalography (M/EEG) are very noisy and inherently multidimensional, i.e. provide a vector of measurements at each single time instant. To cope with noise, researchers traditionally acquire measurements over multiple repetitions (trials) and average them to classify various patterns of activity. This is not optimal because of trial-to-trial variability (waveform variation, jitters). The jitter-adaptive dictionary learning method (JADL) has been developed to better handle for this variability (with a particular emphasis on jitters). JADL is a data-driven method that learns a dictionary (prototype pieces) from a set of signals, but is currently limited to a single channel, which restricts its capacity to work with very noisy data such as M/EEG. We propose an extension to the jitter-adaptive dictionary learning method, that is able to handle multidimensional measurements such as M/EEG.

This work was presented at the conference BaCI 2015 [32].

6.8. Brain Computer Interfaces

6.8.1. *Decoding covert shifts of attention induced by ambiguous visuospatial cues*

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

Simple and unambiguous visual cues (e.g., an arrow) can be used to trigger covert shifts of visual attention away from the center of gaze. The processing of visual stimuli is enhanced at the attended location. Covert shifts of attention modulate the power of cerebral oscillations in the alpha band over parietal and occipital regions. These modulations are sufficiently robust to be decoded on a single trial basis from electroencephalography (EEG) signals. It is often assumed that covert attention shifts are under voluntary control, and that they also occur in more natural and complex environments, but there is no direct evidence to support this assumption. We address this important issue by using random-dot stimuli to cue one of two opposite locations, where a visual target is presented. We contrast two conditions, one in which the random-dot motion is predictive of the target location, and the other, in which it provides ambiguous information. Behavioral results show attention shifts in anticipation of the visual target, in both conditions. In addition, using the common spatial patterns (CSPs) algorithm, we extract EEG power features in the alpha-band (around 10 Hz) that best discriminate the attended location in single trials. We obtain a significant decoding accuracy in 7/10 subjects using a cross-validation procedure applied in the predictive condition. Interestingly, similar accuracy (significant in 5/10 subjects) is obtained when the CSPs trained in the predictive condition are tested in the ambiguous condition. In agreement with this result, we find that the CSPs show very similar topographies in both conditions. These results shed a new light on the behavioral and EEG correlates of visuospatial attention in complex visual environments. This study demonstrates that alpha-power features could be used in brain computer interfaces to decode covert attention shifts in an environment containing ambiguous spatial information.

This work was published in Frontiers in Human Neurosciences [20].

6.8.2. Online extraction and single trial analysis of regions contributing to erroneous feedback detection

Participants: Eoin Thomas, Matthew Dyson, Laurence Casini, Boris Burle.

Understanding how the brain processes errors is an essential and active field of neuroscience. Real time extraction and analysis of error signals provide an innovative method of assessing how individuals perceive ongoing interactions without recourse to overt behaviour. This area of research is critical in modern Brain–Computer Interface (BCI) design, but may also open fruitful perspectives in cognitive neuroscience research. In this context, we sought to determine whether we can extract discriminatory error-related activity in the source space, online, and on a trial by trial basis from electroencephalography data recorded during motor imagery. Using a data driven approach, based on interpretable inverse solution algorithms, we assessed the extent to which automatically extracted error-related activity was physiologically and functionally interpretable according to performance monitoring literature. The applicability of inverse solution based methods for automatically extracting error signals, in the presence of noise generated by motor imagery, was validated by simulation. Representative regions of interest, outlining the primary generators contributing to classification, were found to correspond closely to networks involved in error detection and performance monitoring. We observed discriminative activity in non-frontal areas, demonstrating that areas outside of the medial frontal cortex can contribute to the classification of error feedback activity.

This work was published in NeuroImage [13].

DEMAR Project-Team

5. New Results

5.1. Modelling and identification of the sensory-motor system

5.1.1. Implementation and Validation of a Stride Length Estimation Algorithm, Using a Single Basic Inertial Sensor on Healthy Subjects and Patients Suffering from Parkinson's Disease

Participants: Christine Azevedo Coste, Benoît Sijobert, Mourad Benoussaad [ENIT, Tarbes, France], Christian Geny [CHU Montpellier, Neurology, France], Jennifer Denys [stagiaire M2 STIC SANTE - DEMAR].

Providing a clinical oriented solution, our study presented a gyrometer and accelerometer based algorithm for stride length estimation. Compared to most of the numerous existing works where only an averaged stride length is computed from several IMU, or where the use of the magnetometer is incompatible with everyday use, our challenge here has been to extract each individual stride length in an easy-to-use algorithm requiring only one inertial sensor attached to the subject shank. Our results were validated on healthy subjects and patients suffering from Parkinson's disease (PD). Estimated stride lengths were compared to GAITRite walkway system data: the mean error over all the strides was less than 6 percents for healthy group and 10.3 percents for PD group. This method provides a reliable portable solution for monitoring the instantaneous stride length and opens the way to promising applications ([27]).

5.1.2. Dynamic mapping of upper limb tremor by muscle ultrasonography

Participants: Olivier Tassaert [stagiaire M1 - DEMAR / ICAR], Benjamin Gilles, Olivier Strauss [LIRMM], Christian Geny [CHU Montpellier, Neurology, France], Christine Azevedo Coste.

Focal treatment of action tremor by botulinum toxin injections has been inadequately investigated and at best provides modest relief with significant muscle weakness. Complexity of multi-joint tremulous movements results in non-individualized dosing regimens. Tremor is complex, especially in the upper extremity, and its manifestation can change depending on posture, task, and bodypart. Proper characterization of the tremor based on visual inspection alone is a daunting task for the clinician. Identification of the main trembling muscles task disturbing is challenging because many upper limb muscles are bi-functional. The performance of electromyographic(EMG) pattern-recognition based method in classifying movements strongly depends on arm positions and needs multiple measurements. High density-surface EMG (HD-sEMG) is a non-invasive promising technique to measure electrical muscle activity but has not been used for tremor research because deep muscles could not be investigated Quantification of tremor dynamics by kinematics may be a feasible assessment and guidance tool which can be used to optimize injection conditions for focal tremor therapy. This approach is limited by the redundancy of the upper limb muscle organization Contribution of synergistic muscles toward specific movements over multi joint systems may change with varying position of distal or proximal joints. The choice of injected muscles remains highly subjective and variable. In the study of Rahami, ten different arm or forearm muscles have been injected and improvement was mild and delayed and associated with muscle weakness. In recent years, muscle ultrasonography has become a promising tool for diagnosing neuromuscular disorders . This technique is a non-invasive, low-cost, imaging modality that may be used to characterize normal and pathological muscle tissue but also subtle muscular activity (fasciculatios) in amyotrophic lateral sclerosis. The frequency of tremor remains stable during movement (3 to 8 Hz). We have initiated the investigation of the use of standard ultrasound as a technique to identify muscle groups responsible of upper limb tremor in patient with essential tremor or Parkinson's disease. The feasibility of the overall procedure has been validated: the acquisition procedure on patients, the possibility to track and segment the apparent motion in images using optical flow, and the ability to segment muscle groups by registering a 3D anatomical template.

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

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

Direct electrical stimulation (DES) have been recently introduced in the neurosurgery of slow-growing and infiltrative brain tumors to guide the resection. By generating transient perturbations, this method allows the real-time identification of both cortical areas and subcortical networks that are essential for the function. Thus, as much as possible, non-functional tissue can be removed while minimizing the sequelae. However, there is much controversy as to whether the use of DES during wideawake surgery is the gold standard for studying the brain function. It is sometimes wrongly assumed that electrical microstimulation (EMS) and DES induce similar effects in the nervous tissues and have comparable behavioural consequences. Both of them are used to perform functional brain mapping: EMS for animal fundamental neuroscience experiments, and DES for neurosurgery patients. We tried to shed new light on electrical stimulation (ES) techniques in brain mapping by comparing EMS and DES [1]. In fact, their effects cannot be directly compared - especially in the electrophysiological domain. There is a gap between theory and practice for ES of the brain. We do not know exactly how ES and especially DES influence the electrophysiological state of networks in the brain; a strong biophysical rationale is lacking. In contrast, the gap between EMS and DES highlights the potential for new experimental paradigms in electrical stimulation for functional brain mapping. In view of this gap and recent technical developments in stimulator design, it may now be time to move towards alternative, innovative protocols. Moreover, the understanding of the electrophysiological effects of DES remains an open and key question. Intra-operative EEG (iEEG) recordings were studied to analyze if and how stimulation currents spread at distant sites. Data were collected during an awake brain surgery for one patient. We observed significant changes in the frequency content at different iEEG sites during DES [2]. Subcortical DES led to neuromodulation at more sites than cortical DES (Figure 4). This may be due to (i) a better conduction and propagation following the direct stimulation of large, myelinated axons and (ii) the greater current intensity in subcortical DES. Further research will have to characterize these aspects more carefully and apply cortical and subcortical DES with identical current intensities [30], [31].

5.1.4. Functional Connectivity Analysis of Motor Imagery EEG signal for Brain-computer Interfacing Application: A preliminary study

Participants: Saugat Bhattacharyya, Poulami Ghosh [Jadavpur University, India], Ankita Mazumder [Jadavpur University, India], D.n. Tibarewala [Jadavpur University, India], Mitsuhiro Hayashibe.

The human brain can be considered as a graphical network having different regions with specific functionality and it can be said that a virtual functional connectivity are present between these regions. These regions are regarded as nodes and the functional links are regarded as the edges between them. The intensity of these functional links depend on the activation of the lobes while performing a specific task(e.g. motor imagery tasks, cognitive tasks and likewise). The analysis of these networks are performed by using a very useful mathematical tool called graph theory. Graph theory basically represents the entire functional network with a number of nodes and edges between them and the amount of connectivity existing between two nodes is depicted by assigning weights to the edges between them. In this study we have tried to utilize functional connectivity between different parts of the human brain for classifying a motor imagery task.

Brain connectivity patterns can be determined by using two types of measures, namely, Bivariate and Multivariate. Here we have considered a multivariate measure known as multivariate autoregressive (MVAR) model. One of the most widely investigated connectivity measure is the Directed Transfer Function (DTF). This function basically computes the directional influences between any two given nodes. There are a number of theoretical indices for defining a graph. In this preliminary work, two indices, namely node strength and network density are measured from the DTF values. In the current study, the BCI competition Dataset III is used for computing different multivariate measures.

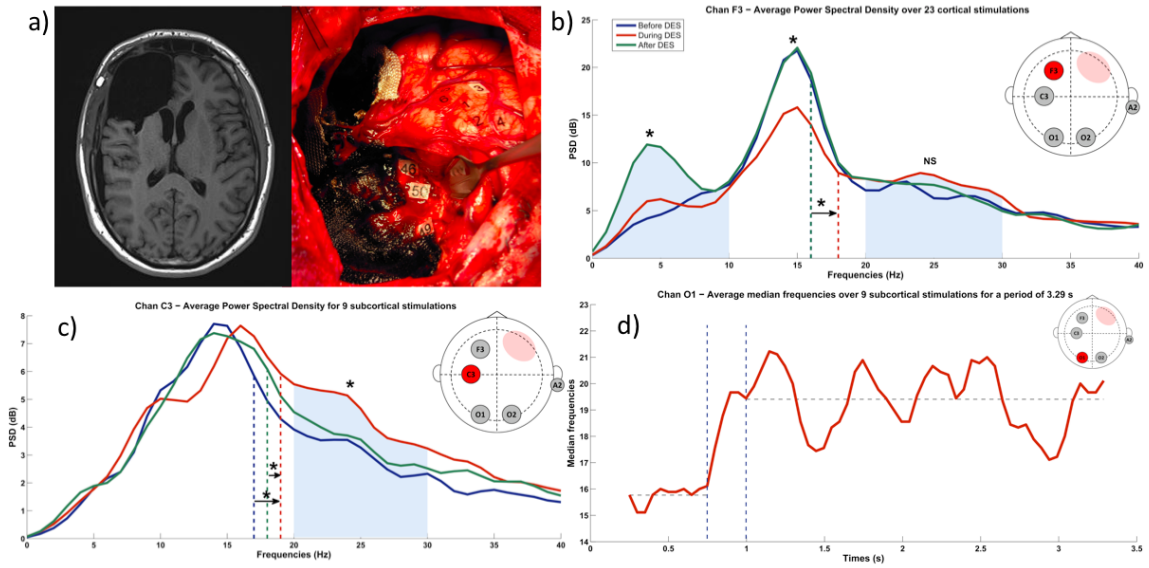


Figure 4. (a) Post-operative MRI of the patientâs brain, showing the right frontal cavity and an intraoperative view of the brain with the main anatomical landmarks. (b) The mean PSD of the iEEG signal, on F3, before, during and after each period of cortical DES. (c) The mean PSD of the iEEG signal, on C3, before during and after each period of subcortical DES. (d) The moving window median frequency averaged over nine subcortical DESs periods for PSD measured at O1.

The inflow-outflow graph of subject 1 while imagining right hand movement in the first training set are given in Fig.5 . Fig. 5 (a) describes the amount of inflow of functional connectivity going out of all the 32 electrodes and these are color coded to indicate the intensity of these inflows. From Fig. 5 (a) it is quite evident that the inflows are maximum in the frontal, temporal and occipital lobes. Figure 5 (b) depicts the functional outflow from the nodes and in contrast to Fig.5 (a) it shows that the outflows are maximum from the Central lobe(Cz). In Fig 5 (c), the direction of the flow between different nodes are shown and it can be seen clearly that majority of the paths are going from Cz to different nodes of the frontal, parietal and temporal lobes.

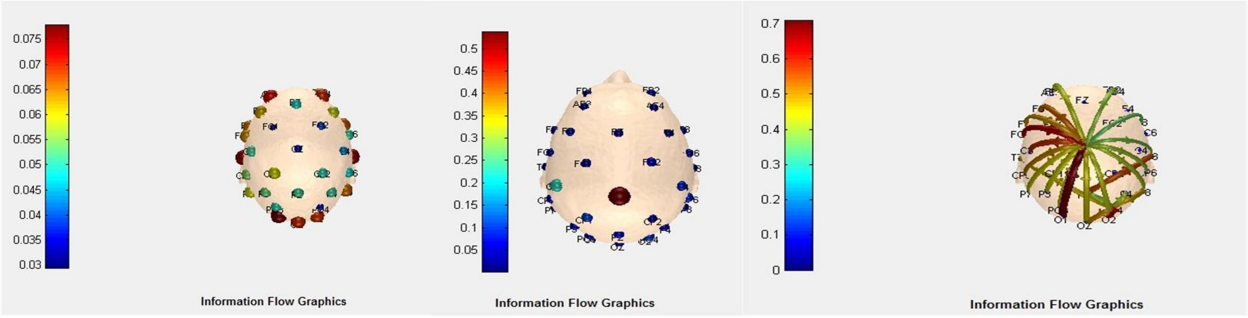


Figure 5. (a) Inflow graph, (b) Outflow graph and (c) Out to inflow graph of the functional connectivity network of the brain while imagining right hand movement.

5.1.5. A Generic Transferable EEG Decoder for Online Detection of Error Potential in Target Selection

Participants: Saugat Bhattacharyya, Amit Konar [Jadavpur University, India], D.n. Tibarewala [Jadavpur University, India], Mitsuhiro Hayashibe.

Detection of error from electroencephalography (EEG) signals as feedback while performing a discrete target selection task is beneficial for general Brain-computer Interfacing (BCI) systems including rehabilitative application. Error Related Potentials (ErrP) are EEG signals which occur when the participant observes an erroneous feedback from the system.

In this study, we have designed a novel scheme for detection of error feedback directly from the EEG signal. For this purpose, we have used a P300-speller dataset from the 'BCI Challenge @ NER 2015' competition hosted at Kaggle. The task involves the subject to select a letter of a word which is followed by a feedback period. The feedback period displays the letter selected and if the selection is wrong, the subject perceives it by the generation of ErrP signal. Our proposed system is designed to detect whether the feedback is erroneous or not. The decoder designed for this task is an ensemble of linear discriminant analysis, quadratic discriminant analysis and logistic regression classifier. The decoder is also transferable in nature as it is should work with single-trial on new subject without any prior subject-specific training.

The block diagram of the BCI system adopted for online ErrP detection from input EEG signals is shown in Fig.6 . The system implements three main processes: i) Pre-processing of the signal, i.e., temporal filtering in the bandwidth [0.1, 10]Hz, ii) Extraction of relevant features corresponding to the mental state from the signal using savitzsky-golay filter and meta-data of the features, and iii) Classification of the features, using our proposed decoder, to detect the intention of the participant from two given states: *Error* and *No-Error*. A switch is incorporated in the design to detect the beginning of feedback period in the trials, which is marked

in the datasets. We have tested the online functionality of the BCI system on the test dataset provided in the website. To simulate a real-time condition, the EEG is continuously streamed until an onset of the feedback period is detected. On detection of the feedback period, the system extracts a pre-defined length of signal for further processing and the rest are rejected.

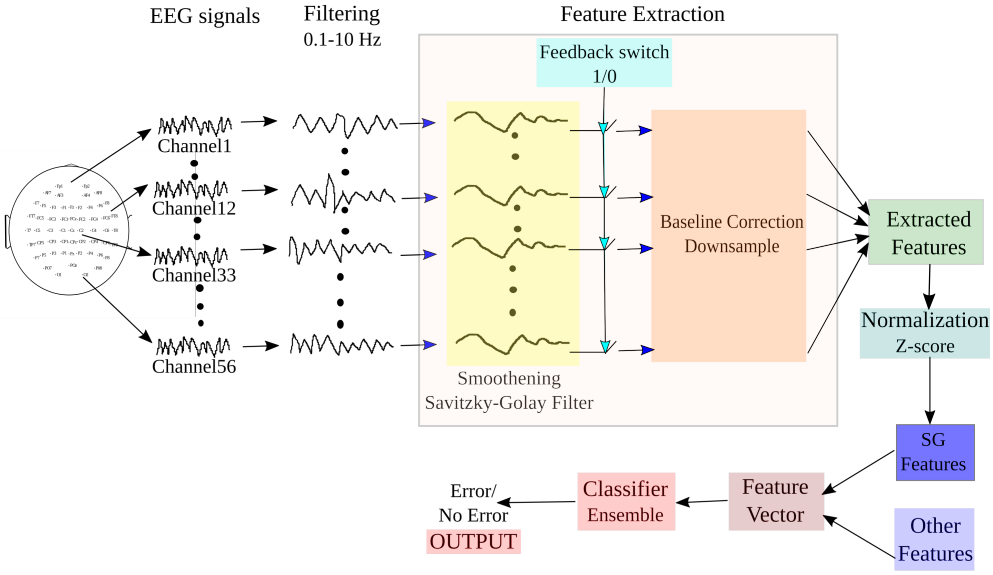


Figure 6. Block diagram of the BCI system adopted for online detection of Error Related Potentials from the input EEG

5.2. Synthesis and Control of Human Functions

5.2.1. FES-cycling and participation to Cybathlon competition

Participants: Christine Azevedo Coste, Benoît Sijobert, Charles Fattal [CRF DIVIO, Dijon, France], Antonio Padilha [UNB, Brasilia, Brazil], Emerson Fachin Martins [UNB, Brasilia, Brazil], David Andreu.

DEMAR and University of Brasilia will jointly participate with two SCI pilots to Cybathlon - FES-Bike competition. Cybathlon intends to promote assistive technologies during a competition. Two trikes will be adapted, an original control strategy will be proposed and two paraplegic individuals (one from Brazil and one from France) will be trained during the upcoming year. The protocol will be submitted to CPP ethical committee for agreement in the coming weeks (http://freewheels.inria.fr/).

5.2.2. PersoStim: A Personalized Closed-loop FES Control of Muscle Activation with Evoked EMG Feedback

Participants: Mitsuhiro Hayashibe, Zhan Li [University of Electronic Science and Technology of China], David Andreu, David Guiraud.

Functional electrical stimulation (FES) is a useful technique for restoring motor functions for spinal cord injured (SCI) patients. Muscle contractions can be artificially driven through delivery of electrical pulses to impaired muscles, and the electrical activity of contracted muscles under stimulus recorded by electromyography (EMG) is called M-wave. The FES-induced muscle activation which is represented by evoked EMG recordings can indicate the muscle state. Accurate control of muscle activation level by FES is the first step toward achieving more complicated FES control tasks.

A new FES closed-loop control strategy, EMG-feedback predictive control (EFPC), was developped to adaptively control stimulation pattern compensating to time-varying muscle state changes such as muscle fatigue and stimulation electrode detachment, along with the consideration of the personalized muscle responses to the electrical stimulation. This software manages a real-time FES system for control of muscle activation by online modulating pulse width of stimulus. The excitation muscle dynamics is modelled by Hammerstain system with stimulus pulse width and eEMG as input and output respectively. The model predictive control strategy is adopted to systematically produce the pulse width command of the stimulator. It is implemented together with Vivaltis portable stimulator. Four reference muscle activation patterns are provided to test and validate the real-time closed-loop FES control system. Real-time control results show promising control performances.

Recently, this software was demonstrated at the event of Rencontre Inria-Industrie 13/10/2015 at Bordeaux. https://www.inria.fr/centre/bordeaux/innovation/rii-sante/demonstrations2

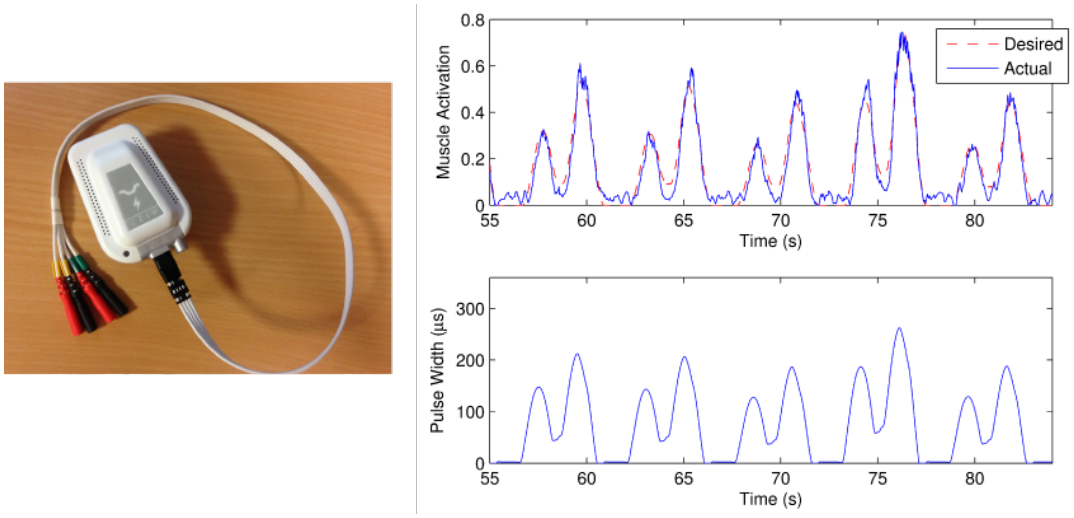


Figure 7. Left: Vivaltis portable stimulator, Right: Real-time control performance of muscle activation with desired dual sinusoidal shaped muscle activation pattern (red dash line is desired muscle activation trajectory and blue solid line is the measured muscle activation under the muscle activation control by FES). The lower plot is the corresponding computed stimulation pulse width.

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

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

Complete spinal cord injury results in loss of movement and sensory sensations but also in function of organs. For example, nearly all spinal cord injured subjects lose their bladder control and are prone to kidney failure if they do not apply intermittent (self-) catheterization. Electrical stimulation of the sacral spinal roots with an implantable neuroprosthesis is one option besides self-catheterization to become continent and control micturition. However, many persons do not ask for this neuroprosthesis since deafferentation and loss of sensory functions and reflexes are serious side effects. Spinal cord stimulation (SCS) is a general term which includes both epidural and intradural stimulation. Originally associated with the treatment of chronic

neurogical pain (in the 1970ies), SCS led also to immediate and profound improvements of sensory and motor functions in recent studies both on SCI patients (only on very few case studies) and rodents. Despite these promising results some limitations have still to be overcome. Among them, the use of small animal models, the empirical aspect of the stimulation procedure and the impact of these protocols on intestinal and urinary functions are critical. To counteract these limits, we want to explore intradural and epidural stimulations in an intermediate model- the house pig- and assess their impact on bladder, guts and genitals. In order to evaluate our approach, we will record EMG signals of lower limbs and sphincters (both urethral and anal), and simultaneously, we will monitor bladder and rectal pressure.

Already preliminary experimental explorations were performed with direct spinal cord stimulation in June (on 2 animals). Experiments were conducted under neurosurgeons involved in the project and urodynamics was recorded together with rectum pressure and sphincters EMG during each stimulation session.

5.3. Neuroprostheses and technology

5.3.1. Selectivity of nerve stimulation using a 12 pole multipolar cuff

Participants: Wafa Tigra, Olivier Rossel, Thomas Guiho, David Guiraud, Christine Azevedo Coste, Hubert Taillades [UM].

Experimentations were performed on 5 rabbits (New Zealand white). A multipolar cuff electrode (12 poles, diameter 3 mm, length 20 mm, 12 oblong contacts of 5mm length) was placed around the sciatic nerve of the rabbit 3 cm above the tibiofibular bifurcation. The nerve was stimulated with increasing intensity. The protocol consisted of the activation of one or more channels of the electrode, the input is a biphasic asymmetric stimulation and the pulse width is modulated in intensity (up to 2.4 mA) and fixed in length (250 μsec, 100 μsec interstim). A stimulus (4 Hz) is used for 2 seconds. 48 configurations of stimulation were tested. Needle electrodes were inserted on the lateral and medial gastrocnemius, soleus, tibialis and extensor digitorum muscles to record EMG signals and were used to evaluate the selectivity capacities of given cuff electrode configuration. The rabbit foot was also attached to a force platform. Inter-fascicular selectivity was observed for the 5 animals. Intra fascicular selectivity was also observed in 3 animals. Placed at a single location, our cuff electrode is capable to activate, selectively, some muscles. Experiments were performed under ethical comittee agreement at the "Plateau Technique Chirurgie Exprimentale" (Montpellier).

5.3.2. A novel EMG interface for individuals with quadriplegia to pilot robot hand grasping

Participants: Wafa Tigra, Benjamin Navarro [LIRMM], Andrea Cherubini [LIRMM], Xavier Gorron [???], Anthony Gelis [PROPARA], Charles Fattal [CRF DIVIO, Dijon, France], David Guiraud, Christine Azevedo Coste.

We have developed and validated a new human-machine interface dedicated to individuals with quadriplegia. We investigated the feasibility of online processing sus-lesional muscle responses, to pilot an assistive device. The ability to voluntary contract a set of selected muscles was assessed in five spinal cord injured subjects through electromyography analysis. Two subjects have also been asked to use the EMG interface to control palmar and lateral grasping of a robot hand (fig.8). These preliminary results sound very promising and open the way to new interface solutions for high level spinal cord injured patients(fig.8).

5.3.3. Wearable 56-pole stimulator

Participants: Arthur Hiairrassary, David Andreu, David Guiraud, Olivier Rossel, Thomas Guiho.

In the context of the EPIONE European project, we have designed and developed, with Axonic, a wearable multichannel stimulator (fig.9), to face phantom limb pain (PLP). This 56-pole neural stimulator is based on four 16-pole stimulation units (each one being connected to a 16-pole TLIFE intra-fascicular electrode) connected to a real-time controller by means of an embedded deterministic network. This controller, in charge of executing FES functions (threshold determination, sensation characterization, etc.), pilots the stimulation units and allows for real-time modulation of the multisite stimulation. The controller can be remotely configured and exploited by the practitioner, by means of a dedicated software (Synergy Neuromodulation Software). But it has been also connected to the controller of the EPFL's artificial hand, in order to link hand touch sensors with neural stimulation to induce natural, meaningful sensations to the amputee.

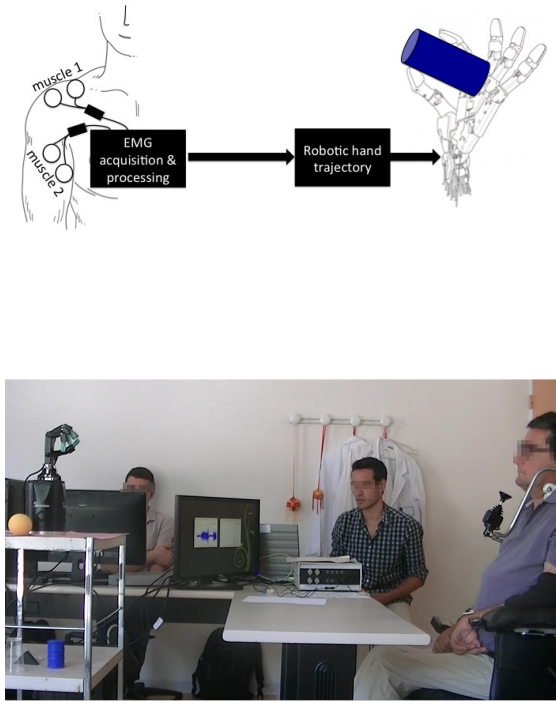


Figure 8. Principle of robot hand control through EMG signals; Setup description and upper arm positioning during EMG recordings.

This stimulator has been deeply validated through animal experiments (rats and pigs, respectively with UAB Barcelone and SMI Alboorg) and is currently used on human at UCBM Rome (http://project-epione.eu/).

Figure 9. 56-pole neural stimulator

5.3.4. CORAIL: Neural Stimulation Integrated Circuit

Participants: Jérémie Salles, Guy Cathébras, Milan Demarcq, David Guiraud, Guillaume Souquet, David Andreu.

DEMAR is currently finishing the development of CORAIL (Current Output Reconfigurable Asic Interface Low power), a new ASIC dedicated to electric neural stimulation.

Its main analog characteristics are:

- 12 independent current output channels;
- a full-scale current of 5 mA with a quantum of 1.3 μA;
- a symmetrical power supply ($\pm$ VHT with ground middlepoint).

This front-end integrated circuit is designed to perform multipolar electrical stimulation of the nerve with highly configurable waveforms in order to achieve selective activation of organs or muscles. In comparison with previously developed current output ASICs, the CORAIL IC embeds new features such as the storage of multiple electrode configurations or the possibility to internally combine poles. These specific aspects of CORAIL and the fact that its elaboration benefited from clinical experience in the team will allow enhanced integration within the whole electrical stimulation environment.

The resulting stimulation ASIC aims to be part of an implanted distributed stimulation system, composed of multiple stimulation units spread across the body, in which CORAIL will be the front-end entity in charge of delivering the current to the electrode. Thus, special care has been paid to its integration in such a network with an emphasis on low power consumption for which different mechanisms have been implemented.

The ASIC is currently undergoing the last phases of its development and a first version is due for fabrication in February 2016.

5.3.5. Tele-Rehabilitation Platform for Gait Training

Participants: Mitsuhiro Hayashibe, Antonio P.l. Bo [Universidade de Brasilia, Brasil], Leslie Casas [Pontificia Universidad Catolica del Peru, Peru], Gonzalo Cucho [Pontificia Universidad Catolica del Peru, Peru], Dante Elias [Pontificia Universidad Catolica del Peru, Peru].

Throughout the world there is an increasing need for better technologies for rehabilitation and assistance. These new solutions must present improved performance in terms of therapy effectiveness, while at the same time minimizing the corresponding costs. In this scenario, computer-aided methods represent a promising alternative for the challenges currently faced by the rehabilitation domain. A tele-rehabilitation platform for gait training in intercontinental circumstances is developed under STIC-AmSud program. This project was joint program 2012-2013 among Inria France, UnB (University of Brasilia) and PCUP (Pontifical Catholic University of Peru) for tele-rehabilitation framework. This system has two mode: Self-modulation control in which the subject can control the speed of the motion therapy with his comfortable training speed and Guidance control mode in which the motion transfer is performed from one therapist to one patient. Guidance control can be performed both with local data transmission and intercontinental data transmission. Fig. 10 shows the case where the motion transfer regarding foot placement was performed with local data transmission. The test with intercontinental data transmission was also realized between France and Peru.

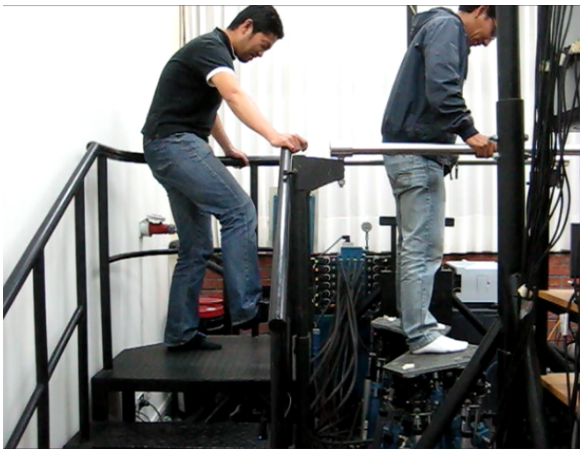


Figure 10. Tele-rehabilitation platform for gait training: Guidance control mode.

5.3.6. Control and scheduling co-design for stimulation systems

Participants: Daniel Simon, David Andreu.

Functional Electrical Stimulation (FES) is used in therapy for rehabilitation or substitution for disabled people. They are control systems using electrodes to interface a digital control system with livings. Hence the whole system gathers continuous-time (muscles and nerves) and discrete-time (controllers and communication links) components. During the design process, realistic simulation remains a precious tool ahead of real experiments to check without danger that the implementation matches the functional and safety requirements. To this aim we are developing a real-time open software simulation system, dedicated to the analysis of FES systems deployed over distributed execution resources and wireless links. The simulation tool is especially devoted to the joint design and analysis of control loops and real-time features.

Realistic simulations are effective tools to design and tune complex systems whose analysis cannot be provided only by theory. Several simulation steps can be explored, from simple functional analysis to HIL, to design, test, tune and validate both the single components of the system and their interactions in a distributed architecture. Simulations are precious, as they allow for non-destructive trials, which must be considered in any domain but this is of particular interest for bio-engineering [42].

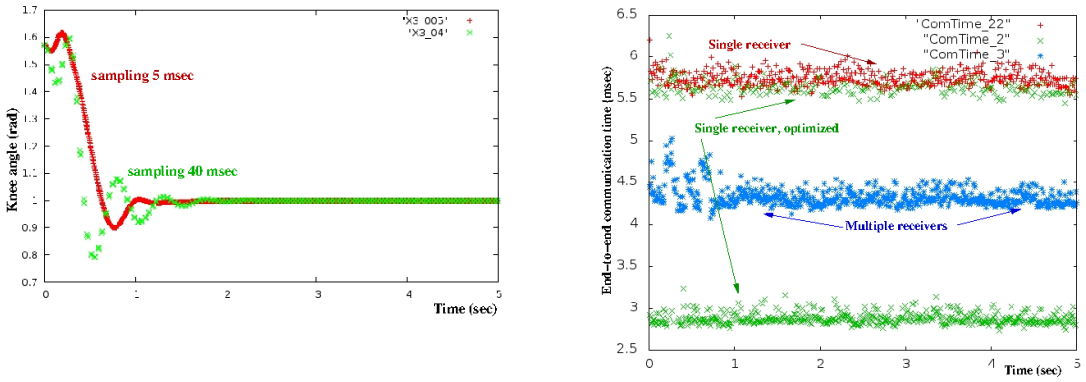


Figure 11. Simulation traces : knee position w.r.t. sampling and communication delays

It is expected that this particular simulator may provide inputs in two main directions. Firstly it allows for preliminary testing and tuning new FES protocols without needing for real experiments with patients, and may help for writing the ethical protocols needed for any experiments involving livings. Secondly it can be used to preliminary evaluation of new technologies or implementations, without costly reworking of existing electronic chips or certified components.

The simulation software is open, so that enhancements w.r.t. to the original release can be added upon request of various designers and to fulfill various objectives.

5.3.7. Control loops design principles for autonomic computing

Participants: Daniel Simon, Eric Rutten [Inria Grenoble Rhône-Alpes], Nicolas Marchand [GIPSA-lab].

Computing systems are becoming more and more dynamically reconfigurable or adaptive, to be flexible w.r.t. their environment and to automate their administration. Autonomic computing proposes a general structure of feedback loop to take this into account. We are particularly interested in approaches where this feedback loop is considered as a case of control loop where techniques stemming from Control Theory can be used to design efficient safe, and predictable controllers. This approach is emerging, with separate and dispersed effort, in different areas of the field of reconfigurable or adaptive computing, at software or architecture level.

We aim at conveying to Computer Scientists the interest and advantages of adopting a Control Theory perspective for the efficient and predictable design of autonomic systems. Compared with open-loop, closed-loop control provides adaptability and robustness, allowing for the design of fault-tolerant systems against varying and uncertain operating conditions. However, there still is a deep need for research in the problems of mapping from high-level objectives in terms of Quality of Service (QoS) or Service Level Objectives (SLO) and abstract models towards lower-levels effective actions on the managed systems [46].

5.4. Others

5.4.1. Do doors opening affect the air contamination in clean surgery? A Prospective, Cross-sectional Study (the ARIBO Project)

Participants: Gabriel Birgand [APHP], Christine Azevedo Coste, Stephane Rukly [INSERM], Roger Pissard-Gibollet [Inria Grenoble Rhône-Alpes], Jean-Christophe Lucet [APHP].

Inappropriate staff behaviours can lead to environmental contamination in the operating room (OR) and subsequent surgical site infection (SSI). This study focused on the continued assessment of OR staff behaviours using doors sensors and their impact on the SSI risk during surgical procedures. This multicentre observational study included 13 ORs in 10 hospitals, 5 University hospitals and 5 private hospitals. Two specialties of clean surgery with cutaneous approach were included: cardiac surgery with procedures requiring a full median sternotomy (CABG or valve replacement surgery); and planned orthopaedic surgery for total hip (THR) or knee replacement (TKR). For each surgical specialty involved, the observed ORs were randomly selected. Doors opening were observed by means of wireless inertial sensors fixed on the doors. For each surgical procedure, 3 microbiological air counts, continuous particles counts of 0.3, 0.5 and 5μm particles, and one bacteriological sample of the wound before skin closure were performed. We collected informations on the OR staff, surgical procedures and surgical environment characteristics. Statistics were performed using univariate and multivariate analysis to adjust on aerolic and architectural characteristics of the OR. We included 34 orthopaedic and 26 cardiac procedures. The mean duration of intervention, from patient entry to exit in the OR, was 5.3 (SD 1.1) h. in cardiac and 2.6 (0.7) h. in orthopaedic surgery. The median number of doors opening was 146 (IQR: 121-183; Min-Max: 86-319) per intervention and 29 (IQR: 23-36; Min-Max: 17, 54) per h. in cardiac surgery and 71.5 (IQR: 58-92; Min-Max: 54-136) per intervention and 29 (IQR: 25-34; Min-Max: 16-65) per h. in orthopaedic procedures. Doors stayed open in average 43 minutes (Min-Max: 19-115) in cardiac and 36 (8-199) in orthopaedic, representing respectively 13.5 percents and 23 percents of the duration of intervention. The highest frequency of doors opening was observed between wound closure and patient exit, median 20.1 openings/h (12.5-32.3) and from patient entry to the incision 13.2 openings/h (8-19). The number and duration of doors opening was significantly different between centres (higher in university hospital, $p<0.01$)). High frequency of openings was observed for doors that should stay closed during procedures (materials store, decontamination room). The number of doors opening from skin incision to wound closure affected significantly the 0.5 and 5μm particles count ($p<0.01$ and 0.02 respectively). This study based on automatic observation suggests a large heterogeneity of doors openings between types of interventions, ORs and hospitals. Data give a standard of doors opening for CABG, THR and TKR. Door openings affected air contamination, potentially jeopardizing operating room sterility. The causes and influences of behaviours in the OR must be evaluated to identify ways to reduce the associated risks.

GALEN Project-Team

7. New Results

7.1. Optimizing Average Precision

Paticipants: Pawan Kumar

Average precision (AP) is one of the most commonly used measures for ranking. However, due to the inefficiency of optimizing it during learning, a common approach is to use surrogate loss functions such as 0-1 loss. We have developped a novel latent AP-SVM classifier [1], that minimizes a carefully designed upper bound on the AP-based loss function over weakly supervised samples. Using publicly available datasets, we demonstrate the advantage of our approach over standard loss-based learning frameworks on three challenging problems: action classification, character recognition and object detection.

7.2. Region-based Semantic Segmentation

Paticipants: Pawan Kumar

In [9] we consider the problem of parameter estimation and energy minimization for a region-based semantic segmentation model. The main problem we face in the context of energy minimization, is the large number of putative pixel-to-region assignments. We address this problem by designing an accurate linear programming based approach for selecting the best set of regions from a large dictionary, which is constructed by merging and intersecting segments obtained from multiple bottom-up over-segmentations. The lack of fully supervised data is tackled by using a latent structural SVM formulation, where the latent variables model any missing information in the human annotation. Using large, publicly available datasets we show that our methods are able to significantly improve the accuracy of the region-based model.

7.3. Parsimonious Labeling

Paticipants: Puneet Dokania, Pawan Kumar

In [22], we propose a new family of discrete energy minimization problems, which we call parsimonious labeling, that is to use as few labels as possible. This allows us to capture useful cues for important computer vision applications such as stereo correspondence and image denoising. Furthermore, we propose an efficient graph-cuts based algorithm for the parsimonious labeling problem that provides strong theoretical guarantees on the quality of the solution. Using both synthetic and standard real datasets, we show that our algorithm significantly outperforms other graph-cuts based approaches.

7.4. Structured Learning and Inference

Paticipants: Jiaqian Yu, Matthew Blaschko

We have developed computationally efficient structured output prediction methods for learning with non-modular losses [19], [29], [40]. We both demonstrate the feasibility of learning with submodular losses, as well as show that learning with multiple correct outputs can lead to NP-hard optimization problems even when learning with a single correct output is feasible.

7.5. Novel graph kernels

Paticipants: Katerina Gkirtzou, Matthew Blaschko

We have developed a novel family of graph kernels that are capable of incorporating local curvature properties of 3D meshes [6]. We have additionally demonstrated their application to the modelling of interdependencies between different brain regions in an fMRI based classification task for predicting cocaine addiction.

7.6. Structured Sparsity Regularization & Statistical Hypothesis Testing

Paticipants: Eugene Belilovsky, Wacha Bounliphone, Katerina Gkirtzou, Andreas Argyriou, Matthew Blaschko

We have developed novel methods for structured sparsity regularization & hypothesis testing. We have applied these methods to fMRI [3], [2], [36] and the analysis of large medical databases [10]. We have also developed novel statistical hypothesis tests for relative dependency [21], [37] and similarity [14]. We have applied these methods to the problem of identifying relative dependencies between languages using a multi-lingual corpus, and for discovering the relative relationships between gliomas and genetic information. Additionally, we have shown the application of relative tests to the problem of model selection in deep generative models, and currently an important question in machine learning.

7.7. High-Order MRF models

Paticipants: Nikos Komodakis, Nikos Paragios

We developped a very general algorithm for structured prediction learning [7] that is able to efficiently handle discrete MRFs/CRFs (including both pairwise and higher-order models) so long as they can admit a decomposition into tractable subproblems. By properly combining dual decomposition with a max-margin learning method, the framework manages to reduce the training of a complex high-order MRF to the parallel training of a series of simple slave MRFs that are much easier to handle.

7.8. Graph-based registration and segmentation

Paticipants: Enzo Ferrante, Vivien Fecamp, Aimilia Gastounioti, Bharat Singh, Stavros Alchatzidis, Nikos Paragios

Deformable image registration plays a fundamental role in many clinical applications. We investigated the use of graphical models in the context of a particular type of image registration problem, known as slice-to-volume registration. We introduced a scalable, modular and flexible formulation that can accommodate low-rank [5] and high order [16] terms, that simultaneously selects the plane and estimates the in-plane deformation through a single shot optimization approach. We applied our models on simulated and real-data in the context of ultrasound and magnetic resonance registration, demonstrating the potential of our methods.

We also developped a novel methodology for graph-based motion-driven segmentation [24] and applied it for carotid plaque segmentation in ultrasound images. We identified the plaque region by exploiting kinematic dependencies between the atherosclerotic and the normal arterial wall. The methodology exploits group-wise image registration towards recovering the deformation field, on which information theory criteria are used to determine dominant motion classes and a map reflecting kinematic dependencies, which is then segmented using Markov random fields.

Moreover, in order to address the problem of general purpose multi-modal deformable registration/fusion we developped a novel and robust method using a metric defined in an appropriate sub-space which is adaptive to the image-content/image-modality [18]. We adopted a graph-based formulation that assumes that intensities of corresponding pixels in the two image domains are related through an unknown piece-wise constant linear function. This relation is propagated to an appropriate sub-space (wavelets coefficients) where a criterion that couples the estimation of the deformation field with optimal transport function on the subspace and the smoothness of the deformation is considered.

7.9. Object Detection from RGB-Depth images

Paticipant: Siddhartha Chandra, Iasonas Kokkinos

In [11] we explore RGB-Depth representations for the training of Deformable Models. and describe strategies to improve an object detection pipeline by introducing viewpoint based mixture components. Our contributions are threefold. First, we use surface-based object representations (3D mesh models) from available 3D object model repositories to learn strongly supervised viewpoint classifiers. Second, we develop a geometric dataset augmentation scheme that uses scene geometry to 'take another look' at the training data, simulating the effect of camera viewpoint changes. Third, to better exploit depth information, we develop a novel depth-based dense feature extraction method that provides a robust statistical description of scene geometry. We evaluate our learned detectors on the common NYU dataset, and demonstrate that each of our advances results in systematic performance improvements over the traditional detection pipeline.

7.10. Deep CNN for Modelling Deformations and Semantic Segmentation

Paticipant: Iasonas Kokkinos

Invariance to deformations in Deep Convolutional Neural Networks (DCNN) is commonly achieved by using multiple 'max-pooling' (MP) layers. In [26] we show that alternative methods of modeling deformations can improve the accuracy and efficiency of DCNNs. For this, (i) we introduce epitomic convolution as an alternative to the common convolution-MP cascade of DCNNs, (ii)we introduce a Multiple Instance Learning algorithm to accommodate global translation and scaling in image classification and (iii) we develop a DCNN sliding window detector that explicitly, but efficiently, searches over the object's position, scale, and aspect ratio. We provide competitive image classification and localization results on the ImageNet dataset and object detection results on Pascal VOC2007.

In [25] we bring together methods from DCNNs and probabilistic graphical models for addressing the task of pixel-level classification ("semantic image segmentation"). We overcome the poor localization property of deep networks by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF). Qualitatively, our "DeepLab" system is able to localize segment boundaries at a level of accuracy which is beyond previous methods.

7.11. Learning Low-level Image Representations with Deep CNN

Paticipant: Iasonas Kokkinos

In [27] we propose a novel framework for learning local image descriptors in a discriminative manner. For this purpose we explore a siamese architecture of DCNNs, with a Hinge embedding loss on the L2 distance between descriptors. Since a siamese architecture uses pairs rather than single image patches to train, there exist a large number of positive samples and an exponential number of negative samples. We propose to explore this space with a stochastic sampling of the training set, in combination with an aggressive mining strategy over both the positive and negative samples which we denote as "fracking". We perform a thorough evaluation of the architecture hyper-parameters, and demonstrate large performance gains compared to both standard CNN learning strategies, hand-crafted image descriptors like SIFT, and the state-of-the-art on learned descriptors: up to 2.5x vs SIFT and 1.5x vs the state-of-the-art in terms of the area under the curve (AUC) of the Precision-Recall curve.

In [4] we explote connections between DCNNs and texture understanding. First, instead of focusing on texture instance and material category recognition, we propose a human-interpretable vocabulary of texture attributes to describe common texture patterns, complemented by a new describable texture dataset for benchmarking. Second, we look at the problem of recognizing materials and texture attributes in realistic imaging conditions, including when textures appear in clutter, developing corresponding benchmarks on top of the recently proposed OpenSurfaces dataset. Third, we revisit classic texture representations, including bag-of-visual-words and the Fisher vectors, in the context of deep learning and show that these have excellent efficiency and generalization properties if the convolutional layers of a deep model are used as filter banks. We obtain in this manner state-of-the-art performance in numerous datasets well beyond textures, an efficient method to apply deep features to image regions, as well as benefit in transferring features from one domain to another.

In [35] we propose a new DCNN architecture that learns pixel embeddings, such that pairwise distances between the embeddings can be used to infer whether or not the pixels lie on the same region. That is, for any two pixels on the same object, the embeddings are trained to be similar; for any pair that straddles an object boundary, the embeddings are trained to be dissimilar. Experimental results show that when this embedding network is used in conjunction with a DCNN trained on semantic segmentation, there is a systematic improvement in per-pixel classification accuracy. Our contributions are integrated in the popular Caffe deep learning framework, and consist in straightforward modifications to convolution routines. As such, they can be exploited for any task involving convolution layers.

7.12. Human-Limb Segmentation for Intelligent Mobility Assistance Robots

Paticipants: Siddhartha Chandra, Stavros Tsogkas, Iasonas Kokkinos

We developed a computer vision component [12] to be used as part of an intelligent robotic assistant. This component exploits RGB and depth information extracted from Kinect sensors mounted on the robot, to accurately segment human limbs, using fully convolutional neural networks (FCNNs). We trained our network using an in-house Human-Limb dataset composed of video frames, and described a scheme for dynamically exploiting RGB and depth data in a single framework for training and testing. Our method demonstrated promising performance, being very efficient at the same time, with a run-time of 8 frames per second on a single GPU.

MIMESIS Team

6. New Results

6.1. Augmented reality for surgery

We have developed a method for real-time augmented reality of internal liver structures during minimally invasive hepatic surgery. Vessels and tumors computed from pre-operative Computed Tomography Angiograms (CTA) scans can be overlaid onto the laparoscopic view for surgery guidance. Compared to current methods, our method is able to locate the in-depth positions of the tumors based on partial three-dimensional liver tissue motion using a real-time biomechanical model. We are pursuing the development of this augmented reality system by using a better biomechanical model, and by relying on parameter optimization and additional per-operative information to further improve accuracy and robustness. In addition, more experiments, and also clinical studies are being performed to precisely measure the benefits and limitations of our approach. This work is strongly related to our involvement in the IHU Strasbourg and is tightly linked to the SOFA-OR project. Many articles were published on this topic [28], [16], [17].

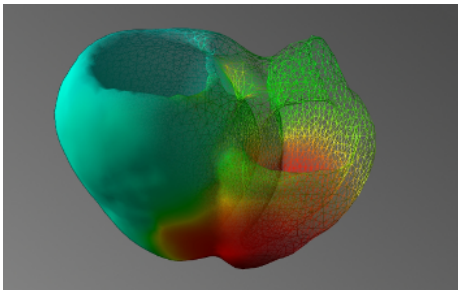


Figure 9. Electrophysiology model of the human heart

6.2. Cardiac electrophysiology

Cardiac arrhythmia is a very frequent pathology that comes from an abnormal electrical activity in the myocardium. This pathology can be treated by catheterization and ablation of the malfunctionning cardiac tissue. The skills required for such interventions are still very challenging to learn, and typically acquired over several years. We first developed a training simulator for interventional electrocardiology and thermo-ablation of these arrhythmias. Based on physical models 9 , this training system reproduces the different steps of the procedure, including endovascular navigation, electrophysiological mapping, pacing and cardiac ablation. Based on a scenario of cardiac arrhythmia, cardiologists assessed the interactivity and the realism of our simulation. This work has been submitted in a journal and is currently under review.

Beyond electrophysiology training, our work around the cardiac electrophysiology also consisted in personalizing our mathematical models. Using the dense electrograms recorded intra-operatively, we presented an accurate and innovative approach to personalize our model, i.e. estimate patient-specific parameters. The modeling in silico of a patient electrophysiology is needed to better understand the mechanism of cardiac arrhythmia.

6.3. Cryoablation

In 2015, we carried on the work around thermal ablation and pre-operative planning based on a thermal Finite Element Model (FEM). The cryoablation technique consists in inserting needles that freeze the surrounding tissues, thus immediately leading to cellular death of the tissues. Cryoablation procedure is used in many medical fields for tumor ablation, and even starts being used in cardiology. In this scope, we built a simulator 10 able to place the cryoprobes and run a simulation representing the evolution of iceballs in living tissues.

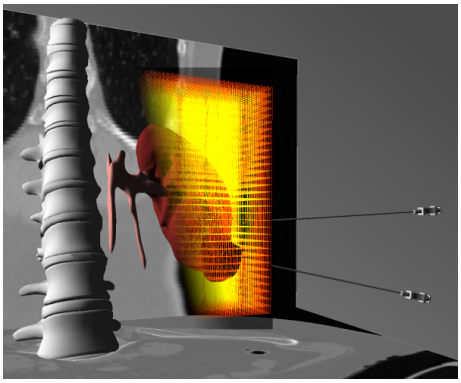


Figure 10. Cryosurgery simulation with the creation of an iceball in the kidney

6.4. Lipofilling reconstructive surgery

We have developed a method to simulate the outcome of reconstructive facial surgery based on fat-filling. Facial anatomy is complex: the fat is constrained between layers of tissues which behave as walls along the face; in addition, connective tissues that are present between these different layers also influence the fat-filling procedure. To simulate the end result, we have proposed a method which couples a 2.5D Eulerian fluid model for the fat and a finite element model for the soft tissues. Both models are coupled using the computation of the mechanical compliance matrix. We had two contributions: a solver for fluids which couples properties of solid tissues and fluid pressure, and an application of this solver to fat-filling surgery procedure simulation. Vincent Majorczyk defended his PhD [14] on this topic in 2015.

6.5. Neurosurgery

Based on an intra-operative registration method, we developed a simulation of a DBS (Deep Brain Stimulation) surgery which can help the surgeon to locate anatomical structures for a safer and a more efficient treatment. The method relies on the biomechanical model of brain shift we developed during the last years. Because some parameters of the model are unknown, we propose to estimate them with an optimization process. The cost function evaluates the distance between the model and the segmentation of pneumocephalus, the only indicator of brain shift visible on an intra-operative CT scan. In 2015, an article about the rest shape of the brain was accepted [19].

6.6. Physics-based registration algorithms

Before targeting the augmented reality for laparoscopic operations, an important step consists in solving the initial alignment problem. Given a pre-operative image of the organ (usually a CT scan) a detailed mesh is constructed. To make the information stored in this mesh available during the operation, the mesh must be registered onto the intra operative view. However, mainly due to the pneumoperitoneum, the organ has

undergone important deformation between the pre-operative images acquisition and the operation. The pre-operative shape and the intra-operative shape of the organ do not correspond. Therefore a non rigid registration is required to align the mesh and the real organ. Our registration algorithms also allowed us to work on means to automatically recover boundary conditions of a patient specific liver.

We created a statistical atlas of the human liver to store the positions of the boundary conditions: the veina cava and the anchor point of the falciform ligament positions. This method was accepted at ABME in 2015 [21]. We also developed a new registration method that evolves automatically from a rigid registration to a non rigid registration to solve the initial alignment problem. The method uses some anatomical features of the liver such as the anchor point position of the falciform ligament.

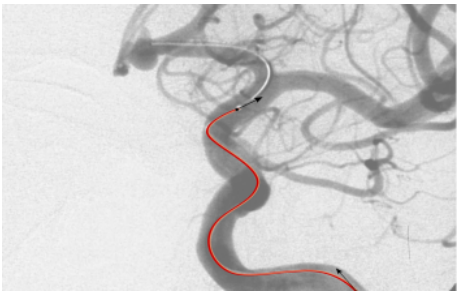


Figure 11. Detecting a catheter in interventional medical images

6.7. Radiation-less guidance during interventional radiology procedures

Significant changes have taken place over the past 20 years in medicine with the development of minimally-invasive procedures. While surgery evolved towards laparoscopy for instance, interventional radiology has become another alternative for many pathologies. Yet, some limitations remain: for percutaneous procedures, soft tissue motion, either due to breathing or deformation induced by the needle, changes the location of the target. When using image guidance, or robotic control, this remains a major obstacle. Regarding catheter-based interventions, the lack of 3D information, and extensive use of X-ray imaging to visualize the path to be followed, are among the main issues. We propose to address these different problems by developing an advanced navigation system which relies on a combination of real-time simulation and information extracted from intra-operative images to assess the current position of the needle. Such a method would have direct applications in pre-operative planning, per-operative guidance, and control for robotics. Our approach will combine advanced modeling of the device, soft tissue deformation, tissue-tool interactions, and planning algorithms 11 .

6.8. Regional anaesthesia

The RASimAs project (Regional Anaesthesia Simulator and Assistant) is a European research project funded by the European Union's 7th Framework Program. It aims at providing a virtual reality simulator and assistant to doctors performing regional anaesthesia by developing the patient-specific Virtual Physiological Human models. In this project, we are in charge of developing a simulation of a needle inserted into a leg using the SOFA framework 12 . We especially focused on the integration of the needle simulation into SOFA. We planned to release the first version of the simulator by January 2016.

In the context of RASimAs, we organized a coding sprint in Strasbourg in April 2015.

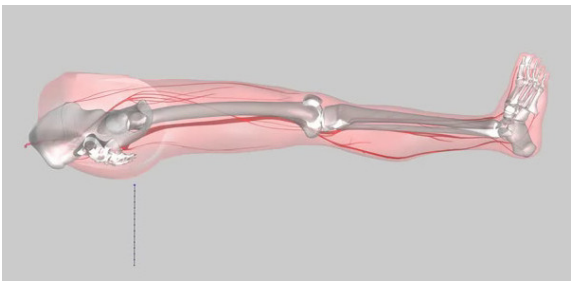


Figure 12. Needle insertion in a muscle in the context of local anaesthesia

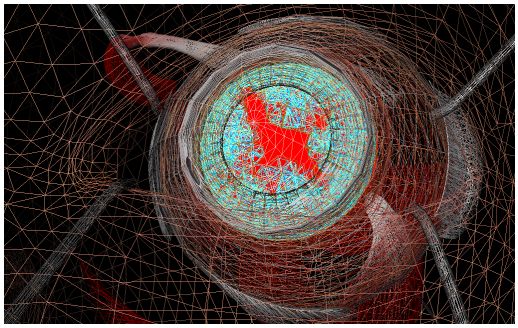


Figure 13. FEM model of the eye used in our simulation of retina surgery

6.9. Training for retina surgery

Retina surgery is an increasingly performed procedure for the treatment of a wide spectrum of retinal pathologies. Yet, as most micro-surgical techniques, it requires long training periods before being mastered. To properly answer requests from clinicians for highly realistic training on one hand, and new requirements for accreditation or recertification from surgical societies on the other hand, we are developing a high-fidelity training system for retinal surgery. This simulator will be built upon our strong scientific expertise in the field of real-time simulation, and a success story for technology transfer in the field of cataract surgery simulation. Members of the consortium have a long expertise in the development of prototypes, as well as collaborations with clinical partners. The simulation system that we propose to develop is based on the Open Source simulation platform SOFA, and relies on expertise from our partners to ensure clinical and industrial relevance. This work is initially funded through the ANR project RESET which started in March 2015. A first version 13 of the training system has been delivered and we made a live demonstration at the Journée Alsacienne d'Ophtalmologie.

6.10. Virtual Cutting

The simulation of cutting is a central interest in the team. We especially work on the simulation of surgical cuts 14 , tearing and other separations of materials induced by surgical tools. On the one hand, we investigated the theoretical aspect: using the standard finite element method (FEM) combined with a re-meshing approach, we replace locally the current structure of the mesh in order to allow for a separation. On the other hand, we detected a separation in the motion of an object provided by a monocular video stream. With that detection, we can provide an augmented reality during the cutting and tearing of a deformable object.

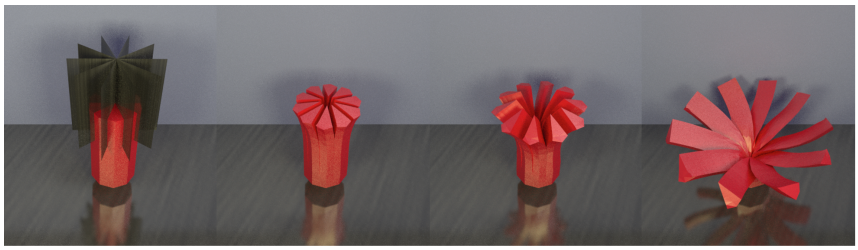


Figure 14. Our cutting algorithm in SOFA

The theoretical aspect of our work has been published in an article both at the conference Computer Graphics International CGI [24] and in the journal "The visual computer" [20]. The application in augmented reality 15 has been published at two conferences: "Augmented Reality during Cutting and Tearing of Deformable Objects", International Symposium on Mixed and Augmented Reality (ISMAR) [30].

To read more about our projects and results, please visit our website: http://mimesis.inria.fr.

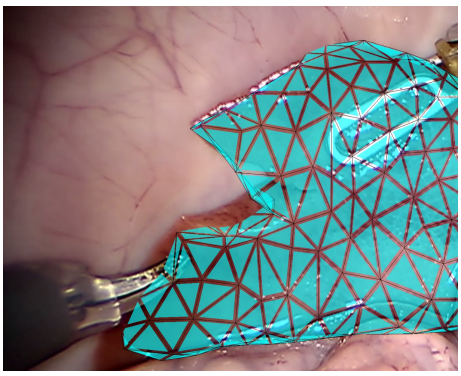


Figure 15. Augmented reality on a liver involving large deformation and cutting, i.e. topological changes

MNEMOSYNE Project-Team

7. New Results

7.1. Overview

Though our view is systemic, our daily research activities are also concerned with the design, at a given scale of description, of models of neuronal structures, each concerned with a specific learning paradigm. Of course, a major challenge is to integrate these elements in a systemic view, i.e. to put a specific emphasis on the way each neuronal structure communicates with the rest of the system and to highlight how its learning paradigms interact with other memory systems.

Among the numerous loops involving the brain, the body and the environment, a basic grid of description corresponds to distinguish "perception aspects of loops", the goal of which is to extract from the inner and outer world sensory invariants helpful to identify and evaluate the current state and to make predictions from previous learning, and "action aspects of loops", the goal of which is to rely on this sensory and emotional information to decide, plan and trigger actions for the benefit of the body.

This year, our team was engaged on the following topics: Concerning perception aspects of loops, we published original models of the amygdala and of the hippocampus and considered their role in pavlovian conditioning and their evaluation as classical models in machine learning. Concerning action aspects of loops, in addition to a critical analysis of the current views of the interactions between the prefrontal cortex and the basal ganglia [15], we have proposed an original model for the formation of habits and have also studied related theoretical problems in machine learning, for data representation. Finally, we also report here more methodological achievements, corresponding to the design of algorithmic ersatz of cerebral subsystems.

7.2. Pavlovian conditioning

Within perception aspects of loops, pavlovian conditioning is a very interesting learning paradigm to study in a systemic view because it is tightly related to other learning paradigms like episodic and semantic memory. This year, we have published papers presenting the biological basis of models of two fundamental structures in pavlovian conditioning, the amygdala [1] and the hippocampus [2]. We have also evaluated their most critical features, when considered as models of machine learning, namely their architecture and implementations at both rate and spiking levels [10] and their robustness to interference [11].

7.3. The formation of habits

Concerning action aspects of loops, we made important extensions to a model of basal ganglia that we developed recently [54] in interaction with another team in our neuroscience lab. In addition to extending the bio-plausibility of this model with an implementation at the spiking level, we have also developed this year a new theoretical framework that provides a novel explanation for the formation and the expression of habits in the cortex of primates by considering the basal ganglia as an implicit supervisor. This has been achieved with a model of basal ganglia running both at the rate and spiking levels. This framework predicts that Hebbian learning and reinforcement learning can be explicitly dissociated by inactivating the output of the basal ganglia during learning and later tested in normal conditions. Experimental results in the monkey confirmed this prediction and explain how a behavioral decision results from both the cooperation (acquisition) and competition (expression) of two distinct but entangled memory systems.

7.4. Beyond symbolic models

Using a biologically plausible model, we have been investigating some external and internal factors related to the stimulus representation that might affect the decision making and action selection [41]. We used a computational model of the cerebral structure Basal Ganglia, inspired and replicated from a model designed in previous studies [54]. One of the questions we attempt to address is to what extent the physical properties of the stimulus affect the decision to overcome the impact of reward associated to the stimuli.

7.5. Algorithmic ersatz of cerebral subsystems

As far as the systemic modeling and simulation of high-level brain functions are concerned (e.g., sensory-motor behavior, action selection and planning, perceptual categorization), we need to confront biologically plausible models at different scales of description with functional models that are not constrained by biological facts but still reproduce the expected functional response. This is mandatory to benchmark bio-physical models with respect to their equivalent in classical machine learning, in order to evaluate the degree of naiveness of their performances and also to build feasible simulation in which detailed biological models can interact with less plausible modules in order to be evaluated in realistic numerical situations.

This year, a set of formalism such as the Friston free-energy minimization general principle, deep-learning and related architectures, and more specific formalisms such as harmonic control or adaptive-subspace self-organizing maps have been studied and reviewed. The next step is to write a review, with the challenge of proposing an unifying view of those, and at a more concrete level, to propose the integration of a relevant subset of the related algorithms as a easily usable toolbox. This can be particularly useful to design global models of cognitive functions, even if biologically-inspired models are not yet available for all their components.

Preliminary key points regarding numerical experimentations have been published in this methodological paper [16].

NEUROMATHCOMP Project-Team

6. New Results

6.1. Neural Networks as dynamical systems

6.1.1. Periodic forcing of stabilized E-I networks: Nonlinear resonance curves and dynamics

Participants: Romain Veltz, Terry Sejnowski [Salk Institute].

Inhibition stabilized networks (ISNs) are neural architectures with strong positive feedback among pyramidal neurons balanced by strong negative feedback from in-hibitory interneurons, a circuit element found in the hippocampus and the primary visual cortex. In their working regime, ISNs produce damped oscillations in the γ-range in response to inputs to the inhibitory population. In order to understand the proper-ties of interconnected ISNs, we investigated periodic forcing of ISNs. We show that ISNs can be excited over a range of frequencies and derive properties of the resonance peaks. In particular, we studied the phase-locked solutions, the torus solutions and the resonance peaks. More particular, periodically forced ISNs respond with (possibly multi-stable) phase-locked activity whereas networks with sustained intrinsic oscillations respond more dynamically to periodic inputs with tori. Hence, the dynamics are surprisingly rich and phase effects alone do not adequately describe the network response. This strengthens the importance of phase-amplitude coupling as opposed to phase-phase coupling in providing multiple frequencies for multiplexing and routing information.

This work has been published in Neural Computation and is available as [29].

6.1.2. A new twist for the simulation of hybrid systems using the true jump method

Participant: Romain Veltz.

The use of stochastic models, in effect piecewise deterministic Markov processes (PDMP), has become increasingly popular especially for the modeling of chemical reactions and cell biophysics. Yet, exact simulation methods, for the simulation of these models in evolving environments, are limited by the need to find the next jumping time at each recursion of the algorithm. We report on a new general method to find this jumping time for the True Jump Method. It is based on an expression in terms of ordinary differential equations for which efficient numerical methods are available. As such, our new result makes it possible to study numerically stochastic models for which analytical formulas are not available thereby providing a way to approximate the state distribution for example. We conclude that the wide use of event detection schemes for the simulation of PDMPs should be strongly reconsidered. The only relevant remaining question being the efficiency of our method compared to the Fictitious Jump Method, question which is strongly case dependent.

This work is available as [55].

6.1.3. On the effects on cortical spontaneous activity of the symmetries of the network of pinwheels in visual area V1

Participants: Romain Veltz, Pascal Chossat, Olivier Faugeras.

This work challenges and extends earlier seminal work. We consider the problem of describing mathematically the spontaneous activity of V1 by combining several important experimental observations including (1) the organization of the visual cortex into a spatially periodic network of hypercolumns structured around pinwheels, (2) the difference between short-range and long-range intracortical connections, the first ones being rather isotropic and producing naturally doubly periodic patterns by Turing mechanisms, the second one being patchy, and (3) the fact that the Turing patterns spontaneously produced by the short-range connections and the network of pinwheels have similar periods. By analyzing the preferred orientation (PO) maps, we are able to classify all possible singular points (the pinwheels) as having symmetries described by a small subset of the

wallpaper groups. We then propose a description of the spontaneous activity of V1 using a classical voltage-based neural field model that features isotropic short-range connectivities modulated by non-isotropic long-range connectivities. A key observation is that, with only short-range connections and because the problem has full translational invariance in this case, a spontaneous doubly periodic pattern generates a 2-torus in a suitable functional space which persists as a flow-invariant manifold under small perturbations, for example when turning on the long-range connections. Through a complete analysis of the symmetries of the resulting neural field equation and motivated by a numerical investigation of the bifurcations of their solutions, we conclude that the branches of solutions which are stable over an extended range of parameters are those that correspond to patterns with an hexagonal (or nearly hexagonal) symmetry. The question of which patterns persist when turning on the long-range connections is answered by (1) analyzing the remaining symmetries on the perturbed torus and (2) combining this information with the Poincaré-Hopf theorem. We have developed a numerical implementation of the theory that has allowed us to produce the predicted patterns of activities, the planforms. In particular we generalize the contoured and non-contoured planforms predicted by previous authors.

This work has been published in Journal of Mathematical Neuroscience and is available as [27].

6.1.4. Biophysical reaction-diffusion model for stage II retinal waves and bifurcations analysis

Participants: Theodora Karvouniari, Bruno Cessac.

Retinal waves are spontaneous waves of spiking activity observed in the retina, during development only, playing a central role in shaping the visual system and retinal circuitry. Understanding how these waves are initiated and propagate in the retina could enable one to control, guide and predict them in the in vivo adult retina as inducing them is expected to reintroduce some plasticity in the retinal tissue and in the projections to the LGN. In this context, we propose a physiologically realistic reaction-diffusion model for the mechanisms of the emergence of stage II cholinergic retinal waves during development. We perform the bifurcation analysis when varying two biophysically relevant parameters, the conductances of calcium and potassium g_{Ca}, g_K respectively. The two main goals of our work are: firstly, reproduce the experimental recordings of developmental retinal waves by simulating our model and secondly, explore the different dynamical behaviours observed when varying these two parameters.

This work is available as [35].

6.1.5. Spatio-Temporal Linear Response of Spiking Neuronal Network Models

Participants: Rodrigo Cofré, Bruno Cessac.

We study the impact of a weak time-dependent external stimulus on the collective statistics of spiking responses in neuronal networks. We extend the current knowledge, assessing the impact over firing rates and cross correlations, to any higher order spatio-temporal correlation [1]. Our approach is based on Gibbs distributions (in a general setting considering non stationary dynamics and infinite memory) [2] and linear response theory. The linear response is written in terms of a correlation matrix, computed with respect to the spiking dynamics without stimulus. We give an example of application in a conductance based integrate-and fire model.

This work is available as [38].

6.1.6. Heteroclinic cycles in Hopfield networks

Participants: Pascal Chossat, Maciej Krupa.

It is widely believed that information is stored in the brain by means of the varying strength of synaptic connections between neurons. Stored patterns can be replayed upon the arrival of an appropriate stimulus. Hence, it is interesting to understand how an information pattern can be represented by the dynamics of the system. In this work, we consider a class of network neuron models, known as Hopfield networks, with a learning rule which consists of transforming an information string to a coupling pattern. Within this class of models, we study dynamic patterns, known as robust heteroclinic cycles, and establish a tight connection between their existence and the structure of the coupling.

This work has been published in Journal of Nonlinear Science and is available as [20].

6.2. Mean field approaches

6.2.1. Confronting mean-field theories to measurements: a perspective from neuroscience

Participant: Bruno Cessac.

Mean-field theories in neuroscience are usually understood as ways to bridge spatial and temporal scales by lumping together the activities of many single neurons, and then explaining or predicting the spatio-temporal variations of mesoscopic or macroscopic quantities measurable with current technologies: EEG, MEG, fMRI, optical imaging, etc. This is very much alike the situation in statistical physics where macroscopic quantities such as pressure, conductivity and so on are explained by the interactions between "microscopic" entities like atoms or molecules.

The situation in neuroscience is different however: the laws governing the microscopic dynamics in physics do not have the same structure as the laws governing neuronal dynamics; for example, interactions between neurons are not symmetric. Moreover, it is yet unclear what the relevant macroscopic quantities are in order to account for, say, visual perception. At the present stage of research, these quantities are considered to be what is measurable with currently available technologies, whereas better theories could reveal new types of phenomenological observables with a higher explanatory power.

We review mean-field methods coming from physics and their consequences on neuronal dynamics predictions.

This work is available as [30], [31], [32].

6.2.2. A Formalism for Evaluating Analytically the Cross-Correlation Structure of a Firing-Rate Network Model

Participants: Diego Fasoli, Olivier Faugeras, Stefano Panzeri.

We introduce a new formalism for evaluating analytically the cross-correlation structure of a finite size firing-rate network with recurrent connections. The analysis performs a first-order perturbative expansion of neural activity equations that include three different sources of randomness: the background noise of the membrane potentials, their initial conditions, and the distribution of the recurrent synaptic weights. This allows the analytical quantification of the relationship between anatomical and functional connectivity, i.e. of how the synaptic connections determine the statistical dependencies at any order among different neurons. The technique we develop is general, but for simplicity and clarity we demonstrate its efficacy by applying it to the case of synaptic connections described by regular graphs. The analytical equations obtained in this way reveal previously unknown behaviors of recurrent firing-rate networks, especially on how correlations are modified by the external input, by the finite size of the network, by the density of the anatomical connections and by correlation in sources of randomness. In particular, we show that a strong input can make the neurons almost independent, suggesting that functional connectivity does not depend only on the static anatomical connectivity, but also on the external inputs. Moreover we prove that in general it is not possible to find a mean-field description à la Sznitman of the network, if the anatomical connections are too sparse or our three sources of variability are correlated. To conclude, we show a very counterintuitive phenomenon, which we call stochastic synchronization, through which neurons become almost perfectly correlated even if the sources of randomness are independent. Due to its ability to quantify how activity of individual neurons and the correlation among them depends upon external inputs, the formalism introduced here can serve as a basis for exploring analytically the computational capability of population codes expressed by recurrent neural networks.

This work is available as [22].

6.2.3. Asymptotic Description of Neural Networks with Correlated Synaptic Weights

Participants: Olivier Faugeras, James Maclaurin.

We study the asymptotic law of a network of interacting neurons when the number of neurons becomes infinite. Given a completely connected network of neurons in which the synaptic weights are Gaussian correlated random variables, we describe the asymptotic law of the network when the number of neurons goes to infinity. We introduce the process-level empirical measure of the trajectories of the solutions to the equations of the finite network of neurons and the averaged law (with respect to the synaptic weights) of the trajectories of the solutions to the equations of the network of neurons. The main result of this article is that the image law through the empirical measure satisfies a large deviation principle with a good rate function which is shown to have a unique global minimum. Our analysis of the rate function allows us also to characterize the limit measure as the image of a stationary Gaussian measure defined on a transformed set of trajectories.

This work is available as [23].

6.2.4. *Clarification and Complement to "Mean-Field Description and Propagation of Chaos in Networks of Hodgkin-Huxley and FitzHugh-Nagumo Neurons"*

Participants: Mireille Bossy, Olivier Faugeras, Denis Talay.

In this work, we clarify the well-posedness of the limit equations to the mean-field N-neuron models proposed in [1] and we prove the associated propagation of chaos property. We also complete the modeling issue in [1] by discussing the well-posedness of the stochastic differential equations which govern the behavior of the ion channels and the amount of available neurotransmitters.

This work is available as [18].

6.3. Neural fields theory

6.3.1. *ERRATUM: A Center Manifold Result for Delayed Neural Fields Equations*

Participants: Romain Veltz, Olivier Faugeras.

Lemma C.1 in [95] is wrong. This lemma is used in the proof of the existence of a smooth center manifold, Theorem 4.4 in that paper. An additional assumption is required to prove this existence. We spell out this assumption, correct the proofs, and show that the assumption is satisfied for a large class of delay functions τ. We also weaken the general assumptions on τ.

This work has been published in SIAM J. Math. Anal. and is available as [28].

6.3.2. *A general framework for stochastic traveling waves and patterns, with application to neural field equations*

Participants: James Inglis, James Maclaurin.

In this work we present a general framework in which to rigorously study the effect of spatio- temporal noise on traveling waves and stationary patterns. In particular, the framework can incorporate versions of the stochastic neural field equation that may exhibit traveling fronts, pulses or stationary patterns. To do this, we first formulate a local SDE that describes the position of the stochastic wave up until a discontinuity time, at which point the position of the wave may jump. We then study the local stability of this stochastic front, obtaining a result that recovers a well-known deterministic result in the small-noise limit. We finish with a study of the long-time behavior of the stochastic wave.

This work has appeared in SIAM J. on Applied Dynamical Systems (SIADS) [49].

6.4. Slow-Fast Dynamics in Neural Models

6.4.1. *From Canards of Folded Singularities to Torus Canards in a Forced van der Pol Equation*

Participants: John Burke [Boston University, USA], Mathieu Desroches, Albert Granados [Technical University of Denmark, Lyngby, Denmark], Tasso Kaper [Boston University, USA], Maciej Krupa, Theodore Vo [Boston University, USA].

In this work, we study canard solutions of the forced van der Pol equation in the relaxation limit for low-, intermediate-, and high-frequency periodic forcing. A central numerical observation made herein is that there are two branches of canards in parameter space which extend across all positive forcing frequencies. In the low-frequency forcing regime, we demonstrate the existence of primary maximal canards induced by folded saddle nodes of type I and establish explicit formulas for the parameter values at which the primary maximal canards and their folds exist. Then, we turn to the intermediate- and high-frequency forcing regimes and show that the forced van der Pol possesses torus canards instead. These torus canards consist of long segments near families of attracting and repelling limit cycles of the fast system, in alternation. We also derive explicit formulas for the parameter values at which the maximal torus canards and their folds exist. Primary maximal canards and maximal torus canards correspond geometrically to the situation in which the persistent manifolds near the family of attracting limit cycles coincide to all orders with the persistent manifolds that lie near the family of repelling limit cycles. The formulas derived for the folds of maximal canards in all three frequency regimes turn out to be representations of a single formula in the appropriate parameter regimes, and this unification confirms the central numerical observation that the folds of the maximal canards created in the low-frequency regime continue directly into the folds of the maximal torus canards that exist in the intermediate- and high-frequency regimes. In addition, we study the secondary canards induced by the folded singularities in the low-frequency regime and find that the fold curves of the secondary canards turn around in the intermediate-frequency regime, instead of continuing into the high-frequency regime. Also, we identify the mechanism responsible for this turning. Finally, we show that the forced van der Pol equation is a normal form-type equation for a class of single-frequency periodically driven slow/fast systems with two fast variables and one slow variable which possess a non-degenerate fold of limit cycles. The analytic techniques used herein rely on geometric desingularisation, invariant manifold theory, Melnikov theory, and normal form methods. The numerical methods used herein were developed in Desroches et al. (SIAM J Appl Dyn Syst 7:1131–1162, 2008, Nonlinearity 23:739–765 2010).

This work has been published in J. Nonlinear Sci. and is available as [19].

6.4.2. *Extending the zero-derivative principle for slow-fast dynamical systems*

Participants: Eric Benoît [Université de La Rochelle, France], Morten Brøns [Technical University of Denmark, Lyngby, Denmark], Mathieu Desroches, Maciej Krupa.

Slow-fast systems often possess slow manifolds, that is invariant or locally invariant sub-manifolds on which the dynamics evolves on the slow time scale. For systems with explicit timescale separation, the existence of slow manifolds is due to Fenichel theory, and asymptotic expansions of such manifolds are easily obtained. In this work, we discuss methods of approximating slow manifolds using the so-called zero-derivative principle. We demonstrate several test functions that work for systems with explicit time scale separation including ones that can be generalized to systems without explicit timescale separation. We also discuss the possible spurious solutions, known as ghosts, as well as treat the Templator system as an example.

This work has been published in ZAMP and is available as [17].

6.4.3. *Canards, folded nodes and mixed-mode oscillations in piecewise-linear slow-fast systems*

Participants: Mathieu Desroches, Antoni Guillamon [Polytechnic University of Catalunya, Barcelona, Spain], Enrique Ponce [University of Sevilla, Spain], Rafel Prohens [University of the Balearic Islands, Palma, Spain], Serafim Rodrigues [Plymouth University, UK], Antonio Teruel [University of the Balearic Islands, Palma, Spain].

Canard-induced phenomena have been extensively studied in the last three decades, both from the mathematical and from the application viewpoints. Canards in slow-fast systems with (at least) two slow variables, especially near folded-node singularities, give an essential generating mechanism for Mixed-Mode oscillations (MMOs) in the framework of smooth multiple timescale systems. There is a wealth of literature on such slow-fast dynamical systems and many models displaying canard-induced MMOs, in particular in neuroscience. In parallel, since the late 1990s several papers have shown that the canard phenomenon can be faithfully reproduced with piecewise-linear (PWL) systems in two dimensions although very few results are available in the

three-dimensional case. This work aims to bridge this gap by analyzing canonical PWL systems that display folded singularities, primary and secondary canards, with a similar control of the maximal winding number as in the smooth case. We also show that the singular phase portraits are compatible in both frameworks. Finally, we show on an example how to construct a (linear) global return and obtain robust PWL MMOs.

This work has been accepted for publication in SIAM Review and is available as [46].

6.4.4. Canard solutions in planar piecewise linear systems with three zones

Participants: Soledad Fernández-García [Inria Paris-Rocquencourt, France], Mathieu Desroches, Maciej Krupa, Antonio Teruel [University of the Balearic Islands, Palma, Spain].

In this work, we analyze the existence and stability of canard solutions in a class of planar piecewise linear systems with three zones, using a singular perturbation theory approach. To this aim, we follow the analysis of the classical canard phenomenon in smooth planar slow-fast systems and adapt it to the piecewise-linear framework. We first prove the existence of an intersection between repelling and attracting slow manifolds, which defines a maximal canard, in a non-generic system of the class having a continuum of periodic orbits. Then, we perturb this situation and prove the persistence of the maximal canard solution, as well as the existence of a family of canard limit cycles in this class of systems. Similarities and differences between the piecewise linear case and the smooth one are highlighted.

This work has been published Dynam. Syst. and is available as [24].

6.4.5. Spike-adding mechanism in parabolic bursters: the role of folded-saddle canards

Participants: Mathieu Desroches, Maciej Krupa, Serafim Rodrigues [Plymouth University, UK].

The present work develops a new approach to studying parabolic bursting, and also proposes a novel four-dimensional canonical and polynomial-based parabolic burster. In addition to this new polynomial system, we also consider the conductance-based model of the Aplysia R15 neuron known as Plant's model, and a reduction of this prototypical biophysical parabolic burster to three variables, including one phase variable, namely Rinzel's theta model. Revisiting these models from the perspective of slow-fast dynamics reveals that the number of spikes per burst may vary upon parameter changes, however the spike-adding process occurs in a brutal (explosive) fashion that involves special solutions called canards. This spike-adding canard explosion phenomenon is analysed by using tools from geometric singular perturbation theory in tandem with numerical bifurcation techniques. We find that the bifurcation structure persists across both parabolic bursters, that is, spikes within the burst are incremented via the crossing of an excitability threshold given by a particular type of canard orbit, namely the strong canard of a folded-saddle singularity. Using these findings, we construct a new polynomial approximation of Plant's model, which retains all the key elements for parabolic bursting, including the canard mediated spike-adding transitions. Finally, we briefly investigate the presence of spike-adding via canards in planar phase models of parabolic bursting, namely the theta model by Ermentrout and Kopell.

This work has been submitted for publication and is available as [47].

6.4.6. Canards and spike-adding transitions in a minimal piecewise-linear Hindmarsh-Rose square-wave burster

Participants: Mathieu Desroches, Soledad Fernández-García [Inria Paris-Rocquencourt, France], Maciej Krupa.

We construct a piecewise-linear (PWL) approximation of the Hindmarsh-Rose (HR) neurone model that is minimal, in the sense that the vector field has the least number of pieces, in order to reproduce all the dynamics present in the original HR model with the classical parameter values. This includes spiking, square-wave bursting, and also special trajectories called canards, which possess long repelling segments and organise the transition between stable bursting patterns with n and $n+1$ spikes. This is the spike-adding canard explosion. We propose a first approximation of the smooth bursting model, using a continuous PWL system, and show that its fast subsystem cannot possess a homoclinic bifurcation, which is necessary to obtain proper square-wave bursting. We then relax the assumption of continuity of the vector field across all zones and show that

we can obtain a homoclinic bifurcation in the fast subsystem. We use the recently developed canard theory for PWL systems in order to reproduce the spike-adding canard explosion feature of the HR model as studied, e.g., in [66].

This work has been submitted for publication and is available as [45].

6.4.7. Ducks in space

Participants: Daniele Avitabile [University of Nottingham, UK], Mathieu Desroches, Edgar Knobloch [University of California at Berkeley, USA], Maciej Krupa.

A subcritical pattern-forming system with nonlinear advection in a bounded domain is recast as a slow-fast system in space and studied using a combination of geometric singular perturbation theory and numerical continuation. Two types of solutions describing the possible location of stationary fronts are identified, one of which is present for all values of the bifurcation parameter while the other is present for zero or sufficiently small inlet boundary conditions but only when the bifurcation parameter is large enough. For slightly larger inlet boundary condition a continuous transition from one type to the other takes place as the bifurcation parameter increases. The origin of the two solution types is traced to the onset of convective and absolute instability on the real line. The role of canard trajectories in the transitions between these states is clarified and the stability properties of the resulting spatial structures are determined. Front location in the convective regime is highly sensitive to the upstream boundary condition and its dependence on this boundary condition is studied using a combination of numerical continuation and Monte Carlo simulations of the partial differential equation. Statistical properties of the system subjected to random or stochastic boundary conditions are interpreted using the deterministic slow-fast spatial-dynamical system.

This work has been submitted for publication and is available as [43].

6.5. Spike Train statistics

6.5.1. Statistical models for spike trains analysis in the retina.

Participant: Bruno Cessac.

Recent advances in multi-electrodes array acquisition have made it possible to record the activity of up to several hundreds of neurons at the same time and to register their collective activity (spike trains). For the retina, this opens up new perspectives in understanding how retinal structure and ganglion cells encode information about a visual scene and what is transmitted to the brain. Especially, two paradigms can be confronted: in the first one, ganglion cells encode information independently of each others; in the second one non linear dynamics and connectivity contribute to produce a population coding where spatio-temporal correlations, although weak, play a significant role in spike coding. Confronting these two paradigms can be done at an experimental and at a theoretical level. On experimental grounds, new methods to analyse the role of weak correlations in spike train statistics are required. On theoretical grounds, mathematical results have been established, in neuronal models, showing how non linear dynamics and connectivity contribute to produce a correlated spike response to stimuli. In the context of the ANR KEOPS project, we have been working on these two aspects and we present our main results.

This work is available as [33].

6.5.2. Spectral dimension reduction on parametric models for spike train statistics

Participants: Cesar Ravello, Ruben Herzog, Bruno Cessac, Maria-Jose Escobar, Adrian Palacios.

It has been shown that the neurons of visual system present correlated activity in response to different stimuli. The role of these correlations is an unresolved subject. These correlations vary according to the stimulus, specially with natural images. To uncover the role of these correlation and characterize the population code, it is necessary to measure the simultaneous activity of large neural populations. This has been achieved thanks to the advent of Multi-Electrode Array technology, opening up a way to better characterize how the brain encodes information in the concerted activity of neurons. In parallel, powerful statistical tools have been developed to accurately characterize spatio-temporal correlations between neurons. Methods based on *Maximum Entropy Principle*, where statistical entropy is maximized under a set of constraints corresponding to specific assumptions on the relevant statistical quantities, have been proved successfully, specially when they consider *spatiotemporal* correlations. They are although limited by (i) **the assumption of stationarity**, (ii) **the many possible choice of constraints**, and (iii) **the huge number of free parameters**. We present our results on these two aspects obtained in the context of ANR KEOPS.

This work is available as [54].

6.6. Visual Neuroscience

6.6.1. Shifting stimulus for faster receptive fields estimation of ensembles of neurons

Participants: Bruno Cessac, Matthias Hennig [University of Edinburg, UK], Gerrit Hilgen [Institute of Neuroscience, Medical School, Newcastle University, Newcastle, UK], Pierre Kornprobst, Daniela Pamplona, Sahar Pirmoradian [University of Edinburg, UK], Evelyne Sernagor [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK].

The Spike Triggered Average (STA) is a classical technique to find a discrete approximation of the Receptive Fields (RFs) of sensory neurons [63], a required analysis in most experimental studies. One important parameter of the STA is the spatial resolution of the estimation, corresponding to the size of the blocks of the checkerboard stimulus images. In general, it is experimentally fixed to reach a compromise: If too small, neuronal responses might be too weak thus leading to RF with low Signal-to-Noise-Ratio; on the contrary, if too large, small RF will be lost, or not described with enough details, because of the coarse approximation. Other solutions were proposed consisting in starting from a small block size and updating it following the neuron response in a closed-loop to increase its response [70], [78], [77]. However, these solutions were designed for single cells and cannot be applied to simultaneous recordings of ensembles of neurons (since each RF has its own size and preferred stimulus).

To solve this problem, we introduced a modified checkerboard stimulus where blocks are shifted randomly in space at fixed time steps. This idea is inspired from super-resolution techniques developed in image processing [84]. The main interest is that the block size can be large, enabling strong responses, while the resolution can be finer since it depends on the shift minimum size. In [52], we show that the STA remains an unbiased RF estimator and, using simulated spike trains from an ensemble of Linear Nonlinear Poisson cascade neurons, it was predicted that this approach improves RF estimation over the neuron ensemble, in terms of resolution and convergence. In [53], we test these predictions experimentally on the RFs estimation of 8460 ganglion cells from two mouse retinas, using recordings performed with a large scale high-density multielectrode array. We compare RFs obtained using (i) the classical checkerboard stimulus with block size of 160μm and (ii) our checkerboard stimulus with block size of 160μm and arbitrary shifts of 40μm in $x-$ and $y-$directions. Results show how spatial resolution can be improved and that our approach allows to recover 51% of the mapped RFs at a resolution of 40μm, while in the classical case, 41% of the RFs could be found at a resolution of only 160μm. Thus, our approach improves not only the quality of the RF estimation but also the amount of successfully mapped RFs in neural ensembles.

This work was presented in [52], [53] and it is being used in current experimental protocols by E. Sernagor (Newcastle University), partner of the EC IP project FP7-ICT-2011-9 no. 600847 (RENVISION).

6.6.2. Using neural mechanisms underlying motion analysis for optical flow estimation

Participants: Manuela Chessa [University of Genoa, DIBRIS, Italy], Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone, Team InVibe], Kartheek Medathati, Fabio Solari [University of Genoa, DIBRIS, Italy].

We explore how motion information, also called optical flow, is estimated from natural moving sequences. Owing to application potential, optical flow estimation has been studied extensively by computer vision. On the other hand the neural mechanisms underlying motion analysis in the visual cortex have been extensively studied almost with little interaction with computer vision community resulting in few mathematical models. Even though there was some early interaction among the two communities for example, methods by Heeger et.al, Sejnowski et. al, comparatively little work has been done in terms of examining or extending the mathematical models proposed in biology in terms of their engineering efficacy on modern optical flow estimation datasets.

Pursuing this idea, in [26], we proposed a neural model inspired from the ones presented in [87], [86] which are popular models of primate velocity encoding. We started from a classical V1-MT feedforward architecture. We modeled V1 cells by motion energy (based on spatio-temporal filtering), and MT pattern cells (by pooling V1 cell responses). The efficacy of this architecture and its inherent limitations in the case of real videos were not known. To answer this question, we proposed a velocity space sampling of MT neurones (using a decoding scheme to obtain the local velocity from their activity) coupled with a multi-scale approach. After this, we explored the performance of our model on the Middlebury dataset. To the best of our knowledge, this is the only neural model in this dataset. The results were promising and suggested several possible improvements, in particular to better deal with discontinuities. An extension was proposed in [40].

We also focused on the decoding the motion energies which is of natural interest for developing biologically inspired computer vision algorithms for dense optical flow estimation. In [37], we addressed this problem by evaluating four strategies for motion decoding: intersection of constraints, maximum likelihood, linear regression on MT responses and neural network based regression using multi scale-features. We characterized the performances and the current limitations of the different strategies, in terms of recovering dense flow estimation using Middlebury benchmark dataset widely used in computer vision, and we highlight key aspects for future developments.

This work was partially funded by the EC IP project FP7-ICT-388 2011-8 no. 318723 (MatheMACS).

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

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

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

In [42], we aim to bridge this gap by providing a computer vision task centric presentation of models primarily originating in biological vision studies. Not only we revisit some of the main features of biological vision and discuss the foundations of existing computational studies modelling biological vision, but also consider three classical computer vision tasks from a biological perspective: image sensing, segmentation and optical flow. Using this task-centric approach, we discuss well-known biological functional principles and compare them with approaches taken by computer vision. Based on this comparative analysis of computer and biological vision, we present some recent models in biological vision and highlight a few models that we think are promising for future investigations in computer vision. To this extent, this paper provides new insights and a starting point for investigators interested in the design of biology-based computer vision algorithms and pave a way for much needed interaction between the two communities leading to the development of synergistic models of artificial and biological vision.

[42] is under review. This work was partially funded by the EC IP project FP7-ICT-388 2011-8 no. 318723 (MatheMACS).

NEUROSYS Project-Team

7. New Results

7.1. From the Microscopic to the Mesoscopic Scale

Participants: Laure Buhry, Axel Hutt, Francesco Giovannini, Jean-Baptiste Schneider
In collaboration with LieJune Shiau (University of Houston)

7.1.1. Memory and Anaesthesia

7.1.1.1. Hippocampal Memory Networks

To improve our understanding of the effects of anaesthesia on the neural correlates of memory, we focussed on how anaesthetics disrupt the interaction between the hippocampus and the cerebral cortex. As a first step towards this objective Francesco Giovannini modelled a hippocampal pyramidal neuron using the Hodgkin-Huxley model capable of exhibiting long-lasting persistent firing activity when subject to a strong transient stimulus [16]. This behaviour is underlay by an intrinsic membrane current activated by the increase of intracellular calcium ions, following the discharge of an action potential by the neuron, in accord with that displayed in neural recordings of hippocampal slice preparations. Connecting these persistent firing neurons in a network comprising strong local excitation yields a wide range of behaviours depending on the interaction between CAN and synaptic currents. Indeed, the network model is capable of displaying rhythmic behaviour in the form of short synchronised bursts with intra-burst frequencies of $20 - 40\,Hz$ and inter-burst frequencies of $3\,Hz$. Furthermore, coupling CAN-equipped pyramidal neurons with a population of fast-spiking inhibitory interneurons yields emerging synchronous activity whose frequency is modulated by the strength of this coupling. These results hint towards a possible mechanism for the generation of memory-related oscillatory activity in the hippocampus.

7.1.1.2. Anaesthetic Effects on Hippocampal Oscillations

We investigated the effects of propofol-mediated tonic inhibition on the synchronous activity elicited in a network of hippocampal inhibitory interneurons. This work was conducted in collaboration with Jean-Baptiste Schneider, as part of hist 2-month internship. We studied the effect of propofol-induced tonic inhibition on the oscillations elicited in a network of hippocampal Hodgkin-Huxley gamma-aminobutyric acid ($GABA$) interneurons by studying the action of propofol on extrasynaptic GABAergic receptors. Our results [15] show that increasing doses of propofol reduce the overall network activity and slow down its oscillations until a critical value at which the synchronisation increases abruptly at values of twice the synchronisation displayed in the absence of tonic inhibition, and the mean firing rate increases. This emergence of synchronous activity mediated by anaesthetic perfusion point towards a possible mechanism for the emergence of paradoxical excitation under general anaesthesia.

In this context, Laure Buhry works with LieJune Shiau (University of Houston) on a better understanding of the models used by the community of computational neuroscientists. The goal is to show in which extent models are comparable or interchangeable. We focus on the comparison of oscillatory mechanisms of neuronal populations in different spiking models, especially in the Hodgkin-Huxley and the adaptive exponential integrate-and-fire model (AdEx). Especially, we have shown that a same number of synaptic connection per neuron is necessary to elicit synchronization in inhibitory neural networks of adaptive exponential integrate and fire neurons as in networks of Hodgkin-Huxley neurons. We have also conducted an extensive study regarding the effects of the different parameters of the AdEx model on the synchronization mechanisms in inhibitory neural networks, particularly in the context of gamma oscillations. A manuscript will be submitted soon to the Journal of Computational Neuroscience.

7.1.1.3. Noise Effects on Neural Rhythms

We have continued working on the effect of additive noise on neural oscillations and have shown that additive noise modulates the frequency of self-sustained neural rhythms [3].

7.2. From the Mesoscopic to the Macroscopic Scale

Participants: Laurent Bougrain, Axel Hutt, Pedro Garcia-Rodriguez, Eric Nichols, Guillaume Serrière, Tamara Tosic, Mariia Fedotenkova, Meysam Hashemi, Benjamin le Golvan, Cecilia Lindig-Leon, Sébastian Rimbert.

7.2.1. Level of Consciousness

7.2.1.1. Spatio-temporal Dynamics in Neural Fields

Neural fields serve as a model for experimental macroscopic activity. We have developed a numerical simulator NeuralFieldSimulator [21]. In addition, we have worked out a neural neural field model that exhibits a sequence of metastable activity states as observed in experimental data [4].

7.2.1.2. Synchronisation in Local Field Potentials under Anaesthesia

We have applied advanced data analysis techniques based on wavelet analysis to detect instantaneous partial synchronisation in experimental data [5].

7.2.1.3. Statistical Frequency-dependent Analysis by Recurrence Plots

Participants : Axel Hutt, Mariia Fedotenkova, Tamara Tosic
In collaboration with Flavio Frohlich, Peter Beim Graben and Kristin K. Sellers
For decades, research in neuroscience has supported the hypothesis that brain dynamics exhibits recurrent metastable states connected by transients, which together encode fundamental neural information processing. To understand the system's dynamics it is important to detect such recurrence domains, but it is challenging to extract them from experimental neuroscience datasets due to the large trial-to-trial variability. We proposed a methodology to extract recurrent metastable states in univariate time series by transforming datasets into their time-frequency representations and computing recurrence plots based on instantaneous spectral power values in various frequency bands [6]. Additionally, a new statistical inference analysis compares different trial recurrence plots with corresponding surrogates to obtain statistically significant recurrent structures. This combination of methods is validated by applying it to two artificial datasets. In a final study of visually-evoked Local Field Potentials in partially anesthetized ferrets, the methodology is able to reveal recurrence structures of neural responses with trial-to-trial variability. Focusing on different frequency bands, the delta-band activity is much less recurrent than alpha-band activity. Moreover, alpha-activity is susceptible to pre-stimuli, while delta-activity is much less sensitive to pre-stimuli. This difference in recurrence structures in different frequency bands indicates diverse underlying information processing steps in the brain.

7.2.2. Motor System

Participants: Laurent Bougrain, Axel Hutt, Benjamin le Golvan, Cecilia Lindig-Leon, Sébastian Rimbert, Guillaume Serrière

7.2.2.1. Motor Patterns during General Anesthesia

Participants: Laurent Bougrain, Axel Hutt, Cecilia Lindig-Leon, Sébastian Rimbert, Guillaume Serrière
The dosage of the anesthetic agent is tricky: too low, it does not achieve a sufficient loss of consciousness and may lead to a partial memorization during surgery and a post-operative trauma; too strong, it is dangerous for people with respiratory or heart problems. To better monitor the effect of the current dosage, we propose to study the dynamics of the motor brain activity during anesthesia. The relationship between motor brain activity and anesthesia is not intensively studied. Yet even if no physical movement by the patient is visually detectable, an electroencephalographic analysis of brain activity in motor areas may reveal an intention movement. This information is important because it demonstrates that the patient is conscious. We started to define a clinical protocol in collaboration with anesthesiologists of the hospital in Nancy to investigate is possibility. To reduce the duration of the protocol, we studied the minimum duration of a motor imagery to allow its detection from EEG recordings [23]. A large number of Brain-Computer Interfaces (BCIs) are based on the detection of motor imagery related features in the electroencephalographic signal. In most BCI experimental paradigms, subjects realize continuous motor imagery, i.e. a prolonged intention of movement, during a time window of a few seconds. Then, the system detects the movement based on the event-related desynchronization (ERD) and the event-related synchronization (ERS) principles. We studied if a discrete motor imagery, corresponding to a single short motor imagery, would allow a better detection of ERD and ERS patterns than a continuous motor

imagery. Indeed, the results of experiments involving 11 healthy subjects suggest that a continuous motor imagery generates a later ERS as well as a more variable and less detectable ERD than discrete motor imagery [11]. This finding suggests an improved experimental paradigm. We deeper investigated the amplitude and latency of EEG Beta activity during real movements, discrete and continuous motor imageries [22].

7.2.2.2. Motor Patterns during Combined Movements

Participants: Laurent Bougrain, Cecilia Lindig-Leon
Imaginary motor tasks cause brain oscillations that can be detected through the analysis of electroencephalographic (EEG) recordings. We studied whether or not the characteristics of the brain activity induced by the combined motor imagery (MI) of both hands can be assumed as the superposition of the activity generated during simple hand MIs. After analyzing the sensorimotor rhythms in EEG signals of five healthy subjects, results show that the imagination of both hands movement generates in each brain hemisphere similar activity as the one produced by each simple hand MI in the contralateral side [8]. Furthermore, during simple hand MIs, brain activity over the ipsilateral hemisphere presents similar characteristics as those observed during the rest condition. Thus, it is shown that the proposed scheme is valid and promising for brain-computer interfaces (BCI) control, allowing to easily detect patterns induced by combined MIs. Based on these results, we proposed a new method to extend the classic Common Spatial Pattern (CSP) algorithm to a multi-class approach which analyses both brain hemispheres separately to solve, together with a stepwise classification strategy, a multi-label BCI problem. After testing the proposed approach over the EEG signals of six healthy subjects performing a four-class multi-label task involving simple and combined hand MIs together with the rest condition, results show that this technique is plausible for BCI control [7]. In terms of accuracy, it outperforms the classical one-vs-one approach by 20% and has the same performance as the one-vs-all method. Nevertheless, to solve a multi-label classification problem involving k classes, the proposed method requires only log2(k) classifiers, whereas the one-vs-one method uses k(k-1)/2 classifiers and the one-vs-all k classifiers, thereby the new approach simplifies the classification task and seems promising for solving multi-label problems involving numerous classes.

7.2.2.3. On-line Detection of the End of Motor Imageries

Participants : Cécilia Lindig-León, Laurent Bougrain and Sébastien Rimbert
Limb movement execution or imagination induce sensorimotor rhythms that can be detected in electroencephalographic (EEG) recordings. We presented the interest of considering not only the beta frequency band but also the alpha band to detect the elicited EEG rebound, i.e. the increasing of oscillatory power synchronization, at the end of motor imageries [9], [19]. This phenomenon can be stronger over the alpha than the beta band and it is experimentally demonstrated [9] that the analysis on the alpha band improves the detection of the end of motor imageries. Moreover a variant method to compute the oscillatory power without referring to a baseline period is proposed; such capacity is useful for self-paced BCI control.

7.2.3. Pain under General Anaesthesia

7.2.3.1. Detection of EEG-signal Features for Pain under General Anaesthesia

Participants : Axel Hutt, Mariia Fedotenkova
In collaboration with Peter Beim Graben and James W. Sleigh
Nowadays, surgical operations are impossible to imagine without general anaesthesia, which involves loss of consciousness, immobility, amnesia and analgesia. Understanding mechanisms underlying each of these effects guarantees well-controlled medical treatment. Our work focuses on analgesia effect of general anaesthesia, more specifically, on patients reaction on nociception stimuli. The study was conducted on dataset consisting of 230 EEG signals: pre- and post-incisional recordings for 115 patients, who received desflurane and propofol. Initial analysis was performed by power spectral analysis, which is a widespread approach in signal processing. Power spectral information was described by fitting the background activity and measuring power contained in delta and alpha bands according to power of background activity. The fact that power spectrum of background activity decays as frequency increasing is well known and thoroughly studied. Here, traditional $1/f^{\alpha}$ behaviour of the decay was replaced by a Lorentzian model to describe the power spectrum of background activity. Due to observed non-stationary nature of EEG signals spectral analysis does not suffice

to reveal significant changes between two states. A further improvement was done by expanding spectra with time information. To obtain time-frequency representations of the signals conventional spectrograms were used as well as a spectrogram reassignment technique. The latter allows to ameliorate readability of a spectrogram by reassigning energy contained in spectrogram to more precise positions. Subsequently, obtained spectrograms were used in recurrence analysis and its quantification by complexity measure. Recurrence analysis allows to describe and visualise dynamics of a system and discover structural patterns contained in the data. Structure of each recurrence plot is characterised by Lempel–Ziv complexity measure [5], which shows a difference between pre- and post-incision [13].

PARIETAL Project-Team

7. New Results

7.1. Semi-Supervised Factored Logistic Regression for High-Dimensional Neuroimaging Data

Imaging neuroscience links human behavior to aspects of brain biology in ever-increasing datasets. Existing neuroimaging methods typically perform either discovery of unknown neural structure or testing of neural structure associated with mental tasks. However, testing hypotheses on the neural correlates underlying larger sets of mental tasks necessitates adequate representations for the observations. We therefore propose to blend representation modelling and task classification into a unified statistical learning problem. A multinomial logistic regression is introduced that is constrained by factored coefficients and coupled with an au-toencoder. We show that this approach yields more accurate and interpretable neural models of psychological tasks in a reference dataset, as well as better generalization to other datasets.

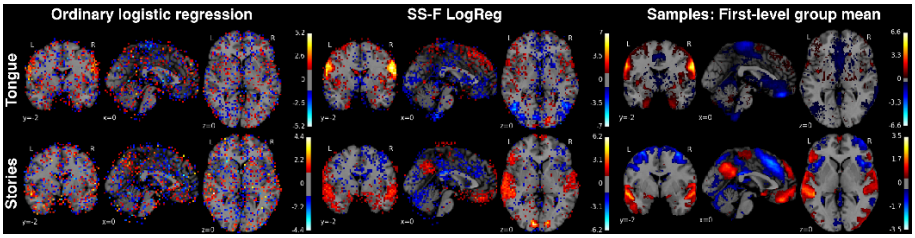


Figure 3. Classification weight maps. The voxel predictors corresponding to 2 exemplary (of 18 total) psychological tasks (rows) from the Human Connectome Project dataset. Left column: multinomial logistic regression (same implementation but without bottleneck or autoencoder), middle column: Semi-Supervised Factored Logistic Regression (SSFLogReg), right column: voxel-wise average across all samples of whole-brain activity maps from each task. SSFLogReg puts higher absolute weights on relevant structure, lowers ones on irrelevant structure, and yields BOLD-typical local con- tiguity (without enforcing an explicit spatial prior). More information can be found in [50].

More information can be found in [50].

7.2. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain

Here we present NeuroVault — a web based repository that allows researchers to store, share, visualize, and decode statistical maps of the human brain. NeuroVault is easy to use and employs modern web technologies to provide informative visualization of data without the need to install additional software. In addition, it leverages the power of the Neurosynth database to provide cognitive decoding of deposited maps. The data are exposed through a public REST API enabling other services and tools to take advantage of it. NeuroVault is a new resource for researchers interested in conducting meta- and coactivation analyses.

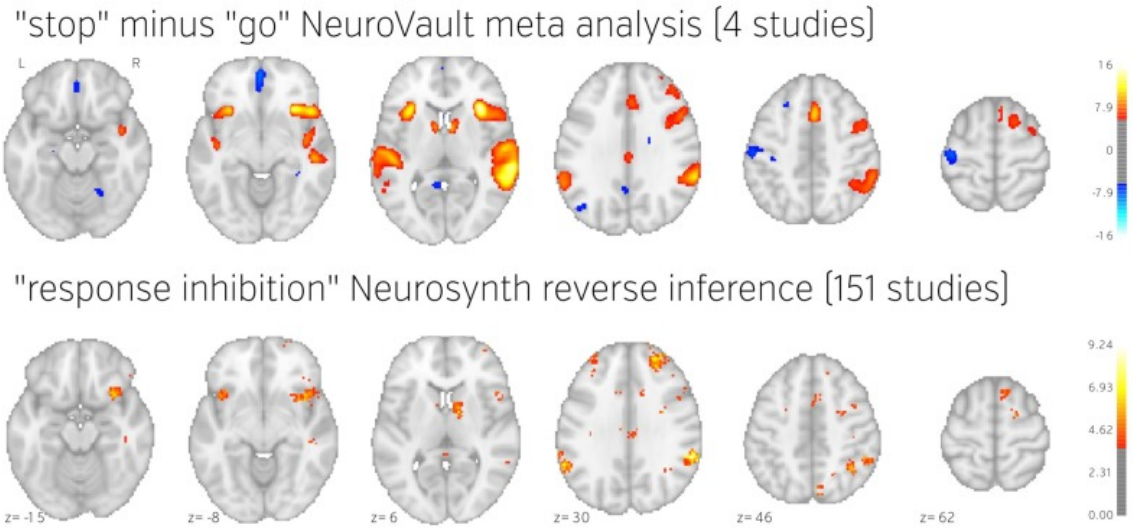


Figure 4.

Comparison of image based and coordinate based meta analysis of response inhibition. Meta analysis based on unthresholded statistical maps obtained from NeuroVault (top row) managed to recover the pattern of activation obtained using traditional methods despite including much fewer studies. NeuroVault map has been thresholded at $z = 6$, response inhibition map has been thresholded at $z = 1.77$ (the threshold values were chosen for visualization purposes only, but both are statistically significant at $p < 0.05$). Unthresholded versions of these maps are available at http://neurovault.org/collections/439/

More information can be found in [18] and [17].

7.3. FAASTA: A fast solver for total-variation regularization of ill-conditioned problems with application to brain imaging

The total variation (TV) penalty, as many other analysis-sparsity problems, does not lead to separable factors or a proximal operator with a closed-form expression, such as soft thresholding for the $\ell 1$ penalty. As a result, in a variational formulation of an inverse problem or statistical learning estimation, it leads to challenging non-smooth optimization problems that are often solved with elaborate single-step first-order methods. When the data-fit term arises from empirical measurements, as in brain imaging, it is often very ill-conditioned and without simple structure. In this situation, in proximal splitting methods, the computation cost of the gradient step can easily dominate each iteration. Thus it is beneficial to minimize the number of gradient steps. We present fAASTA, a variant of FISTA, that relies on an internal solver for the TV proximal operator, and refines its tolerance to balance computational cost of the gradient and the proximal steps. We give benchmarks and illustrations on “brain decoding”: recovering brain maps from noisy measurements to predict observed behavior. The algorithm as well as the empirical study of convergence speed are valuable for any non-exact proximal operator, in particular analysis-sparsity problems.

Convergence of currently available optimization algorithms, for 3 scenarios, with weak, medium and strong regularization, where medium regularization corresponds to the value chosen by cross-validation. These are log-log plots with the 0 defined as the lowest energy value reached across all algorithms.

More information can be found in [47].

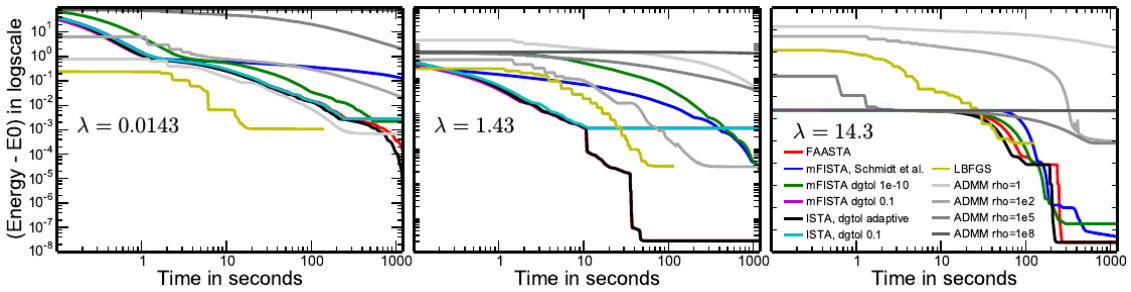


Figure 5.

7.4. Bootstrapped Permutation Test for Multiresponse Inference on Brain Behavior Associations.

Despite that diagnosis of neurological disorders commonly involves a collection of behavioral assessments, most neuroimaging studies investigating the associations between brain and behavior largely analyze each behavioral measure in isolation. To jointly model multiple behavioral scores, sparse mul-tiresponse regression (SMR) is often used. However, directly applying SMR without statistically controlling for false positives could result in many spurious findings. For models, such as SMR, where the distribution of the model parameters is unknown, permutation test and stability selection are typically used to control for false positives. In this paper, we present another technique for inferring statistically significant features from models with unknown parameter distribution. We refer to this technique as bootstrapped permutation test (BPT), which uses Studentized statistics to exploit the intuition that the variability in parameter estimates associated with relevant features would likely be higher with responses permuted. On synthetic data, we show that BPT provides higher sensitivity in identifying relevant features from the SMR model than permutation test and stability selection, while retaining strong control on the false positive rate. We further apply BPT to study the associations between brain connec-tivity estimated from pseudo-rest fMRI data of 1139 fourteen year olds and be-havioral measures related to ADHD. Significant connections are found between brain networks known to be implicated in the behavioral tasks involved. Moreover , we validate the identified connections by fitting a regression model on pseudo-rest data with only those connections and applying this model on resting state fMRI data of 337 left out subjects to predict their behavioral scores. The predicted scores are shown to significantly correlate with the actual scores of the subjects, hence verifying the behavioral relevance of the found connections.

Real data results: Statistically significant connectivity differences between populations (a) Significant network connections found on pseudo-rest fMRI data. (b) Pearson's correlation between predicted and actual scores with p-values noted. Each set of three bars (top to bottom) correspond to spatial working memory strategy, spatial working memory between errors, and rapid visual information processing accuracy scores. Significance is declared at $p < 0.05$.

More information can be found in [43].

7.5. Total Variation meets Sparsity: statistical learning with segmenting penalties

Prediction from medical images is a valuable aid to diagnosis. For instance, anatomical MR images can reveal certain disease conditions, while their functional counterparts can predict neuropsychiatric phenotypes. However, a physician will not rely on predictions by black-box models: understanding the anatomical or functional features that underpin decision is critical. Generally, the weight vectors of classifiers are not

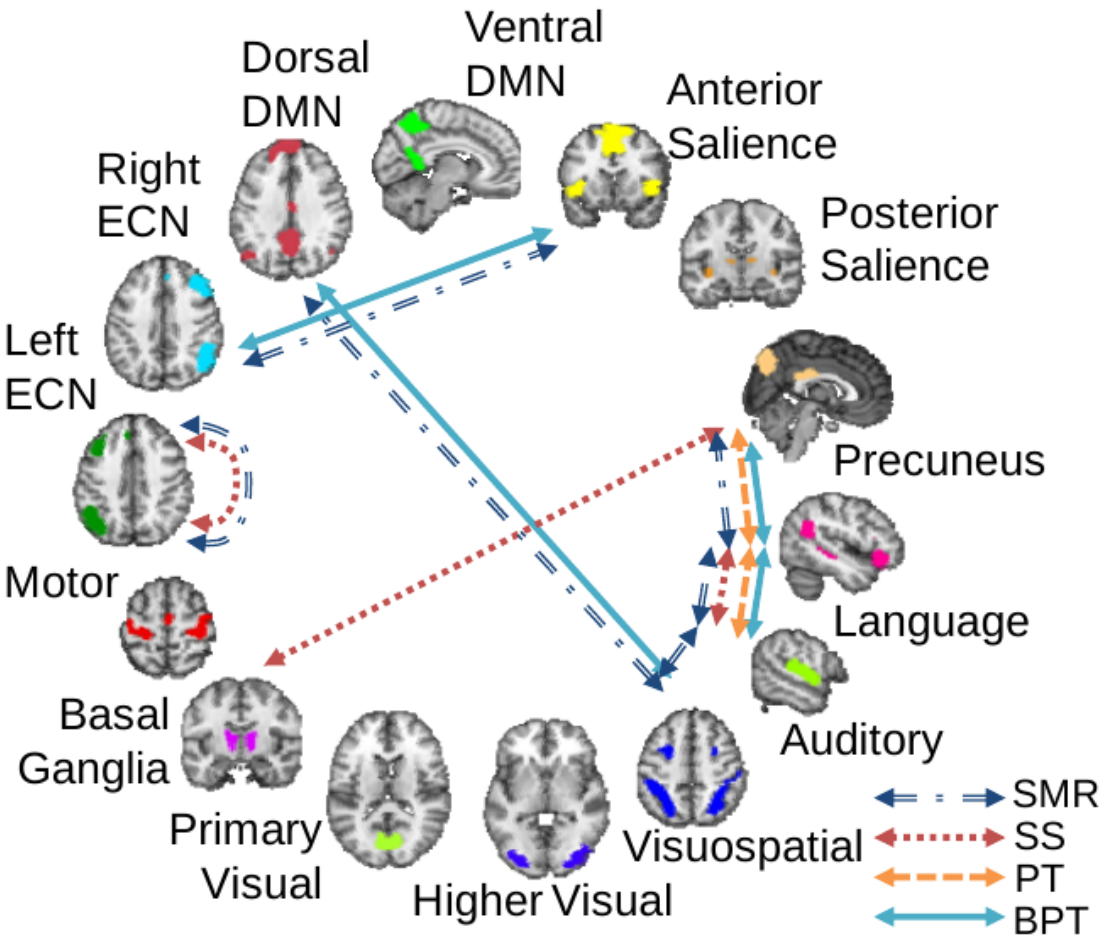

Dorsal DMN
Ventral DMN
Anterior Salience
Right ECN
Posterior Salience
Left ECN
Precuneus
Motor
Language
Basal Ganglia
Auditory
Primary Visual
Higher Visual
Visuospatial
SMR
SS
PT
BPT

Figure 6.

easily amenable to such an examination: Often there is no apparent structure. Indeed, this is not only a prediction task, but also an inverse problem that calls for adequate regularization. We address this challenge by introducing a convex region-selecting penalty. Our penalty combines total-variation regularization, enforcing spatial contiguity, and 1 regularization, enforcing sparsity, into one group: Voxels are either active with non-zero spatial derivative or zero with inactive spatial derivative. This leads to segmenting contiguous spatial regions (inside which the signal can vary freely) against a background of zeros. Such segmentation of medical images in a target-informed manner is an important analysis tool. On several prediction problems from brain MRI, the penalty shows good segmentation. Given the size of medical images, computational efficiency is key. Keeping this in mind, we contribute an efficient optimization scheme that brings significant computational gains.

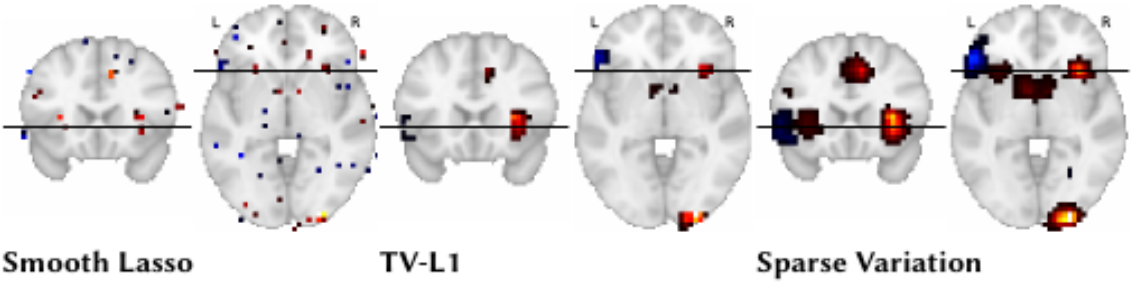


Figure 7.

Weight vectors from estimating gain on the mixed gambles task for three sparse methods: Graphnet, TV-l1 and Sparse Variation. This inter- subject analysis shows broader regions of activation. Mean correlation scores on held out data: Graphnet: 0.128, TV-l1 : 0.147, Sparse Variation: 0.149. One can see that both TV-l1 and Sparse Variation regularizations yield more interpreatble patterns than Graphnet.

More information can be found in [40].

7.6. Improving sparse recovery on structured images with bagged clustering

The identification of image regions associated with external variables through discriminative approaches yields ill-posed estimation problems. This estimation challenge can be tackled by imposing sparse solutions. However, the sensitivity of sparse estimators to correlated variables leads to non-reproducible results, and only a subset of the important variables are selected. In this paper, we explore an approach based on bagging clustering-based data compression in order to alleviate the instability of sparse models. Specifically, we design a new framework in which the estimator is built by averaging multiple models estimated after feature clustering, to improve the conditioning of the model. We show that this combination of model averaging with spatially consistent compression can have the virtuous effect of increasing the stability of the weight maps, allowing a better interpretation of the results. Finally, we demonstrate the benefit of our approach on several predictive modeling problems.

Z-score obtained across bootstraps for two discriminative tasks, using the candidate approaches. Higher values hint at lower variability across bootstrap replications. SCLR decreases the variability and yields larger standardized effects.

More information can be found in [42].

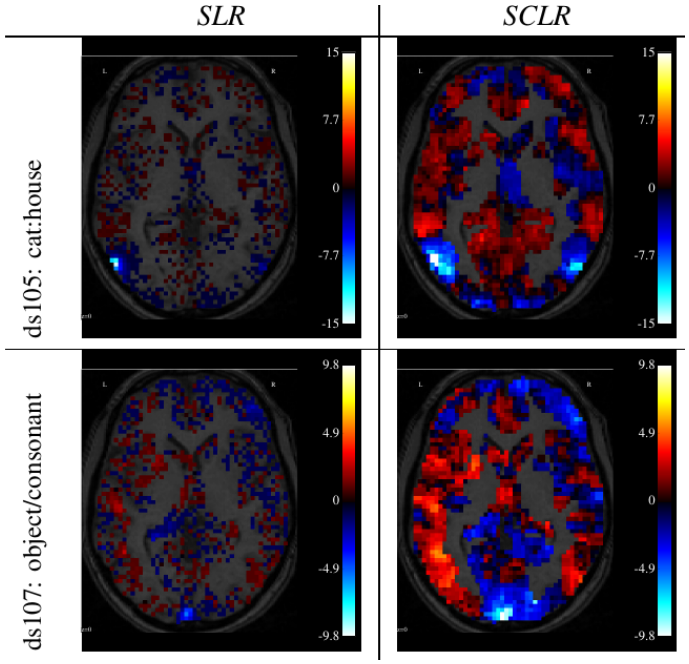


Figure 8.

7.7. Integrating Multimodal Priors in Predictive Models for the Functional Characterization of Alzheimer's Disease

Functional brain imaging provides key information to characterize neurodegenerative diseases, such as Alzheimer's disease (AD). Specifically, the metabolic activity measured through fluorodeoxyglu-cose positron emission tomography (FDG-PET) and the connectivity extracted from resting-state functional magnetic resonance imaging (fMRI), are promising biomarkers that can be used for early assessment and prognosis of the disease and to understand its mechanisms. FDG-PET is the best suited functional marker so far, as it gives a reliable quantitative measure, but is invasive. On the other hand, non-invasive fMRI acquisitions do not provide a straightforward quantification of brain functional activity. To analyze populations solely based on resting-state fMRI, we propose an approach that leverages a metabolic prior learned from FDG-PET. More formally, our classification framework embeds population priors learned from another modality at the voxel-level, which can be seen as a regularization term in the analysis. Experimental results show that our PET-informed approach increases classification accuracy compared to pure fMRI approaches and highlights regions known to be impacted by the disease.

Overview of the proposed classification pipeline: The inputs are ROI-to- voxel connectivities computed from the rs-fMRI time-series. FDG-PET model weights are integrated as prior for the classification. Then, predictions of all ROIs are the inputs of a stacking model to predict the clinical group.

More information can be found in [44].

7.8. Inverse problems with time-frequency dictionaries and Gaussian non-white noise

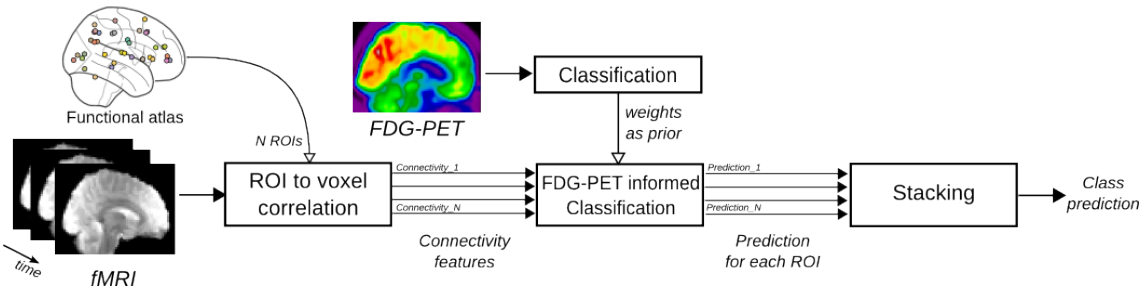


Figure 9.

Sparse regressions to solve ill-posed inverse problems have been massively investigated over the last decade. Yet, when noise is present in the model, it is almost exclusively considered as Gaussian and white. While this assumption can hold in practice it rarely holds when observations are time series as they are corrupted by auto-correlated and colored noise. In this work we study sparse regression under the assumption of non white Gaussian noise and explain how to run the inference using proximal gradient methods. We investigate an application in brain imaging: the problem of source localization using magneto- and electroencephalography (M/EEG) which allow functional brain imaging with high temporal resolution. We use a time-frequency representation of the source waveforms and a sparse regularization which promotes focal sources with smooth and transient activations. Our approach is evaluated using simulations comparing it to strategies that assume the noise is white or to simple prewhitening.

More information can be found in [30].

7.9. Variable density sampling based on physically plausible gradient waveform. Application to 3D MRI angiography

Performing k-space variable density sampling is a popular way of reducing scanning time in Magnetic Resonance Imaging (MRI). Unfortunately, given a sampling trajectory, it is not clear how to traverse it using gradient waveforms. In this paper, we actually show that existing methods can yield large traversal time if the trajectory contains high curvature areas. Therefore, we consider here a new method for gradient waveform design which is based on the projection of unrealistic initial trajectory onto the set of hardware constraints. Next, we show on realistic simulations that this algorithm allows implementing variable density trajectories resulting from the piecewise linear solution of the Travelling Salesman Problem in a reasonable time. Finally, we demonstrate the application of this approach to 2D MRI reconstruction and 3D angiography in the mouse brain.

Full k-space acquisition with an EPI sequence (a) and corresponding reference image (f). Comparison between an exact parameterization of the TSP trajectory (b) and projection from Travelling Salesman Problem trajectory onto the set of constraints (c),(d). In experiments (b,c), the number of measured locations is fixed to 9% (r = 11.2), whereas in (b,d), the time to traverse the curve is fixed to 62 ms. (e): Spiral trajectory with acquisition of the k-space center. (g-j): Reconstructed images corresponding to sampling strategies (b-e).

More information can be found in [38].

7.10. A projection method on measures sets.

We consider the problem of projecting a probability measure π on a set MN of Radon measures. The projection is defined as a solution of the following variational problem:

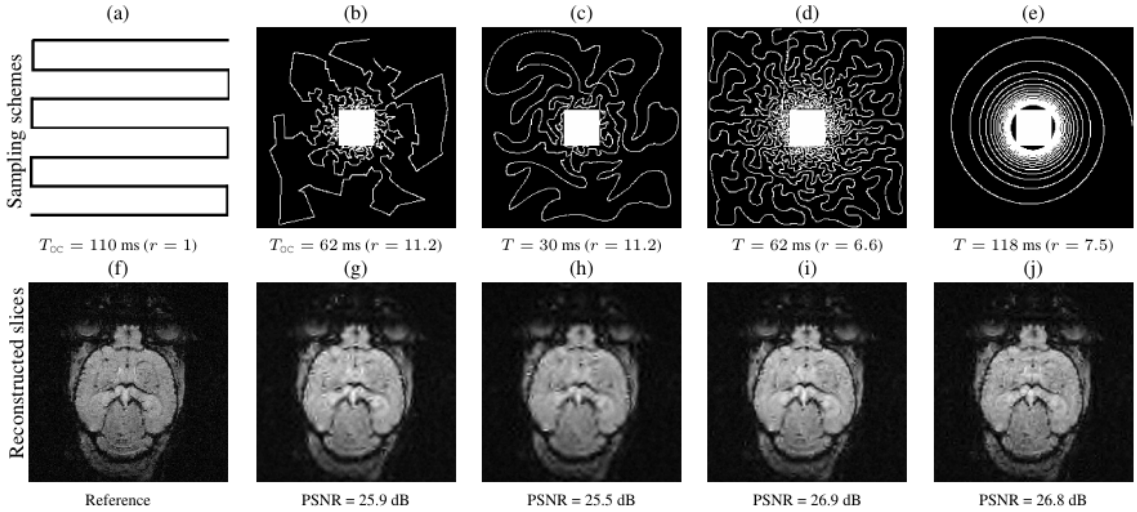


Figure 10.

$$inf_{\mu \in \mathcal{M}_N} \|h(\mu - \pi)\|_2^2,$$

where $h \in L^2(\Omega)$ is a kernel, $\Omega \subset \mathbb{R}^d$ and denotes the convolution operator. To motivate and illustrate our study, we show that this problem arises naturally in various practical image rendering problems such as stippling (representing an image with N dots) or continuous line drawing (representing an image with a continuous line). We provide a necessary and sufficient condition on the sequence $(\mathcal{M}_N)_{N \in \mathbb{N}}$ that ensures weak convergence of the projections $(\mu_N^*)_{N \in \mathcal{N}}$ to π. We then provide a numerical algorithm to solve a discretized version of the problem and show several illustrations related to computer-assisted synthesis of artistic paintings/drawings.

Projection of a lion image onto $P_N^{1,\infty}$ with $N = 8,000$. The figure depicts the resulting line with several values of the iterates of our Algorithm.

More information can be found in [55].

Figure 11.

POPIX Team

7. New Results

7.1. Model identifiability

We have discussed the question of model identifiability within the context of nonlinear mixed effects models. Although there has been extensive research in the area of fixed effects models, much less attention has been paid to random effects models. In this context we distinguish between theoretical identifiability, in which different parameter values lead to non-identical probability distributions, structural identifiability which concerns the algebraic properties of the structural model, and practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. We have explored a number of pharmacokinetic models which are known to be non-identifiable at an individual level but can become identifiable at the population level if a number of specific assumptions on the probabilistic model hold. Essentially if the probabilistic models are different, even though the structural models are non-identifiable, then they will lead to different likelihoods.

7.2. Model of tumor growth

Both molecular profiling of tumors and longitudinal tumor size data modeling are relevant strategies to predict cancer patients' response to treatment. Herein we have proposed a model of tumor growth inhibition integrating a tumor's genetic characteristics that successfully describes the time course of tumor size in patients with low-grade gliomas treated with first-line temozolomide chemotherapy. The model captures potential tumor progression under chemotherapy by accounting for the emergence of tissue resistance to treatment following prolonged exposure to temozolomide. Using information on individual tumors' genetic characteristics, in addition to early tumor size measurements, the model was able to predict the duration and magnitude of response, especially in those patients in whom repeated assessment of tumor response was obtained during the first 3 months of treatment. Combining longitudinal tumor size quantitative modeling with a tumor"s genetic characterization appears as a promising strategy to personalize treatments in patients with low-grade gliomas

7.3. Methods for PDEs based model

We have extended the methodologies previously developed for ordinary differential equations (ODE) to partial differential equations (PDE). A finite element method solver for a given family of PDEs has been developed. This solver can now be used with a prototype version of Monolix, a platform for population modeling of longitudinal data. We have implemented the well-known Lagrange finite element method in one, two and three dimensions of the space.

SISTM Project-Team

7. New Results

7.1. Time-Course Gene Set Analysis for Longitudinal Gene Expression Data

The application of TcGSA methodology has revealed the commitment of inflammatory pathways and T-cell pathway in response of DC-based vaccine.

VISAGES Project-Team

7. New Results

7.1. Image Computing: Detection, Segmentation, Registration and Analysis

7.1.1. Symmetric Block-Matching Registration for the Distortion Correction of Echo-Planar Images

Participants: Renaud Hédouin, Olivier Commowick, Elise Bannier, Christian Barillot.

We introduce a new approach to correct geometric and intensity distortion of Echo Planar Images (EPI) from images acquired with opposite phase encoding directions. A new symmetric block-matching registration algorithm has been developed for this purpose relying on new adapted transformations between blocks and a symmetric optimization scheme to ensure an opposite symmetric transformation. Our results show the ability of our algorithm to robustly recover EPI distortion while obtaining sharper results than the popular TOPUP algorithm [24], [34].

7.1.2. Quantitative analysis of T2/T2 relaxation time alteration*

Participants: Benoit Combès, Anne Kerbrat, Olivier Commowick, Christian Barillot.

T2 and T2* relaxometric data becomes a standard tool for the quantitative assessment of brain tissues and of their changes along time or after the infusion of a contrast agent. Being able to detect significant changes of T2/T2* relaxation time is an important issue. Generally, such a task is performed by comparing the variability level in the regions of interest to the variability in the normal appearance white matter. However, in the case of T2 and T2* relaxometry, this solution is highly problematic. Indeed the level of noise in the normal appearance white matter is significanlty smaller than the level of noise in more intense region (e.g. MS lesions). Our aim is to provide a Bayesian analysis of T2/T2* relaxometry estimation and alteration. More specifically, we build posterior distributions for the relaxation time and the relaxation offset by elucidating the dedicated Jeffreys priors. Then the resulting posterior distributions can be evaluated using a Monte Carlo Markov Chain algorithm. Such an analysis has three main advantages over the classical estimation procedure. First it allows in a simple way to compute many estimators of the posterior including the mode, the mean, the variance and confidence intervals. Then, it allows to include prior information. Finally, becauce one can extract confidence interval from the posterior, testing properly whether the true relaxometry time is included within a certain range of value given a confidence level is simple.

7.1.3. MRI quantitative imaging: Myelin Water Fraction (MWF) quantification in Multiple Sclerosis

Participants: Olivier Commowick, Elise Bannier, Christian Barillot.

Multi-echo T2 relaxometry is potentially a relevant imaging method for MWF quantification in the study of multiple sclerosis (MS). However, to ensure accurate estimation, a large number of echoes are still required that can drive to very long acquisitions. In practice, 32 echo times ranging from 10 ms to 320 ms and an echo spacing (ESP) of 10 ms are used1. Analysis of the decay curve of the consecutive echoes allows the estimation of the T2 spectrum. The proposed approach makes use of recent spatial regularization methods for MWF estimation from clinically compatible acquisitions (typically 11 echoes acquired within 6 minutes). The algorithms were evaluated on both synthetic and clinical data. This work was done during the internship of Lucas Soustelle [32], [29].

7.1.4. Classification of Multiple Sclerosis Lesions using Adaptive Dictionary Learning

Participants: Hrishikesh Deshpande, Pierre Maurel, Christian Barillot.

This work presents a sparse representation and an adaptive dictionary learning based method for automated classification of Multiple Sclerosis (MS) lesions in Magnetic Resonance (MR) images. Manual delineation of MS lesions is a time-consuming task, requiring neuroradiology experts to analyze huge volume of MR data. This, in addition to the high intra- and inter-observer variability necessitates the requirement of automated MS lesion classification methods. Among many image representation models and classification methods that can be used for such purpose, we investigate the use of sparse modeling. In the recent years, sparse representation has evolved as a tool in modeling data using a few basis elements of an over-complete dictionary and has found applications in many image processing tasks including classification. We propose a supervised classification approach by learning dictionaries specific to the lesions and individual healthy brain tissues, which include White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF). The size of the dictionaries learned for each class plays a major role in data representation but it is an even more crucial element in the case of competitive classification. Our approach adapts the size of the dictionary for each class, depending on the complexity of the underlying data. The algorithm is validated using 52 multi-sequence MR images acquired from 13 MS patients. The results demonstrate the effectiveness of our approach in MS lesion classification.

This work has been published in the journal of Computerized Medical Imaging and Graphics, Elsevier, 2015 [15]. Part of this work is published as a conference paper in ISBI 2015 [22].

7.1.5. Robust Detection of Multiple Sclerosis Lesions

Participants: Yogesh Karpate, Olivier Commowick, Christian Barillot.

Multiple sclerosis (MS) is a disease with heterogeneous evolution among the patients. Quantitative analysis of longitudinal Magnetic Resonance Images (MRI) provides a spatial analysis of the brain tissues which may lead to the discovery of biomarkers of disease evolution. Better understanding of the disease will lead to a better discovery of pathogenic mechanisms, allowing for patient-adapted therapeutic strategies. To characterize MS lesions, we have proposed two new approaches. The first one consists in a novel paradigm to detect white matter lesions based on a statistical framework [26]. It aims at studying the benefits of using multi-channel MRI to detect statistically significant differences between each individual MS patient and a database of control subjects. This framework consists in two components. First, intensity standardization is conducted to minimize the inter-subject intensity difference arising from variability of the acquisition process and different scanners. The intensity normalization maps parameters obtained using a robust Gaussian Mixture Model (GMM) estimation not affected by the presence of MS lesions. The second part studies the comparison of multi-channel MRI of MS patients with respect to an atlas built from the control subjects, thereby allowing us to look for differences in normal appearing white matter, in and around the lesions of each patient. Experimental results demonstrate that our technique accurately detects significant differences in lesions consequently improving the results of MS lesion detection.

Then we have presented an automatic algorithm for the detection of multiple sclerosis lesions (MSL) from multi-sequence magnetic resonance imaging (MRI) [25]. We built a probabilistic classifier that can recognize MSL as a novel class, trained only on Normal Appearing Brain Tissues (NABT). Patch based intensity information of MRI images is used to train a classifier at the voxel level. The classifier is in turn used to compute a probability characterizing the likelihood of each voxel to be a lesion. This probability is then used to identify a lesion voxel based on simple Otsu thresholding. The proposed framework was evaluated on 16 patients and our analysis reveals that our approach is well suited for MSL detection and outperforms other benchmark approaches.

7.2. Image processing on Diffusion Weighted Magnetic Resonance Imaging

7.2.1. Interpolation and Averaging of Multi-Compartment Model Images

Participants: Renaud Hédouin, Olivier Commowick, Christian Barillot.

Multi-compartment diffusion models (MCM) are increasingly used to characterize the brain white matter microstructure from diffusion MRI. We address the problem of interpolation and averaging of MCM images as a simplification problem based on spectral clustering. As a core part of the framework, we propose novel solutions for the averaging of MCM compartments. Evaluation is performed both on synthetic and clinical data, demonstrating better performance for the "covariance analytic" averaging method. We then present an MCM template of normal controls constructed using the proposed interpolation [23].

7.2.2. The DTI Challenge: Toward Standardized Evaluation of Diffusion Tensor Imaging Tractography for Neurosurgery

Participants: Olivier Commowick, Sylvain Prima.

Diffusion tensor imaging (DTI) tractography reconstruction of white matter pathways can help guide brain tumor resection. However, DTI tracts are complex mathematical objects and the validity of tractography-derived information in clinical settings has yet to be fully established. To address this issue, the DTI Challenge was initiated, an international working group of clinicians and scientists whose goal was to provide standardized evaluation of tractography methods for neurosurgery. The purpose of this empirical study was to evaluate different tractography techniques in the first DTI Challenge workshop. Eight international teams from leading institutions reconstructed the pyramidal tract in four neurosurgical cases presenting with a glioma near the motor cortex. Tractography methods included deterministic, probabilistic, filtered, and global approaches. Standardized evaluation of the tracts consisted in the qualitative review of the pyramidal pathways by a panel of neurosurgeons and DTI experts and the quantitative evaluation of the degree of agreement among methods. The evaluation of tractography reconstructions showed a great inter-algorithm variability. Although most methods found projections of the pyramidal tract from the medial portion of the motor strip, only a few algorithms could trace the lateral projections from the hand, face, and tongue area. In addition, the structure of disagreement among methods was similar across hemispheres despite the anatomical distortions caused by pathological tissues. The DTI Challenge provides a benchmark for the standardized evaluation of tractography methods on neurosurgical data. This study [18] suggests that there are still limitations to the clinical use of tractography for neurosurgical decision making.

7.2.3. Diffusion MRI abnormalities detection with orientation distribution functions: A multiple sclerosis longitudinal study

Participants: Olivier Commowick, Jean-Christophe Ferré, Gilles Edan, Christian Barillot.

We proposed a new algorithm for the voxelwise analysis of orientation distribution functions between one image and a group of reference images [13]. It relies on a generic framework for the comparison of diffusion probabilities on the sphere, sampled from the underlying models. We demonstrated that this method, combined to dimensionality reduction through a principal component analysis, allows for more robust detection of lesions on simulated data when compared to classical tensor-based analysis. We then demonstrated the efficiency of this pipeline on the longitudinal comparison of multiple sclerosis patients at an early stage of the disease: right after their first clinically isolated syndrome (CIS) and three months later. We demonstrated the predictive value of ODF-based scores for the early detection of lesions that will appear or heal.

7.3. EEG and MR Imaging

7.3.1. On the feasibility and specificity of simultaneous EEG and ASL MRI at 3T

Participants: Elise Bannier, Marsel Mano, Isabelle Corouge, Lorraine Perronnet, Christian Barillot.

Brain functional imaging can be performed using several approaches, including EEG, BOLD and ASL MRI. To date, only a few studies have addressed the issue of connecting EEG signal to ASL perfusion. ASL imaging relies on control and label RF pulses, generating alternate gradient patterns as well as higher SAR. The aim of this study was to assess ASL-EEG at 3T in terms of safety as well as EEG and MR signal quality [19].

7.3.2. Symmetrical EEG-FMRI Imaging by Sparse Regularization

Participants: Pierre Maurel, Nicolas Raillard, Saman Noorzadeh, Christian Barillot.

This work [28] considers the problem of brain imaging using simultaneously recorded electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). To this end, we introduce a linear coupling model that links the electrical EEG signal to the hemodynamic response from the blood-oxygen level dependent (BOLD) signal. Both modalities are then symmetrically integrated, to achieve a high resolution in time and space while allowing some robustness against potential decoupling of the BOLD effect. The novelty of the approach consists in expressing the joint imaging problem as a linear inverse problem, which is addressed using sparse regularization. We consider several sparsity-enforcing penalties, which naturally reflect the fact that only few areas of the brain are activated at a certain time, and allow for a fast optimization through proximal algorithms. The significance of the method and the effectiveness of the algorithms are demonstrated through numerical investigations on a spherical head model. This is a joint work with T.Oberlin and R.Gribonval.

7.4. Applications in Neuroradiology and Neurological Disorders

7.4.1. Brain perfusion gender differences using ASL in young adults

Participants: Léa Itmi, Pierre Maurel, Isabelle Corouge, Jean-Christophe Ferré, Christian Barillot.

The use of population models is becoming increasingly important in cerebral imaging, particularly using Arterial Spin Labeling perfusion imaging. Therefore, it is important to know the limits of the models before applying them, to guarantee the reliability of the results. It is now well-known that brain perfusion, in particular cerebral blood flow (CBF), changes with age, and this effect needs to be taken into account when evaluating brain perfusion images. But gender differences have not been well studied yet. It is known that female brain perfusion is, in average, higher than male brain perfusion, but only few studies have investigated whether some regional perfusion differences exist or not. This work aims to assess whether, as for the age, gender differences should be taken into account when analyzing brain perfusion images. We then focus on adult subjects and study the CBF gender differences. We compared the raw CBF means and the means after normalization, we also investigated perfusion asymetries. We used atlases for the region comparisons and the General Linear Model for the voxel level. Our results confirmed that women have a higher CBF than men, and showed that this difference can be suppressed with a normalization process, but no specific major regional difference or asymmetry was found.

7.4.2. Arterial Spin Labeling Motor Activation Presurgical Mapping for Brain Tumor Resection

Participants: Isabelle Corouge, Elise Bannier, Jean-Christophe Ferré.

Functional Arterial Spin Labeling (fASL) has demonstrated its greater specificity as a marker of neuronal activity than the reference BOLD fMRI for motor activation mapping in healthy volunteers. Motor fASL is yet to be investigated in the context of tumors, under the assumption that fASL would be less sensitive to venous contamination induced by the hemodynamics remodeling in the tumor vicinity than BOLD fMRI. As the arterial transit time may be shortened in activation areas, this preliminary study explores the ability of fASL to map the motor areas at different post-labeling delays (PLD) in healthy subjects and patient with brain tumor [21].

7.4.3. Dynamic assessment of macrophages infiltration and tissue damage in MS lesions

Participants: Anne Kerbrat, Benoit Combès, Olivier Commowick, Jean-Christophe Ferré, Elise Bannier, Christian Barillot, Gilles Edan.

Inflammation is a dynamic and complex process that could be beneficial when it supports tissue repair but also detrimental when excessive, leading to worsen tissue injury. In multiple sclerosis, it is well known from pathological and MRI studies that the prognostic between white matter lesions differed at the lesion level. Thus, 10 to 30% of T2 hyperintense lesions are seen as area of persistent hypointensity on T1-w images. These T1 hypointensity are areas of pathologically confirmed severe axonal loss. Complementary, quantitative MRI such as Diffusion imaging, magnetization transfer imaging and relaxometry can quantify and characterize

tissue changes on MRI before, during, and after the evolution of a new MRI-detected lesion. They are related to damage to myelin and axons. However, identifying in vivo the dynamic pathophysiological processes that leads to these various degree of demyelination and axonal loss in MS lesions remained challenging. In recent year, molecular and cellular imaging of the inflammatory process have been developed. Although some techniques remains at the pre-clinical level, MRI using non targeted USPIO as contrast agent can be used in MS patients. USPIO are phagocyted in periphery by macrophages and migrate to the central nervous system to characterize in vivo macrophages infiltrations within lesions. The association of cellular imaging and longitudinal quantitative MRI consist of a great opportunity to assess more specifically the overall process. In a recent study from our group, we demonstrated that infiltration of activated macrophages evidenced by USPIO enhancement, was present at the onset of MS and associated with higher local loss of tissue structure [17]. This year, we pursued this work by analyzing a longitudinal study with USPIO infusion every 3 months, associated with quantitative MRI assessment including MTI, diffusion imaging and relaxometry with the objectives of describing relationships between macrophages infiltration and quantitative MRI metrics reflecting tissue structure along time.

7.4.4. The effect of water suppression on the hepatic lipid quantification, as assessed by the LCModel, in a preclinical and clinical scenario

Participant: Elise Bannier.

This work investigates the effect of water suppression on the hepatic lipid quantification, using the LCModel. MR spectra with and without water suppression were acquired in the liver of mice at 4.7 T and patients at 3 T, and processed with the LCModel. The Cramer-Rao Lower Bound (CRLB) values of the seven lipid resonances were determined to assess the impact of water suppression on hepatic lipid quantification. A paired t test was used for comparison between the CRLBs obtained with and without water suppression. For the preclinical data, in the high (low) fat fraction subset an overall impairment in hepatic lipid quantification, i.e. an increase of CRLBs (no significant change of CRLBs) was observed in spectra acquired with water suppression. For the clinical data, there were no substantial changes in the CRLB with water suppression. Because (1) the water suppression does not overall improve the quantification of the lipid resonances and (2) the MR spectrum without water suppression is always acquired for fat fraction calculation, the optimal data-acquisition strategy for liver MRS is to acquire only the MR spectrum without water suppression. For quantification of hepatic lipid resonances, it is advantageous to perform MR spectroscopy without water suppression in a clinical and preclinical scenario (at moderate fields) [14].

7.5. Management of Information in Neuroimaging

Participants: Michael Kain, Olivier Commowick, Elise Bannier, Inès Fakhfakh, Justine Guillaumont, Florent Leray, Yao Yao, Christian Barillot.

The major topic that is addressed in this period concern the sharing of data and processing tools in neuroimaging (through the “Programme d’Investissement d’Avenir” project such as OFSEP and FLI-IAM) which led to build a suitable architecture to share images and processing tools, started from the NeuroBase project (supported by the French Ministry of Research). Our overall goal within these projects is to set up a computer infrastructure to facilitate the sharing of neuroimaging data, as well as image processing tools, in a distributed and heterogeneous environment. These consortium gathered expertise coming from several complementary domains of expertise: image processing in neuroimaging, workflows and GRID computing, ontology development and ontology-based mediation. This enables a large variety of users to diffuse, exchange or reach neuroimaging information with appropriate access means, in order to be able to retrieve information almost as easily as if the data were stored locally by means of the "cloud computing" Storage as a Service (SaaS) concept As an example, the Shanoir environment has been succesfully deployed to the Neurinfo platform were it is routinely used to manage images of the research studies. It is also currently being deployed for two large projets: OFSEP ("Observatoire Français de la Sclérose en Plaques") where up to 30000 patients will be acquired on a ten years frame, and the Image Analysis and Mangment (IAM) node of the France Life Imaging national infrastructure (FLI-IAM). Our team fulfills multiple roles in this nation-wide FLI project. Christian Barillot

is the chair of the IAM node, Olivier Commowick is participating in the working group workflow and image processing and Michael Kain is the technical manager of the node. Apart from the team members, software solutions like medInria and Shanoir are part of the final software platform.

AIRSEA Team

7. New Results

7.1. Modeling for Oceanic and Atmospheric flows

7.1.1. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Mehdi-Pierre Daou, Laurent Debreu, Florian Lemarié, Charles Pelletier, Antoine Rousseau.

7.1.1.1. Coupling heterogeneous models in hydrodynamics

The coupling of models of different kinds is gaining more and more attention, due in particular to a need for more global modeling systems encompassing different disciplines (e.g. multi-physics) and different approaches (e.g. multi-scale, nesting). In order to develop such complex systems, it is generally more pragmatic to assemble different modeling units inside a user friendly modelling software platform rather than to develop new complex global models.

In the context of hydrodynamics, global modeling systems have to couple models of different dimensions (1D, 2D or 3D) and representing different physics (Navier-Stokes, hydrostatic Navier-Stokes, shallow water...). We have been developing coupling approaches for several years, based on so-called Schwarz algorithms. Our recent contributions address the development of absorbing boundary conditions for Navier-Stokes equations [1], and of interface conditions for coupling hydrostatic and nonhydrostatic Navier-Stokes flows [2]. In the context of our partnership with with ARTELIA Group (PhD thesis of Medhi Pierre Daou), implementations of Schwarz coupling algorithms have been performed for hydrodynamics industrial codes (Mascaret, Telemac and OpenFoam), using the PALM coupling software. A first implementation has been realized in an academic test case, and a second one is presently under implementation in a much more realistic context.

7.1.1.2. Ocean-atmosphere coupling

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approximation of one [12]. For the last few years we have been actively working on the analysis of Schwarz waveform relaxation to apply this type of iterative coupling method to air-sea coupling [59], [60], [58]. In the context of the simulation of tropical cyclone, sensitivity tests to the coupling method have been carried out using ensemble simulations (through perturbations of the coupling frequency and initial conditions). We showed that the use of the Schwarz iterative coupling methods leads to a significantly reduced spread in the ensemble results (in terms of cyclone trajectory and intensity), thus suggesting that a source of error is removed w.r.t coupling methods en vogue in existing coupled models [61].

Motivated by this encouraging result, our activities over the last year can be divided into three topics

1. *Stability and consistency analysis of existing coupling methods*: in [12] 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.
2. *Study of physics-dynamics coupling*: during the PhD-thesis of Charles Pelletier (funded by Inria) the scope is on including the formulation of physical parameterizations in the theoretical analysis of the coupling. The first months of this Ph-D were dedicated to the study of the parameterization schemes to compute air-sea fluxes. A thorough sensitivity analysis showed that several parameters within existing schemes have no influence on the resulting fluxes. A simplified scheme retaining most the complexity of complicated parameterizations has thus been designed. This new scheme has also the advantage to be more adequate to conduct the mathematical analysis of the coupling.

3. *Design of a coupled single column model*: in order to focus on specific problems of ocean-atmosphere coupling, a work on simplified equation sets has been started. The aim is to implement a one-dimensional (in the vertical direction) coupled model with physical parameterizations representative of those used in realistic models. Thanks to this simplified coupled model the objective is to develop a benchmark suite for coupled models evaluation.

These three topics are addressed through strong collaborations between the applied mathematics and the climate community As an illustration, the PhD-thesis of Charles Pelletier is in collaboration with the LSCE (Laboratoire des Sciences du Climat et de l'Environnement).

Moreover a PPR (*Projet à partenariat renforcé*) called SIMBAD (SIMplified Boundary Atmospheric layer moDel for ocean modeling purposes) is funded by Mercator-Ocean for the next three years (from march 2015 to march 2018). The aim of this project in collaboration with Meteo-France, Ifremer, LMD, and LOCEAN is to derive a metamodel to force high-resolution oceanic operational models for which the use of a full atmospheric model is not possible due to a prohibitive computational cost.

7.1.1.3. Data assimilation for coupled models

In the context of operational meteorology and oceanography, forecast skills heavily rely on proper combination of model prediction and available observations via data assimilation techniques. Historically, numerical weather prediction is made separately for the ocean and the atmosphere in an uncoupled way. However, in recent years, fully coupled ocean-atmosphere models are increasingly used in operational centers to improve the reliability of seasonal forecasts and tropical cyclones predictions. For coupled problems, the use of separated data assimilation schemes in each medium is not satisfactory since the result of such assimilation process is generally inconsistent across the interface, thus leading to unacceptable artefacts. Hence, there is a strong need for adapting existing data assimilation techniques to the coupled framework. As part of our ERACLIM2 contribution, R. Pellerej started a PhD on that topic late 2014. So far, three general data assimilation algorithms, based on variational data assimilation techniques, have been developed and applied to a simple coupled problem. The dynamical equations of the considered problem are coupled using an iterative Schwarz domain decomposition method. The aim is to properly take into account the coupling in the assimilation process in order to obtain a coupled solution close to the observations while satisfying the physical conditions across the air-sea interface. Preliminary results shows significant improvement compared to the usual approach on this simple system.

7.1.2. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Laurent Debreu, Florian Lemarié.

In 2015, we worked on the stability constraints for oceanic numerical models ([13]). The idea is to carry a deep analysis of these constraints in order to propose new time stepping algorithms for ocean models. Except for vertical diffusion (and possibly the external mode and bottom drag), oceanic models usually rely on explicit time-stepping algorithms subject to Courant-Friedrichs-Lewy (CFL) stability criteria. Implicit methods could be unconditionally stable, but an algebraic system must be solved at each time step and other considerations such as accuracy and efficiency are less straightforward to achieve. Depending on the target application, the process limiting the maximum allowed time-step is generally different. In this paper, we introduce offline diagnostics to predict stability limits associated with internal gravity waves, advection, diffusion, and rotation. This suite of diagnostics is applied to a set of global, regional and coastal numerical simulations with several horizontal/vertical resolutions and different numerical models. We show that, for resolutions finer that 1/2°, models with an Eulerian vertical coordinate are generally constrained by vertical advection in a few hot spots and that numerics must be extremely robust to changes in Courant number. Based on those results, we review the stability and accuracy of existing numerical kernels in vogue in primitive equations oceanic models with a focus on advective processes and the dynamics of internal waves. We emphasize the additional value of studying the numerical kernel of oceanic models in the light of coupled space-time approaches instead of studying the time schemes independently from spatial discretizations. From this study, we suggest some guidelines for the development of temporal schemes in future generation multi-purpose oceanic models.

The increase of model resolution naturally leads to the representation of a wider energy spectrum. As a result, in recent years, the understanding of oceanic submesoscale dynamics has significantly improved. However, dissipation in submesoscale models remains dominated by numerical constraints rather than physical ones. Effective resolution is limited by the numerical dissipation range, which is a function of the model numerical filters (assuming that dispersive numerical modes are efficiently removed). In [16], we present a Baroclinic Jet test case set in a zonally reentrant channel that provides a controllable test of a model capacity at resolving submesoscale dynamics. We compare simulations from two models, ROMS and NEMO, at different mesh sizes (from 20 to 2 km). Through a spectral decomposition of kinetic energy and its budget terms, we identify the characteristics of numerical dissipation and effective resolution. It shows that numerical dissipation appears in different parts of a model, especially in spatial advection-diffusion schemes for momentum equations (KE dissipation) and tracer equations (APE dissipation) and in the time stepping algorithms. Effective resolution, defined by scale-selective dissipation, is inadequate to qualify traditional ocean models with low-order spatial and temporal filters, even at high grid resolution. High- order methods are better suited to the concept and probably unavoidable. Fourth-order filters are suited only for grid resolutions less than a few kilometers and momentum advection schemes of even higher-order may be justified. The upgrade of time stepping algorithms (from filtered Leapfrog), a cumbersome task in a model, appears critical from our results, not just as a matter of model solution quality but also of computational efficiency (extended stability range of predictor-corrector schemes). Effective resolution is also shaken by the need for non scale-selective barotropic mode filters and requires carefully addressing the issue of mode splitting errors. Possibly the most surprising result is that submesoscale energy production is largely affected by spurious diapycnal mixing (APE dissipation). This result justifies renewed efforts in reducing tracer mixing errors and poses again the question of how much vertical diffusion is at work in the real ocean.

7.1.3. Better Parameterization of the Coastline for Ocean Models

Participants: Eric Blayo, Eugene Kazantsev, Florian Lemarié, Pierre Marchand.

We aim at the development of finer approximations of lateral boundaries and boundary conditions for NEMO, by investigating and comparing analytical and optimal control approaches.

Regarding the analytical approach, we focused on a 2D shallow water formulation, and revisited the properties of the energy and enstrophy conserving schemes in the presence of a coastline. This led us to highlight a number of problems with the enstrophy conserving scheme (sensitivity to the choice of a slip or a noslip boundary condition, non conservation of the enstrophy, numerical instability). We also proposed a corrected scheme near the boundary for the continuity equation and new values for ghost points derived from the energy conservation in order for the energy conserving scheme to take into account a coastline with some inclination with regard to the numerical grid. We also investigated the viscous case, and proposed an implementation of slip and no slip boundary conditions for the viscous term in such a case of an inclined coastline.

These results are under comparison with the optimal control approach 7.3.2 realised for the Nemo model in a similar configuration.

7.2. Model reduction / multiscale algorithms

7.2.1. Intrusive sensitivity analysis, reduced models

Participants: Maëlle Nodet, Clémentine Prieur.

Another point developed in the team for sensitivity analysis is model reduction. To be more precise regarding model reduction, the aim is to reduce the number of unknown variables (to be computed by the model), using a well chosen basis. Instead of discretizing the model over a huge grid (with millions of points), the state vector of the model is projected on the subspace spanned by this basis (of a far lesser dimension). The choice of the basis is of course crucial and implies the success or failure of the reduced model. Various model reduction methods offer various choices of basis functions. A well-known method is called “proper orthogonal decomposition" or “principal component analysis". More recent and sophisticated methods also exist and may be studied, depending on the needs raised by the theoretical study. Model reduction is a natural

way to overcome difficulties due to huge computational times due to discretizations on fine grids. In [55], the authors present a reduced basis offline/online procedure for viscous Burgers initial boundary value problem, enabling efficient approximate computation of the solutions of this equation for parametrized viscosity and initial and boundary value data. This procedure comes with a fast-evaluated rigorous error bound certifying the approximation procedure. The numerical experiments in the paper show significant computational savings, as well as efficiency of the error bound.

When a metamodel is used (for example reduced basis metamodel, but also kriging, regression, ...) for estimating sensitivity indices by Monte Carlo type estimation, a twofold error appears: a sampling error and a metamodel error. Deriving confidence intervals taking into account these two sources of uncertainties is of great interest. We obtained results particularly well fitted for reduced basis metamodels [56]. In [54], the authors provide asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. These results were also adapted to problems related to more general models such as Shallow-Water equations, in the context of the control of an open channel [8].

Let us come back to the output of interest. Is it possible to get better error certification when the output is specified. A work in this sense has been accepted, dealing with goal oriented uncertainties assessment [7].

A collaboration has been started with Christophe Prieur (Gipsa-Lab) on the very challenging issue of sensitivity of a controlled system to its control parameters [8].

7.2.2. Multigrid Methods for Variational Data Assimilation.

Participants: Laurent Debreu, François-Xavier Le Dimet, Arthur Vidard.

In order to lower the computational cost of the variational data assimilation process, we investigate the use of multigrid methods to solve the associated optimal control system. On a linear advection equation, we study the impact of the regularization term on the optimal control and the impact of discretization errors on the efficiency of the coarse grid correction step. We show that even if the optimal control problem leads to the solution of an elliptic system, numerical errors introduced by the discretization can alter the success of the multigrid methods. The view of the multigrid iteration as a preconditioner for a Krylov optimization method leads to a more robust algorithm. A scale dependent weighting of the multigrid preconditioner and the usual background error covariance matrix based preconditioner is proposed and brings significant improvements. This work is summarized in ([5]).

7.3. Dealing with uncertainties

7.3.1. Sensitivity Analysis for Forecasting Ocean Models

Participants: Eric Blayo, Laurent Gilquin, Céline Helbert, François-Xavier Le Dimet, Elise Arnaud, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Laurence Viry, Federico Zertuche.

7.3.1.1. Scientific context

Forecasting geophysical systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

7.3.1.2. Estimating sensitivity indices

A first task is to develop tools for estimated sensitivity indices. In variance-based sensitivity analysis, a classical tool is the method of Sobol' [68] which allows to compute Sobol' indices using Monte Carlo integration. One of the main drawbacks of this approach is that the estimation of Sobol' indices requires the use of several samples. For example, in a d-dimensional space, the estimation of all the first-order Sobol' indices requires $d+1$ samples. Some interesting combinatorial results have been introduced to weaken this defect, in particular by Saltelli [66] and more recently by Owen [64] but the quantities they estimate still require $O(d)$ samples.

In a recent work [71] we introduce a new approach to estimate all first-order Sobol' indices by using only two samples based on replicated latin hypercubes and all second-order Sobol' indices by using only two samples based on replicated randomized orthogonal arrays. We establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, we propose to apply this new approach to a marine ecosystem model of the Ligurian sea (northwestern Mediterranean) in order to study the relative importance of its several parameters. The calibration process of this kind of chemical simulators is well-known to be quite intricate, and a rigorous and robust — i.e. valid without strong regularity assumptions — sensitivity analysis, as the method of Sobol' provides, could be of great help. The computations are performed by using CIGRI, the middleware used on the grid of the Grenoble University High Performance Computing (HPC) center. We are also applying these estimates to calibrate integrated land use transport models. As for these models, some groups of inputs are correlated, Laurent Gilquin extended the approach based on replicated designs for the estimation of grouped Sobol' indices [6].

We can now wonder what are the asymptotic properties of these new estimators, or also of more classical ones. In [54], the authors deal with asymptotic properties of the estimators. In [52], the authors establish also a multivariate central limit theorem and non asymptotic properties.

7.3.1.3. Sensitivity analysis with dependent inputs

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. For the moment, there does not exist conclusive results in that direction. Our aim is to define an analogue of Hoeffding decomposition [53] in the case where input parameters are correlated. Clémentine Prieur supervised Gaëlle Chastaing's PhD thesis on the topic (defended in September 2013) [44]. We obtained first results [45], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs. We then adapted various algorithms for the estimation of these new indices. These algorithms make the assumption that among the potential interactions, only few are significant. Two papers have been recently accepted [43], [46]. We also considered (see the paragraph 7.3.1) the estimation of groups Sobol' indices, with a procedure based on replicated designs. These indices provide information at the level of groups, and not at a finer level, but their interpretation is still rigorous.

Céline Helbert and Clémentine Prieur supervised the PhD thesis of Simon Nanty (funded by CEA Cadarache, and defended in October, 2015). The subject of the thesis is the analysis of uncertainties for numerical codes with temporal and spatio-temporal input variables, with application to safety and impact calculation studies. This study implied functional dependent inputs. A first step was the modeling of these inputs, and a paper has been submitted [63]. The whole methodology proposed during the PhD is under advanced revision [36].

7.3.1.4. Multy-fidelity modeling for risk analysis

Federico Zertuche's PhD concerns the modeling and prediction of a digital output from a computer code when multiple levels of fidelity of the code are available. A low-fidelity output can be obtained, for example on a coarse mesh. It is cheaper, but also much less accurate than a high-fidelity output obtained on a fine mesh. In this context, we propose new approaches to relieve some restrictive assumptions of existing methods ([57], [65]): a new estimation method of the classical cokriging model when designs are not nested and a nonparametric modeling of the relationship between low-fidelity and high-fidelity levels. The PhD takes place in the REDICE consortium and in close link with industry. The first part of the thesis was also dedicated to the development of a case study in fluid mechanics with CEA in the context of the study of a nuclear reactor.

The second part of the thesis was dedicated to the development of a new sequential approach based on a course to fine wavelets algorithm. Federico Zertuche presented his work at the annual meeting of the GDR Mascot Num in 2014 [72].

7.3.1.5. Data assimilation and second order sensitivity analysis

Basically, in the deterministic approach, a sensitivity analysis is the evaluation of a functional depending on the state of the system and of parameters. Therefore it is natural to introduce an adjoint model. In the framework of variational data assimilation the link between all the ingredients (observations, parameters and other inputs of the model is done through the optimality system (O.S.), therefore a sensitivity will be estimated by deriving the O.S. leading to a second order adjoint. This is done in the paper [15] in which a full second order analysis is carried out on a model of the Black Sea.

This methodology has been applied to

- **Oil Spill**. These last years have known several disasters produced by wrecking of ships and drifting platforms with severe consequences on the physical and biological environments. In order to minimize the impact of these oil spills its necessary to predict the evolution of oil spot. Some basic models are available and some satellites provide images on the evolution of oil spots. Clearly this topic is a combination of the two previous one: data assimilation for pollution and assimilation of images. A theoretical framework has been developed with Dr. Tran Thu Ha (iMech).
- **Data Assimilation in Supercavitation (with iMech)**. Some self propelled submarine devices can reach a high speed thanks to phenomenon of supercavitation: an air bubble is created on the nose of the device and reduces drag forces. Some models of supercavitation already exist but are working on two applications of variational methods to supercavitation:
  - Parameter identification : the models have some parameters that can not be directly measured. From observations we retrieve the unknown parameters using a classical formalism of inverse problems.
  - Shape Optimization. The question is to determine an optimum design of the shape of the engine in order to reach a maximum speed.

7.3.2. Optimal Control of Boundary Conditions

Participants: Christine Kazantsev, Eugene Kazantsev.

A variational data assimilation technique is applied to the identification of the optimal boundary conditions for a simplified configuration of the NEMO model. A rectangular box model placed in mid-latitudes, and subject to the classical single or double gyre wind forcing, is studied. The model grid can be rotated on a desired angle around the center of the rectangle in order to simulate the boundary approximated by a staircase-like coastlines. The solution of the model on the grid aligned with the box borders was used as a reference solution and as artificial observational data. It is shown in [9], [10] that optimal boundary has a rather complicated geometry which is neither a staircase, nor a straight line. The boundary conditions found in the data assimilation procedure bring the solution toward the reference solution allowing to correct the influence of the rotated grid (see fig. 1).

Adjoint models, necessary to variational data assimilation, have been produced by the TAPENADE software, developed by the SCIPORT team. This software is shown to be able to produce the adjoint code that can be used in data assimilation after a memory usage optimization.

7.3.3. Non-Parametric Estimation for Kinetic Diffusions

Participants: Clémentine Prieur, Jose Raphael Leon Ramos.

This research is the subject of a collaboration with Venezuela and is partly funded by an ECOS Nord project.

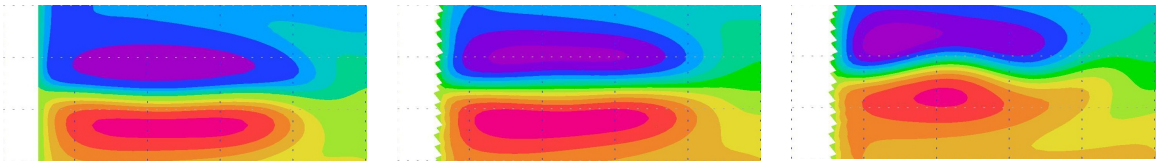


Figure 1. Sea surface elevation: reference solution on the aligned grid (left), solutions on the 30° *rotated grid with optimal (center) and conventioal (right) boundary conditions.*

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 [40] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [41] (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 [41], in collaboration with J.R. Leon (Caracas, Venezuela) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [42], also recently accepted.20

Note that Professor Jose R. Leon (Caracas, Venezuela) is now funded by an international Inria Chair and will spend one year in our team, allowing to collaborate further on parameter estimation.

7.3.4. Multivariate Risk Indicators

Participants: Clémentine Prieur, Patricia Tencaliec.

Studying risks in a spatio-temporal context is a very broad field of research and one that lies at the heart of current concerns at a number of levels (hydrological risk, nuclear risk, financial risk etc.). Stochastic tools for risk analysis must be able to provide a means of determining both the intensity and probability of occurrence of damaging events such as e.g. extreme floods, earthquakes or avalanches. It is important to be able to develop effective methodologies to prevent natural hazards, including e.g. the construction of barrages.

Different risk measures have been proposed in the one-dimensional framework . The most classical ones are the return level (equivalent to the Value at Risk in finance), or the mean excess function (equivalent to the Conditional Tail Expectation CTE). However, most of the time there are multiple risk factors, whose dependence structure has to be taken into account when designing suitable risk estimators. Relatively recent regulation (such as Basel II for banks or Solvency II for insurance) has been a strong driver for the development of realistic spatio-temporal dependence models, as well as for the development of multivariate risk measurements that effectively account for these dependencies.

We refer to [47] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [67] proposed a generalization of the concept of return period in dimension greater than or equal to two. Michele et al. proposed in a recent study [48] to take into account the duration and not only the intensity of an event for designing what they call the dynamic return period. However, few studies address the issues of statistical inference in the multivariate context. In [49], [51], we proposed non parametric estimators of a multivariate extension of the CTE. As might be expected, the properties of these estimators deteriorate when considering extreme risk levels. In collaboration with Elena Di Bernardino (CNAM, Paris), Clémentine Prieur is working on the extrapolation of the above results to extreme risk levels.

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [50]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervises the PhD thesis of Patricia Tencaliec. We are working on risk assessment, concerning flood data for the Durance drainage basin (France). The PhD thesis started in October 2013. A first paper on data reconstruction has been accepted [18]. It was a necessary step as the initial series contained many missing data.

7.4. Assimilation of Images

Participants: François-Xavier Le Dimet, Maëlle Nodet, Arthur Vidard, Nelson Feyeux, Vincent Chabot, Nicolas Papadakis.

7.4.1. Direct assimilation of image sequences

At the present time the observation of Earth from space is done by more than thirty satellites. These platforms provide two kinds of observational information:

- Eulerian information as radiance measurements: the radiative properties of the earth and its fluid envelops. These data can be plugged into numerical models by solving some inverse problems.
- Lagrangian information: the movement of fronts and vortices give information on the dynamics of the fluid. Presently this information is scarcely used in meteorology by following small cumulus clouds and using them as Lagrangian tracers, but the selection of these clouds must be done by hand and the altitude of the selected clouds must be known. This is done by using the temperature of the top of the cloud.

MOISE was the leader of the ANR ADDISA project dedicated to the assimilation of images, and is a member of its follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France) that ended in 2013.

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [69], which was recently extended [17]. Thanks to the nonlinear brightness assumption, we proposed an algorithm to estimate the motion between two images, based on the minimization of a nonlinear cost function. We proved its efficiency and robustness on simulated and experimental geophysical flows [38]. As part of the ANR project GeoFluids, we are investigating new ways to define distance between a couple of images. One idea is to compare the gradient of the images rather than the actual value of the pixels. This leads to promising results. Another idea, currently under investigation, consists in comparing main structures within each image. This can be done using, for example, a wavelet representation of images. Both approaches have been compared, in particular their relative merits in dealing with observation errors, in a paper published early 2015 [4] and presented in several international conferences [21], [28].

In recent developments [11] we have also used "Level Sets" methods to describe the evolution of the images. The advantage of this approach is that it permits, thanks to the level sets function, to consider the images as a state variable of the problem. We have derived an Optimality System including the level sets of the images.

7.4.2. Optimal transport for image assimilation

Within the optimal transport project TOMMI funded by the ANR white program (started mid 2011), a new optimization scheme based on proximal splitting method has been proposed to solve the dynamic optimal transport problem. We investigate the use of optimal transport based distances for data assimilation. The study is still under investigation in the framework of N. Feyeux's PhD, but preliminary encouraging results have already been presented in [20] and an article is in preparation on this topic.

7.5. Tracking of Mesoscale Convective Systems

Participant: Clémentine Prieur.

We are interested in the tracking of mesoscale convective systems.A particular region of interest is West Africa. Data and hydrological expertise is provided by T. Vischel and T. Lebel (LTHE, Grenoble).

A first approach involves adapting the multiple hypothesis tracking (MHT) model originally designed by the NCAR (National Centre for Atmospheric Research) for tracking storms [70] to the data for West Africa. With A. Makris (working on a post-doctoral position), we proposed a Bayesian approach [62], which consists in considering that the state at time t is composed on one hand by the events (birth, death, splitting, merging) and on the other hand by the targets' attributes (positions, velocities, sizes, ...). The model decomposes the state into two sub-states: the events and the targets positions/attributes. The events are updated first and are conditioned to the previous targets sub-state. Then given the new events the target substate is updated. A simulation study allowed to verify that this approach improves the frequentist approach by Storlie et al. (2009). It has been tested on simulations [62] and investigated in the specific context of real data on West Africa [35]. Using PHD (probability hypothesis density) filters adapted to our problem, generalizing recent developments in particle filtering for spatio-temporal branching processes (e.g. [39]) could be an interesting alternative to explore. The idea of a dynamic, stochastic tracking model should then provide the base for generating rainfall scenarios over a relatively vast area of West Africa in order to identify the main sources of variability in the monsoon phenomenon.

7.6. Land Use and Transport Models Calibration

Participants: Thomas Capelle, Laurent Gilquin, Clémentine Prieur, Arthur Vidard, Peter Sturm, Elise Arnaud.

Given the complexity of modern urban areas, designing sustainable policies calls for more than sheer expert knowledge. This is especially true of transport or land use policies, because of the strong interplay between the land use and the transportation systems. Land use and transport integrated (LUTI) modelling offers invaluable analysis tools for planners working on transportation and urban projects. Yet, very few local authorities in charge of planning make use of these strategic models. The explanation lies first in the difficulty to calibrate these models, second in the lack of confidence in their results, which itself stems from the absence of any well-defined validation procedure. Our expertise in such matters will probably be valuable for improving the reliability of these models. To that purpose we participated to the building up of the ANR project CITiES led by the STEEP EPI. This project started early 2013 and two PhD about sensitivity analysis and calibration were launched late 2013. This work led to conference papers [24], [23] and a two published journal paper [3], [6]

ANGE Project-Team

7. New Results

7.1. Modelling of complex flows

7.1.1. Non-hydrostatic models

Participant: Martin Parisot.

A new shallow water type model involving non-hydrostatic effects is derived in [37]. Under the assumption that the horizontal velocity is close to its vertical mean value, the model enables to recover the energy from the Euler system before integration. Link with the non-hydrostatic published in [18] is identified. Compared to the aforementioned models, the new system consists of more equations (6). However, the numerical strategy presented in the paper does not induce extra computational time.

7.1.2. Seismic activities: energy radiated by elastic waves

Participants: Anne Mangeney, Jacques Sainte-Marie.

Estimating the energy loss in elastic waves during an impact is an important problem in seismology and in industry. Three complementary methods to estimate the elastic energy radiated by bead impacts on thin plates and thick blocks from the generated vibration are proposed in [30]. The first two methods are based on the direct wave front and are shown to be equivalent. The third method makes use of the diffuse regime. These methods are shown to be relevant to establish the energy budget of an impact. The radiated elastic energy estimated with the presented methods is quantitatively validated by Hertz's model of elastic impact.

7.1.3. Layer-averaged Euler and Navier-Stokes systems

Participants: Marie-Odile Bristeau, Bernard Di Martino, Cindy Guichard, Jacques Sainte-Marie.

In [25] we propose a strategy to approximate incompressible free surface Euler and Navier-Stokes models. The main advantage of the proposed models is that the water depth is a dynamical variable of the system and hence the model is formulated over a fixed domain.

The proposed strategy extends previous works approximating the Euler and Navier-Stokes systems using a multilayer description. Here, the needed closure relations are obtained using an energy-based optimality criterion instead of an asymptotic expansion. Moreover, the layer-averaged description is successfully applied to the Navier-Stokes system with a general form of the Cauchy stress tensor.

7.2. Applications to marine energies

7.2.1. Partially free surface flow

Participants: Martin Parisot, Fabien Wahl.

In view of taking into account interactions with buoys, a new formulation of the shallow water model is derived with a constraint corresponding to a static roof. A relaxation approach is considered to adapt the standard numerical schemes. A particular attention is paid to the energy law whether it be for the original model with constraint or the relaxed version.

7.2.2. Swell energy

Participants: Sebastian Reyes-Riffo, Julien Salomon.

The internship consisted in designing an optimisation algorithm to determine advantageous topographies in view of producing energy from swell. This approach corresponds to the coupling between a shallow water type model with iterative updates of the topography. Stability of the numerical scheme is a critical point and requires the tuning of parameters.

7.3. Analysis of models in Fluid Mechanics

7.3.1. Weak solutions of multilayer models

Participants: Bernard Di Martino, Ethem Nayir, Yohan Penel.

Proving the existence of global weak solutions is a difficult problem for Navier-Stokes type equations, particularly in case of a degenerate viscosity (viscosity term can vanish if density or thickness goes to zero). In some recent works, Vasseur and Yu [46], have proved this existence for 2D shallow water equations. For the multilayer model, a collaboration with Boris Haspot (Univ. Paris-Dauphine) lead to stability results for the system with a focus on the difficulty to construct a sequence of approximate solutions that conserve all a priori estimates.

7.3.2. Strong solutions of multilayer models

Participants: Emmanuel Audusse, Ethem Nayir, Yohan Penel.

The existence and uniqueness of strong solutions of the multilayer model proposed in [41] was previously proven in the case of boundary conditions. We extended this result to an unbounded domain for short times, overcoming the issue of integrability often barely evoked in similar investigations. Current works deal with the long time existence by a continuation process which requires a particular care of the short time solution at the end of its existence interval.

7.3.3. Hyperbolic problems under constraints

Participant: Nicolas Seguin.

In [21], we study a family of linear hyperbolic systems whose solution must satisfy a constraint (e.g. a simplified model of river flows taking risk of flooding into account). We analyse relaxed models based on a penalisation. This theoretical approach could be used to derive numerical methods.

7.3.4. Entropy-satisfying finite volume schemes

Participant: Nicolas Seguin.

In [44], we carry out an analysis of 1st-order entropy-satisfying finite volume schemes for hyperbolic systems. More precisely, we investigate the numerical dissipation on unstructured meshes under relevant stability conditions. This results in a minimal convergence order towards smooth solutions.

7.3.5. Global existence for Green-Naghdi type equations

Participant: Dena Kazerani.

In [31], we consider the Cauchy problem for the Green-Naghdi equations with viscosity, for small initial data. It is well-known that adding a second order diffusion term to a hyperbolic system leads to the existence of global smooth solutions, as soon as the hyperbolic system is symmetrizable and the so-called Kawashima-Shizuta condition is satisfied. In a previous work, we have proved that the Green-Naghdi equations can be written in a symmetric form, using the associated Hamiltonian. This system being dispersive, in the sense that it involves third order derivatives, the symmetric form is based on symmetric differential operators. We use this structure for an appropriate change of variable to prove that adding viscosity effects through a second order term leads to global existence of smooth solutions, for small data. We also deduce that constant solutions are asymptotically stable.

7.4. Numerical methods for free-surface flows

7.4.1. Godunov schemes for the low Froude regime

Participants: Emmanuel Audusse, Do Minh Hieu, Yohan Penel.

We investigated in [29] the behaviour of collocated Godunov type finite volume schemes when applied to the 1d linear wave equation with Coriolis force in collaboration with S. Dellacherie and P. Omnes (CEA). Accuracy for short time and stability were proven for different versions of the classical Godunov schemes, including some schemes already proposed in the literature (Bouchut *et al*, [42]). Next step will be to include linear advection and then to study the fully non linear shallow water model. Then results will be extended to 2d problems for which geometrical constraints should be taken into account.

7.4.2. Numerical method for non-hydrostatic models

Participants: Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, Jacques Sainte-Marie.

In [1], a numerical method based on a prediction-correction scheme in one dimension has been developed and compared to experimental data and analytical solutions. The issue is then to extend the method in higher dimensions. We propose a variational framework for the resolution of a non-hydrostatic Saint-Venant type model with bottom topography. This model is a shallow water type approximation of the free surface incompressible Euler system and slightly differs from the Green-Naghdi model. The resolution of the incompressibility constraint leads to an elliptic problem involving the non-hydrostatic part of the pressure. This step uses a variational formulation of a shallow water version of the incompressibility condition. Several numerical experiments are performed to confirm the relevance of our approach. This work is exposed in [18].

7.4.3. Uncertainties with the topography

Participants: Emmanuel Audusse, Nicole Goutal, Philippe Ung.

We propose to study the uncertainty related to the Saint-Venant system. A perturbation is introduced in the bottom topography such that the topography deviation is characterized by two parameters: its amplitude and its smoothness. In particular, we extend the work previously done with periodic boundary conditions and suggest a treatment of the physical ones. In doing so, we are interested in the influence of the topography deviation on the hydraulic quantities, and in particular, we numerically exhibit a relationship between the spatial correlations of the topography and the water height. Furthermore, we complete the study by a comparison of the outputs between the two flow regimes – fluvial and torrential.

7.4.4. Coupled Stokes-Exner model

Participant: Nora Aïssiouene.

In the framework of the 2015 CEMRACS session (Coupling Multi-Physics Models involving Fluids), we explored an approach to model the sediment transport. In [17], we consider a coupling between the Exner equation and the Stokes system to model sediments in geophysical flow phenomena. We focus on a model without free surface and used some numerical tests to evaluate the relevance of the method. The fluid structure interaction theory and methods have been applied on the coupled system and the objective is to test the proposed method which can be extend to a free surface model. The library Feel++ and the high computing performance embedded have been used to test the solution method. Therefore, the goal of this project is to understand the impact of the sediment transport on the flow using Navier-Stokes with a free surface system coupled with the standard Exner equation. This work has been done in collaboration with Tarik Amtout, Matthieu Brachet, Emmanuel Frenod, Romain Hild, Christophe Prud'homme, Antoine Rousseau and Stéphanie Salmon.

7.5. Software developments and assessments

7.5.1. Improvements in the FRESHKISS3D code

Participants: Marie-Odile Bristeau, David Froger, Raouf Hamouda, Jacques Sainte-Marie.

Several tasks have been achieved in the FRESHKISS3D software:

- The parallelisation of FRESHKISS3D with MPI is achieved for the Eulerian description and the explicit time scheme.
- The paddle wheel vertical effect is now taken into account.
- Vertical and time dependent flow rates can be customised.
- Unit tests have been improved and functional tests have been added.
- Software dependencies are packaged in SED-Paris repository.
- Online documentation is being written.
- A prototype of the software implemented in Cython is under discussion.
- Code executing time's loop is being refactored into multiple classes.
- Various improvements (build system, continuous integrations, coding rules) have been provided.

CASTOR Project-Team

5. New Results

5.1. Plasma boundary reconstruction

Participants: Jacques Blum, Cédric Boulbe, Blaise Faugeras.

A new fast and stable algorithm has been developped for the reconstruction of the plasma boundary from discrete magnetic measurements taken at several locations surrounding the vacuum vessel. The resolution of this inverse problem takes two steps. In the first one we transform the set of measurements into Cauchy conditions on a fixed contour Γ_O close to the measurement points. This is done by least square fitting a truncated series of toroidal harmonic functions to the measurements. The second step consists in solving a Cauchy problem for the elliptic equation satisfied by the flux in the vacuum and for the overdetermined boundary conditions on Γ_O previously obtained with the help of toroidal harmonics. It is reformulated as an optimal control problem on a fixed annular domain of external boundary Γ_O and fictitious inner boundary Γ_I. A regularized Kohn-Vogelius cost function depending on the value of the flux on Γ_I and measuring the discrepancy between the solution to the equation satisfied by the flux obtained using Dirichlet conditions on Γ_O and the one obtained using Neumann conditions is minimized. The method presented here has led to the development of a software, called VacTH-KV, which enables plasma boundary reconstruction in any Tokamak (see [14]).

5.2. Free boundary - Transport Solver - Controller coupling

Participants: Cédric Boulbe, Blaise Faugeras, Jean François Artaud [IRFM CEA Cadarache], Vincent Basiuk [IRFM CEA Cadarache], Emiliano Fable [Max-Planck-Institut für Plasmaphysik, Garching], Philippe Huyn [IRFM CEA Cadarache], Eric Nardon [IRFM CEA Cadarache], Jakub Urban [IPP, Academy of Sciences of the Czech Republic, Prague].

Last year, a first version of the workflow coupling a free boundary equilibrium code, the European transport solver ETS and a plasma shape and position controller had been developped. In 2015, this new tool has been tested and improved. An experiment realized on the Tokamak TCV and called "yoyo" shot has been successfully simulated. This work has been realised in the framework of the Eurofusion Work Package: Code Development for integrated modelling project.

5.3. A finite element method with overlapping meshes for free-boundary toroidal plasma equilibria in realistic geometry

Participants: Holger Heumann, Francesca Rappetti.

Existing finite element implementations for the computation of free-boundary toroidal plasma equilibria approximate the flux function by piecewise polynomial, globally continuous functions. Recent numerical results for the self-consistent coupling of equilibrium and resistive diffusion in the spirit of Grad-Hogan suggest the necessity of higher regularity. Enforcing continuity of the gradient in finite elements methods on triangular meshes, leads to a drastic increase in the number of unknowns, since the degree of the polynomial approximation needs to be increased beyond four. Therefore existing implementations for the fixed boundary problem resort to (curvilinear) quadrilateral meshes and approximation spaces based on cubic Hermite splines. Fine substructures in the realistic geometry of a tokamak, such as air-gaps, passive structures and the vacuum vessel prevent the use of quadrilateral meshes for the whole computational domain, as it would be necessary for the free-boundary problem.

In this work we propose a finite element method that employs two meshes, one of quadrilaterals in the vacuum domain and one of triangles outside, which *overlap* in a narrow region around the vacuum domain. This approach gives the flexibility to achieve easily and at low cost higher order regularity for the approximation of the flux function in the domain covered by the plasma, while preserving accurate meshing of the geometric details exterior to the vacuum. The continuity of the numerical solution in the region of overlap is weakly enforced by relying on the mortar projection. A publication is in preparation.

5.4. Inverse transient plasma equilibrium problem

Participants: Holger Heumann, Jacques Blum.

The inverse transient plasma equilibrium problem aims at precomputing the trajectories of externally applied voltages in the poloidal field coils of a tokamak. A basic implementation of this problem in 2011/2012 in CEDRES++ during Holger Heumann's PostDoc at Inria, provided first insight into the capabilities and also difficulties of this approach. Application engineers are highly interested in this application, but realistic cases will require more evolved numerical methods to reduce the computational time and memory requirements. In 2014 we implemented the inverse transient plasma equilibrium problem in FEEQS.M to facilitate our search for better algorithms. In 2015 we started working on realistic test cases for the upcoming WEST tokamak. In order to make such problems accessible by the current version of our code, we had to split the time interval of interest into 5 subintervals, on which we solve 5 inverse problems. Only by the initial condition the problem on a subinterval is connected to its predecessor. Next we faced some serious convergence problem of the optimisation algorithms for some of these problems. These led us to do extensive benchmark runs with different optimisation algorithms and implementation, including both Gradient and SQP-type methods, either with handcoded or MATLAB-native implementations. As a result we envisage for 2016 the incorporation of the SQP implementation of Jean-Charles Gilbert, which seems to be perfectly adapted to optimal control problems such as ours. Another improvement was achieved in reducing the actual number of free control parameters and to replace piecewise linear control trajectories with high order polynomials.

5.5. High order for the axisymmetric magnetohydrodynamic equilibrium problem

Participants: Holger Heumann, Lukas Drescher [TU Berlin], Kersten Schmidt [TU Berlin].

We implemented a higher order finite element method (FEM) for solving numerically axisymmetric magneto-hydrodynamic (MHD) equilibrium problems. The focus is on high accuracy and the capabilities of high-order FEM implementations for faster calculations. High order FEM for elliptic problems, such as the considered MHD equilibrium problem, is well established and understood. This work uses the hp-FEM software CONCEPTS developed at ETH Zürich/TU Berlin. Further, we developed a novel method for computing accurately the so-called flux surface averages, that are important in transient MHD calculations. This new method circumvents the expensive and very technical computation of line-integrals and fits seamlessly into the high order finite element method. A publication is in preparation.

5.6. Towards automated magnetic divertor design for optimal heat exhaust

Participants: Holger Heumann, Maarten Blommaert [FZ, Jülich (Germany)], Martine Baelmans [KU Leuven (Belgium)], Nicolas R. Gauger [TU, Kaiserslautern (Germany)], Detlev Reiter [FZ, Jülich (Germany)].

Avoiding excessive structure heat loads in future fusion tokamaks is regarded as one of the greatest design challenges. In this joint effort, we aim at developing a tool to study how the severe divertor heat loads can be mitigated by reconfiguring the magnetic confinement. For this purpose, the free boundary equilibrium code FEEQS.M was integrated with a plasma edge transport code to work in an automated fashion. A practical and efficient adjoint based sensitivity calculation was proposed to evaluate the sensitivities of the integrated code. The sensitivity calculation was applied to a realistic test case and compared with finite difference sensitivity calculations.

The integration of the free boundary equilibrium solver FEEQS.M allowed to assess the validity of a previous simplified model introduced by M. Bloomaert. It was found that the absence of plasma response currents significantly limits the accuracy of this simplified model.

The novel procedure was applied to obtain first results for the new WEST (Tungsten Environment in Steady-state Tokamak) divertor currently under construction in the Tore Supra tokamak at CEA. The sensitivities and the related divertor optimization paths are strongly affected by the extension of the magnetic model (see [24]).

5.7. Bohm boundary conditions

Participants: Richard Pasqueti, Sebastian Minjeaud.

Focusing on a minimal model proposed in the late 2000's by the IRFM (Cadarache), an algorithm has been proposed to enforce at the plates the inequality $M \geq 1$, where M is the parallel Mach number. The algorithm is implemented in the FBGKI code, but still requires improvements to enhance the robustness of the numerical method (see [18]).

5.8. High order approximation of dispersive equations and conservation of invariants

Participants: Richard Pasqueti, Sebastian Minjeaud.

Focusing on the Korteweg-de Vries (KdV) equation, algorithms have been proposed to handle high order derivative terms (third order for KdV) with C^0 elements and to preserve invariants (mass and momentum for KdV) through the time-scheme (see [33]).

5.9. Taylor-Galerkin stabilized Finite Element

Participants: José Costa, Boniface Nkonga.

The theoretical part of Taylor-Galerkin/Variational multi-scales (TG/VMS) strategy applied to MHD and reduced MHD modeling has been achieved last year. The final method amounts to adding in the finite element formulation, a self-adjoint operator associated to the most critical hyperbolic component of the system to be solved. The design of the critical contours and the identification of associated waves to be stabilized is problem dependent and related to the Jacobian matrix. This year we have continued the investigations for the design and improvement of the stabilization started in 2015. For application to plasma configurations with X-point, we have designed a numerical strategy that preserved the initial equilibrium without flows. The Bohm boundary condition on open flux walls has been formulated and is now under validation.

5.10. Toward full MHD numerical modeling with $C1$ finite element.

Participants: José Costa, Giorgio Giorgiani, Hervé Guillard, Boniface Nkonga.

In this context the single fluid full MHD model is considered and the divergence free constraint on the magnetic field is achieved by introduction of a vector potential . The use of the vector potential has the additional advantage that the toroidal component is the magnetic flux of the Grad-Shafranov equilibrium. However, using the vector potential as variable introduces higher order derivatives in the system and classical C0 finite elements cannot be directly applied. This is why our finite element strategies uses shape/test functions whose derivatives have global continuity in space (smooth finite elements). The global approach uses cross product shape/test functions between poloidal(2D) and the Toroidal(1D). In the 2D poloidal plane, discretization uses either quadrangular or triangular elements. In order to derive efficient strategies for the full MHD in the vector potential formulation, the Gauge condition on the vector potential and the boundary conditions have been enforced by penalization. For the Gauge condition it gives rise to element contributions but also boundary integrals that should be computed on curved surfaces that sometime fitted the magnetic surfaces. Equations are formulated in semi-conservative form such as to apply integration by parts. Therefore, boundary conditions can be viewed as evolution of fluxes or variables. Integral formulation on the boundary is very useful for higher order finite elements and also easier for the treatment of corners. Indeed in this context the boundary conditions are edge/surfaces oriented and boundary corners are driven by the neighborhood edge penalization. This strategy is the one that will be used for future developments.

2D Quadrangular Cubic Bezier Finite Elements:
This finite element has been used for a while for reduced MHD models in the software Jorek. Reduced MHD uses the projection of the momentum equation in a space orthogonal to the equilibrium. When full MHD models are used, the momentum equation needs to be projected in the equilibrium space and this projection should be consistent with the Grad-Shafranov equilibrium that is used to compute the initial state. This has been achieved by a proper computation of the JxB contribution in the momentum equation, taking into account the poloidal variation of the toroidal component of the magnetic field. After a detailed analysis, we have performed this year some implementations and numerical validations. An Inria report is under preparation.
2D Triangular Powell-Sabin Finite Elements:
In order to avoid some mesh singularities when using quadrangular meshes for complex geometries and flux surfaces shapes, triangular elements are a possible option. It is not so easy to derive smooth finite elements on triangles with reduced number of degrees of freedom (ddl). The Bell reduced-quintic finite elements we have considered in the previous years have too many unknowns (6 per vertex). Powell-Sabin splines are piecewise quadratic polynomials with a global C1-continuity and 3 unknowns per vertex, they have a local support, they form a convex partition of unity, they are stable, and they have a geometrically intuitive interpretation involving control triangles. Construction of the Powell-Sabin splines needs some geometrical tools that have been developed: Minimum area enclosing triangle of a set of control points (sequential and parallel). This construction is applied to each vertex of the triangular mesh and used to derive the local shape/test functions. These Powell-Sabin splines have been used successfully in the area of computer aided geometric design for the modeling and fitting of surfaces. We have used the Powell-Sabin (PS) splines for the approximation of elliptic partial differential equations (including Grad-Shafranov) in a rectangular domain. In this context, the optimal rate of convergence (order 3) has been recovered. This year, validations have been performed for hyperbolic 2D Euler equations with VMS stabilization. The context of the 3D toroidal geometries has been considered and implemented. Preliminary validations are satisfactory. An Inria report is also under preparation.

5.11. Genuinely multidimensional Riemann Solver

Participants: Jeaniffer Vides, Boniface Nkonga.

Multidimensional Riemann solvers were pioneered by Abgrall. Abgrall, Maire, Nkonga, Després and Loubere have extensively developed them especially as node-solvers for Lagrangian hydrodynamics. Another strain of work comes from explorations by Wendroff and Balsara who took a space-time approach. In this work, the resolved state is obtained via space-time integration over a wave model, just as was done by Wendroff and Balsara. However, an algebraic approach is used for the development of the fluxes. It is, therefore, shown that the multidimensional fluxes can be obtained by application of jump conditions at the boundaries of the wave model. The problem is of course over determined with the result that the shock jump conditions are only satisfied approximately in a least squares sense. Even so, this work gives us new perspective on multidimensional Riemann solvers. The literal satisfaction of the shock jump conditions (up to least squares approximation) makes it easier to understand multidimensional Riemann solvers as a natural extension of the one-dimensional Riemann solvers. Contributions have also been made on the development of a minimalist wave model, which might help in reducing dissipation. Further innovations are reported on the assembling of fluxes based on the structure of the wave model, and those innovations are potentially useful. For MHD the CT approach consists of constraining the transport of magnetic field so that the divergence is always kept zero. The method relies on exploiting the dualism between the flux components and the electric field. Since the electric field is needed at the edges of the mesh, the multidimensional Riemann solver can also provide the electric field. By running an extensive set of simulations, it is shown that the multidimensional Riemann solver is robust and can be used to obtain divergence-free formulations for MHD that perform well on several stringent calculations. The work performed this year was to improve this strategy by enriching of substructures the description of the strongly interaction of waves. These improvements were done in collaboration with the invited professor D. Balsara. This work has resulted in an article to be published in the Journal of Computational Physics in 2016.

5.12. Multi scales approximations of "Shallow water" flows.

Participants: Jeaniffer Vides, Boniface Nkonga, Sergey Gavrilyuk, Kseniya Ivanova.

The terminology "Shallow water" is used to characterize thin flows on curved surfaces. It is customary for this type of flows; to use the incompressible Navier-Stokes equations to asymptotically derive reduced models for the evolution of the depth integrated speed and the thickness of the flow. Reduced models are mainly hyperbolic and finite volume method are often used for their numerical approximation. Approximation strategies are generally structured as follow:

- Construction of a global coordinate system associated with an assumed analytical surface.
- Reduction of the model relatively to the global coordinate system
- Approximation of the surface by a finite number of elements.
- Approximation of the reduced model using the discrete surface.

In the context of real applications, it is presumptuous to expect an analytical formulation of the surface. From the data provided by observation satellites, we can usually extract a discrete description of the surfaces that drives thin flow. Therefore, it is more practical to use the discrete description as the starting point of the resolution strategy. This is the angle of approach that we have considered. We locally define two mesh scales: the element scale and the cell scale. The discrete mapping and the reduced model are defined at the element scale and the average values that evolve in time are defined at the cell scale. First applications have been successfully performed. Our efforts have been extended to include relevant physics at each scale, including sheared flows. We have used a multi-dimensional formulation. An Inria report is under preparation.

5.13. Asymptotic theory of reduced MHD models

Participant: Hervé Guillard.

In the study of fusion plasma, one of the fundamental model used for stability studies is the magnetohydrodynamic (MHD) model. Many theoretical and numerical works in this field use specific approximations of this model known as *reduced* MHD models. The derivation of these reduced MHD models has been formulated as a special instance of the theory of singular limit of hyperbolic system of partial differential equations with a large operator. This formulation allows to use the general results of this theory and to prove rigorously that reduced MHD models are valid approximations of the full MHD equations [29]. In particular, it is proven that the solutions of the full MHD system converge to the solutions of an appropriate reduced model. These results substantiate the intuitive physical idea that in the presence of a strong magnetic field, motion in the plane perpendicular to the plasma is nearly incompressible.

5.14. Finite volume approximations for fusion plasma

Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

The MHD model used for plasma studies in tokamak is very often based on the magnetic vector potential form of the equations where the vector potential satisfies $\nabla \times \mathbf{A} = \mathbf{B}$ with $\mathbf{B}$ the magnetic field and only a small number of numerical models use the conservative formulation based on $\mathbf{B}$. One of the shortcomings of this latter formulation is the necessity to enforce numerically the divergence free constraint on the magnetic field that can be difficult to achieve and/or computationally costly. Another difficulty is that the equilibrium solution of the MHD equation given by the Grad-Shafranov equation is not an exact solution of the discrete equation.
We have begun to investigate the use of the $\mathbf{B}$ formulation for tokamak studies. The divergence free constraint is taken into account by a projection at each time step on a rotated gradient field. This step ensures a strict respect of the divergence free constraint while being extremely cheap since the scalar field is simply advected by the flow. Preliminary numerical experiments show that this approach can have some interest. The design of a well-balanced solver will be the next step of these studies.

CLIME Project-Team

6. New Results

6.1. Simulation, observation and state estimation for analysis and forecast

The objective of Clime is the merging of simulation and observations, with data assimilation methods, for state estimation in environmental applications. However, this aim previously requires, as seen in some of the next subsection, to collect the observations and carry out the simulations.

6.1.1. Assimilation of drifter data in the East Mediterranean Sea

Participants: Julien Brajard, Milad Fakhri [CNRS, Lebanon], Daniel Hayes [Oceanography Centre, Cyprus], Leila Issa [Lebanese American University, Lebanon], Laurent Mortier [LOCEAN], Pierre-Marie Poulain [Oceanography Institute of Trieste, Italy].

Surface velocity fields of the ocean in the Eastern Levantine Mediterranean are estimated by blending altimetry and surface drifters data. The method is based on a variational assimilation approach for which the velocity is corrected by matching real drifters positions with those predicted by a simple advection model, while taking into account the wind effect. The velocity correction is done in a time-continuous fashion by assimilating at once a whole trajectory of drifters using a sliding time window. A divergence-free regularization term is added to the cost function mnimized during the assimilation process in order to constrain the velocity field. First results show that with few drifters, the method improves the estimation of the surface velocity: an eddy between the Lebanese coast and Cyprus is better assessed and the values of velocities along the Lebanese coast are more accurate.

6.1.2. Traffic simulation

Participants: Vivien Mallet, Vincent Aguiléra [CEREMA], Ruiwei Chen [CEREA].

The ANR project ESTIMAIR aims at propagating uncertainties in the complete simulation chain of air quality at urban scale. A key step in the chain lies in traffic assignment and the computation of the corresponding emissions. We take part to the simulation of traffic in the streets of Clermont-Ferrand metropolitan area, with the dynamic traffic assignment model LADTA. The simulations are evaluated against observations from loop counters and also against the simulations of the reference static model VISUM.

From the traffic assignment, the emissions are computed for nitrogen dioxide and particulate matter, using COPERT IV formulae. Preliminary work shows large uncertainties in the emissions due to the fleet composition.

6.1.3. Observation of noise pollution

Participants: Vivien Mallet, Raphaël Ventura, Valérie Issarny [MiMove], Pierre-Guillaume Raverdy [Ambientic], Fadwa Rebhi [MiMove].

Exposure to noise pollution is highly variable in space. As a consequence, it is very difficult to determine individual exposure using only numerical simulations of noise levels. Together with the MiMove Inria project-team, we take part to the SoundCity project that aims at collecting noise observations from smartphones and better evaluating the individual exposure. We assist MiMove in the development of an Android application that automatically senses noise along the day and collects the data (when the user agrees) for the improvement of simulated noise maps. Clime especially contributes to the calibration of the application. Comparisons between the measurements of smartphones and a sound meter allow us to estimate the bias of the main smartphones available on the market.

The SoundCity application was launched in July 2015 with Bernard Jomier, deputy mayor responsible for health, disability, and relations with Paris public hospital system, during a press conference organized by Paris City. The application received a positive coverage in the media, so that the application gained about 2500 users. About one million observations are collected every four days and ongoing work tries to process these data to correct Paris noise maps.

6.1.4. Evaluation of fire models

Participants: Jérémy Lefort, Vivien Mallet, Jean-Baptiste Filippi [CNRS].

In the field of forest fires risk management, important challenges exist in terms of people and goods preservation. Answering to strong needs from different actors (firefighters, foresters), researchers focus their efforts to develop operational decision support system tools that may forecast wildfire behavior. This requires the evaluation of model performance.

We carry out the evaluation of several fire propagation models based on over 500 real fires. We use the data as they would be available in operational conditions, so as to avoid any tuning that would be incompatible with real-time forecasting. The study shows significant performance difference between the models, despite the poor data quality.

6.2. Image assimilation

Sequences of images, such as satellite acquisitions, display structures evolving in time. This information is recognized of major interest by forecasters (meteorologists, oceanographers, etc.) in order to improve the information provided by numerical models. However, the satellite images are mostly assimilated in geophysical models on a point-wise basis, discarding the space-time coherence visualized by the evolution of structures such as clouds. Assimilating in an optimal way image data is of major interest and this issue should be considered in two ways:

- from the model's viewpoint, the location of structures on the observations is used to control the state vector.
- from the image's viewpoint, a model of the dynamics and structures is built from the observations.

6.2.1. Model error and motion estimation

Participants: Isabelle Herlin, Dominique Béréziat [UPMC].

Data assimilation technics are used to retrieve motion from image sequences. These methods require a model of the underlying dynamics, displayed by the evolution of image data. In order to quantify the approximation linked to the chosen dynamic model, an error term is included in the evolution equation of motion and a weak formulation of 4D-Var data assimilation is designed. The cost function to be minimized depends simultaneously on the initial motion field, at the beginning of the studied temporal window, and on the error value at each time step. The result allows to assess the model error and analyze its impact on motion estimation. The approach is used to estimate geophysical forces (gravity, Coriolis, diffusion) from images in order to better assess the surface dynamics and forecast the displacement of structures like oilspill.

6.2.2. Tracking of structures from an image sequence

Participants: Isabelle Herlin, Yann Lepoittevin, Dominique Béréziat [UPMC].

The research concerns an approach to estimate velocity on an image sequence and simultaneously segment and track a given structure. It relies on the underlying dynamics' equations of the studied physical system. A data assimilation method is designed to solve evolution equations of image brightness, those of motion's dynamics. The method is for instance applied on meteorological satellite data, in order to track tropical clouds on image sequences and estimate their motion, as seen on Fig. 2 .

Data assimilation is performed either with a 4D-Var variational approach or with a Kalman ensemble method. In the last case, the initial ensemble is obtained from a set of optical flow methods of the literature with various parameters values.

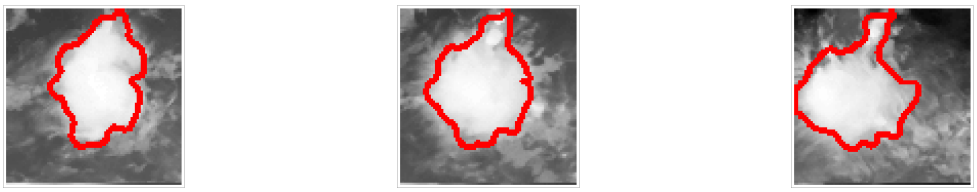


Figure 2. Tracking a tropical cloud. Frames 3, 9, 18 of the sequence.

Various ways for representing the structures are studied and compared.

- For the variational approach, we consider: 1- a distance map modeling the tracked structures, which is added to the state vector, 2- anisotropic regularization terms, which are added to the cost function minimized during the assimilation process, 3- covariances between pixels, which are included in the background error covariance matrix.
- For the filtering approach, we focus either on domain decomposition or on explicit localization, which are both related to the displayed structures.

6.2.3. Applying POD on a model output dabase for defining a reduced motion model

Participants: Isabelle Herlin, Etienne Huot.

Dimension reduction may be obtained by determining a small size reduced basis computed by Proper Orthogonal Decomposition (POD) of a motion fields database and applying the Galerkin projection. This database is constructed for characterizing accurately the surface circulation of the studied area, so that linear combinations of the basis elements obtained by POD accurately describe the motion function observed on satellite image sequences. The database includes the geostrophic motion fields obtained from Sea Level Anomaly reanalysis maps that are available from the MyOcean European project website (http://marine.copernicus.eu/). Fig. 3 displays such SLA maps and the associated motion fields.

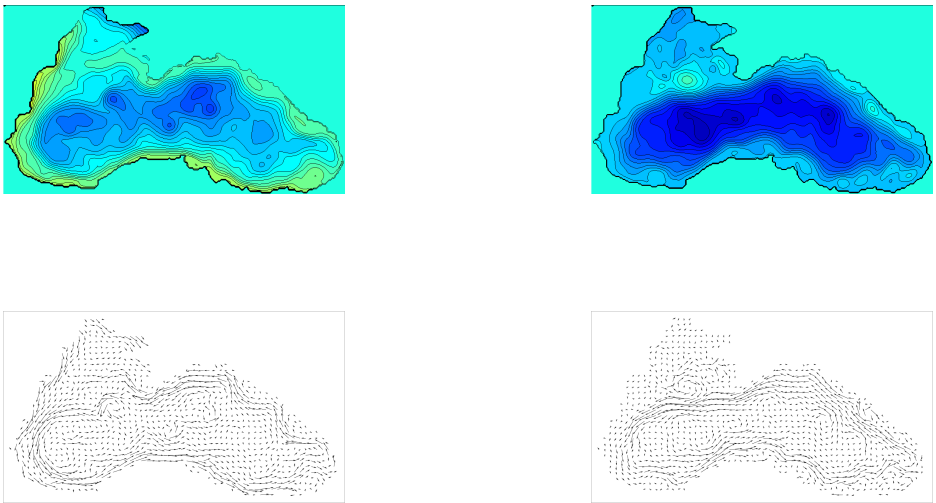


Figure 3. Top: reanalysis of SLA. Bottom: geostrophic motion.

Image assimilation with the POD reduced model allows estimating motion as displayed on Fig. 4 .

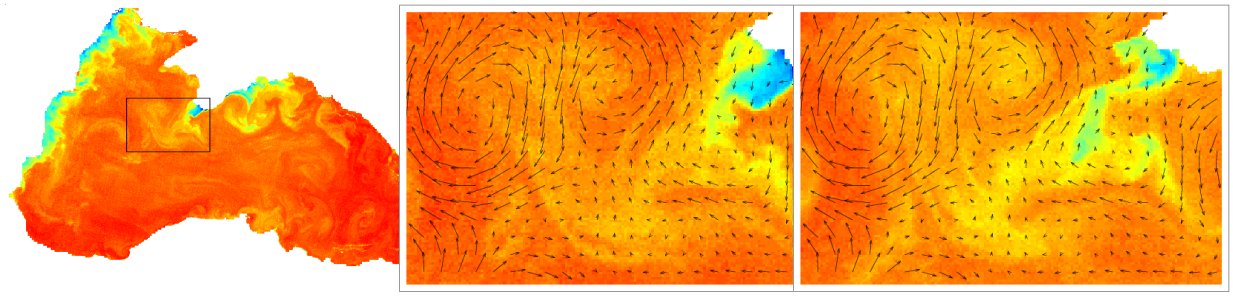


Figure 4. Zoom on a region of interest and motion estimation superposed on two consecutive images.

6.2.4. Rain nowcasting from radar image acquisitions

Participants: Isabelle Herlin, Yann Lepoittevin.

This research concerns the design of an operational method for rainfall nowcasting that aims at mitigating flash floods. The nowcasting method is composed of two main components:

- a data assimilation method, based on radar images, estimates the state of the atmosphere: this is the estimation phase.
- a forecast method uses this estimation to extrapolate the state of the atmosphere in the future: this is the forecast phase.

The method is transfered to the industrial company Weather Measures.

Current research concerns the use of object components in the state vector in order to get an improved motion estimation and a better localization of endangered regions. Assimilation of pluviometers measures in the nowcasting system is also investigated.

6.3. Uncertainty quantification and risk assessment

The uncertainty quantification of environmental models raises a number of problems due to:

- the dimension of the inputs, which can easily be 10^5-10^8 at every time step;
- the dimension of the state vector, which is usually 10^5-10^7;
- the high computational cost required when integrating the model in time.

While uncertainty quantification is a very active field in general, its implementation and development for geosciences requires specific approaches that are investigated in Clime. The project-team tries to determine the best strategies for the generation of ensembles of simulations. In particular, this requires addressing the generation of large multimodel ensembles and the issue of dimension reduction and cost reduction. The dimension reduction consists in projecting the inputs and the state vector to low-dimensional subspaces. The cost reduction is carried out by emulation, i.e., the replacement of costly components with fast surrogates.

6.3.1. Application of sequential aggregation to meteorology

Participants: Paul Baudin, Vivien Mallet, Gilles Stoltz [CNRS].

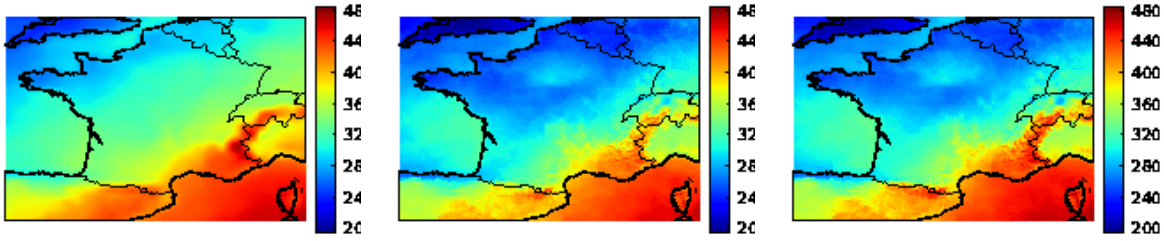
Nowadays, it is standard procedure to generate an ensemble of simulations for a meteorological forecast. Usually, meteorological centers produce a single forecast, out of the ensemble forecasts, computing the ensemble mean (where every model receives an equal weight). It is however possible to apply aggregation methods. When new observations are available, the meteorological centers also compute analyses. Therefore, we can apply the ensemble forecast of analyses, which consists in weighting the ensemble of forecasts to better forecast the forthcoming analyses. Before any forecast, the weights are updated with past observations and past forecasts. The performance of the aggregated forecast is guaranteed, in the long run, to perform at least as well as any linear combination of the forecasts with constant weights.

Ensembles of forecasts for mean sea level pressure, from the THORPEX Interactive Grand Global Ensemble, are aggregated with a forecast error decreased by 18% compared to the best individual forecast. The approach is also proved to be efficient for wind speed. The contribution of the ensembles (from different meteorological centers) to the performance increase are evaluated.

6.3.2. Sequential aggregation with uncertainty quantification and application to photovoltaics production

Participants: Paul Baudin, Vivien Mallet, Jean Thorey, Christophe Chaussin [EDF R&D], Gilles Stoltz [CNRS].

We study the aggregation of ensembles of solar radiations and photovoltaic productions. The aggregated forecasts show a 20% error decrease compared to the individual forecasts. They are also able to retrieve finer spatial patterns than the ones found in the individual forecasts (see Figure 5).



(a) (b) (c)

Figure 5. Yearly average of the map of downward shortwave solar radiation in $\mathrm{W m^{-2}}$*, for an ensemble mean (a), for our aggregated forecasts (b) and observed (c).*

An important issue is the estimation of the uncertainties associated with the aggregated forecasts. We devise a new approach to predict a probability density function or a cumulative distribution function instead of a single aggregated forecast. In practice, the aggregation procedure aims at forecasting the cumulative distribution function of the observations which is simply a Heaviside function centered at the observed value. Our forecast is the weighted empirical cumulative distribution function based on the ensemble of forecasts. The method guarantees that, in the long run, the forecast cumulative distribution function has a Continuous Ranked Probability Score (CRPS) at least as good as the best weighted empirical cumulative function with weights constant in time.

The CRPS is a classical score to evaluate the probabilistic forecasts. However, applying the CRPS on weighted empirical distribution functions (derived from the weighted ensemble) introduces a bias because of which minimizing the CRPS does not produce the optimal weights. Thus, we propose an unbiased version of the CRPS which relies on clusters of members and is strictly proper.

6.3.3. Sensitivity analysis in the dispersion of radionuclides

Participants: Sylvain Girard, Vivien Mallet, Irène Korsakissok [IRSN].

We carry out a sensitivity analysis of the dispersion of radionuclides during Fukushima disaster. We considered the dispersion at regional scale, with the Eulerian transport model Polair3D from Polyphemus. Simulations of the atmospheric dispersion of radionuclides involve large uncertainties originating from the limited knowledge of meteorological input data, composition, amount and timing of emissions and some model parameters. We studied the relative influence of each uncertain input on several outputs. In practice, we used the variance-based sensitivity analysis method of Sobol. This method requires a large number of model evaluations which are not achievable directly due to the high computational cost of the model. To circumvent this issue, we built a mathematical approximation of the model using Gaussian process emulation.

In previous studies, the uncertainties in the meteorological forecasts were crudely modeled by homogeneous and constant perturbations on the fields. Hence, we started investigating the use of ensembles of meteorological forecasts instead of just one base meterological forecast. Including such ensembles allows to better represent the directions along which meteorological uncertainties should lie.

6.3.4. Fire risk assessment

Participants: Jérémy Lefort, Vivien Mallet, Jean-Baptiste Filippi [CNRS].

During days with extreme weather conditions, every wildland fire must be fought within minutes of its occurrence. This means that sufficient firefighting force is available at the right place and at the right time. In practice, firefighters wait at different critical locations, so that they can act quickly. For efficient preventive positioning of the firefighters, forecasting the risks of ignition of large fires is essential. This requires to predict where a fire may start, to estimate its potential size, to evaluate fighting scenarios and to anticipate which urban or protected areas may be under threat.

We designed a surrogate propagation model based on Gaussian process emulation of the model ForeFire. This surrogate model is fast enough to be run all over a region with high fire risk, e.g., Corsica. It can even be used for Monte Carlo simulations, with perturbations in the meteorological conditions and vegetation state, over Corsica. It is then possible to generate a risk map that identifies all the locations where a new fire can grow large.

6.3.5. Ensemble variational data assimilation

Participants: Julien Brajard, Isabelle Herlin, Marc Bocquet [CEREA], Jérôme Sirven [LOCEAN], Olivier Talagrand [LMD, ENS], Sylvie Thiria [LOCEAN].

The general objective of ensemble data assimilation is to produce an ensemble of analysis from observations and a numerical model which is representative of the uncertainty of the system. In a bayesian framework, the ensemble represents a sampling of the state vector probability distribution conditioned to the available knowledge of the system, denoted the a-posteriori probability distribution.

Ensemble variational data assimilation (EnsVar) consists in producing such an ensemble, by perturbating N times the observations according to their error law, and run a standard variationnal assimilation for each perturbation. An ensemble of N members is then produced. In the case of linear models, there is a theoretical guarantee that this ensemble is a sampling of the a-posteriori probability. But there is no theoretical result in the non-linear case.

Numerical experiments using non-linear numerical models suggest that the conclusion reached for linear models still stands for non-linear toy models.

The objective of this work is to study the ability of EnsVar to produce "good" ensemble (i.e. that sampled the a posteriori probablility) on a more realistic model: a shallow-water model. Some statistical properties of the ensemble are presented, and the sensitivity to the main features of the assimilation system (number, distribution of observations, size of the assimilation window, ...) are also studied.

COFFEE Project-Team (section vide)

FLUMINANCE Project-Team

6. New Results

6.1. Fluid motion estimation

6.1.1. Stochastic uncertainty models for motion estimation

Participants: Etienne Mémin, Abed Malti.

In this study we have proposed a stochastic formulation of the brightness consistency used principally in motion estimation problems. In this formalization the image luminance is modeled as a continuous function transported by a flow known only up to some uncertainties. Stochastic calculus then enables to built conservation principles which take into account the motion uncertainties. These uncertainties defined either from isotropic or anisotropic models can be estimated jointly to the motion estimates. Such a formulation, besides providing estimates of the velocity field and of its associated uncertainties, allows us to naturally define a linear multiresolution scale-space framework. The corresponding estimator, implemented within a local least squares approach, has shown to improve significantly the results of the corresponding deterministic estimator (Lucas and Kanade estimator). This fast local motion estimator provides results that are of the same order of accuracy than state-of-the-art dense fluid flow motion estimator for particle images. The uncertainties estimated supply a useful piece of information in the context of data assimilation. This ability has been exploited to define multiscale incremental data assimilation filtering schemes. The development of an efficient GPU based version of this estimator has been investigated through the Inria ADT project FLUMILAB

6.1.2. 3D flows reconstruction from image data

Participants: Kai Berger, Cédric Herzet, Abed Malti.

Our work focuses on the design of new tools for the estimation of 3D turbulent flow motion in the experimental setup of Tomo-PIV. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms.

This year, we keep on our investigation on the problem of efficient volume reconstruction. Our work takes place within the context of some modern optimization techniques. First, we focussed our attention on the family of proximal and splitting methods and showed that the standard techniques commonly adopted in the TomoPIV literature can be seen as particular cases of such methodologies. Recasting standard methodologies in a more general framework allowed us to propose extensions of the latter: i) we showed that the parcimony characterizing the sought volume can be accounted for without increasing the complexity of the algorithms (e.g., by including simple thresholding operations); ii) we emphasized that the speed of convergence of the standard reconstruction algorithms can be improved by using Nesterov's acceleration schemes; iii) we also proposed a totally novel way of reconstructing the volume by using the so-called "alternating direction of multipliers method" (ADMM) . This work has led to the publication of two contributions at the international conference on particle image velocimetry (PIV) in 2015.

On top of this work, we also focussed on another crucial step of the volume reconstruction problem, namely the pruning of the model. The pruning task consists in identifying some positions in the volume of interest which cannot contains any particle. Removing this position from the problem can then potentially allow for a dramatic dimensionality reduction. This year, we provide a methodological answer to this problem through the prism of the so-called "screening" techniques which have been proposed in the community of machine learning. Our work has led to the submission of one contribution to the international conference on acoustics, speech and signal processing.

6.1.3. Sparse-representation algorithms

Participant: Cédric Herzet.

The paradigm of sparse representations is a rather new concept which turns out to be central in many domains of signal processing. In particular, in the field of fluid motion estimation, sparse representation appears to be potentially useful at several levels: i) it provides a relevant model for the characterization of the velocity field in some scenarios; ii) it plays a crucial role in the recovery of volumes of particles in the 3D Tomo-PIV problem.

Unfortunately, the standard sparse representation problem is known to be NP hard. Therefore, heuristic procedures have to be devised to access to the solution of this problem. Among the popular methods available in the literature, one can mention orthogonal matching pursuit (OMP), orthogonal least squares (OLS) and the family of procedures based on the minimization of ℓ_p norms. In order to assess and improve the performance of these algorithms, theoretical works have been undertaken in order to understand under which conditions these procedures can succeed in recovering the "true" sparse vector.

This year, we contributed to this research axis by deriving conditions of success for the algorithms mentioned above when the amplitudes of the nonzero coefficients in the sparse vector obey some decay. In a TomoPIV context, this decay corresponds to the fact that not all the particles in the fluid diffuse the same quantity of light (notably because of illumination or radius variation). In particular, we show that the standard coherence-based guarantees for OMP/OLS can be relaxed by an amount which depends on the decay of the nonzero coefficients. Our work have led to the acceptation of a paper in the journal IEEE Transactions on Information Theory.

6.2. Tracking, Data assimilation and model-data coupling

6.2.1. Sequential smoothing for fluid motion

Participants: Anne Cuzol, Etienne Mémin.

In parallel to the construction of stochastic filtering techniques for fluid motions, we have proposed a new sequential smoothing method within a Monte-Carlo framework. This smoothing aims at reducing the temporal discontinuities induced by the sequential assimilation of discrete time data into continuous time dynamical models. The time step between observations can indeed be long in environmental applications for instance, and much longer than the time step used to discretize the model equations. While the filtering aims at estimating the state of the system at observations times in an optimal way, the objective of the smoothing is to improve the estimation of the hidden state between observation times. The method is based on a Monte-Carlo approximation of the filtering and smoothing distributions, and relies on a simulation technique of conditional diffusions. The proposed smoother can be applied to general non linear and multidimensional models. It has been applied to a turbulent flow in a high-dimensional context, in order to smooth the filtering results obtained from a particle filter with a proposal density built from an Ensemble Kalman procedure. This conditional simulation framework can also be used for filtering problem with low measurement noise. This has been explored through a collaboration with Jean-Louis Marchand (ENS Bretagne) in the context of vorticity tracking from image data.

6.2.2. Stochastic fluid flow dynamics under uncertainty

Participants: Etienne Mémin, Valentin Resseguier.

In this research axis we aim at devising Eulerian expressions for the description of fluid flow evolution laws under uncertainties. Such an uncertainty is modeled through the introduction of a random term that allows taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also uncertainties attached to non-uniform grid discretization. This model is mainly based on a stochastic version of the Reynolds transport theorem. Within this framework various simple expressions of the drift component can be exhibited for different models of the random field carrying the uncertainties we have on the flow. We aim at using such a formalization within image-based data assimilation framework and to derive appropriate stochastic versions of geophysical flow dynamical modeling. This formalization has been published in the journal Geophysical

and Astrophysical Fluid Dynamics [9]. Numerical simulation on divergence free wavelets basis of 3D viscous Taylor-Green vortex and Crow instability have been performed within a collaboration with Souleymane Kadri-Harouna. Besides, we explore in the context of Valentin Resseguier's PhD the extension of such framework to oceanic models and to satellite image data assimilation. This PhD thesis takes place within a fruitful collaboration with Bertrand Chapron (CERSAT/IFREMER). This year we have more deeply explored several uncertainty representations of classical geophysical models for ocean and atmosphere. This study have led to very promising stochastic representation for the Quasi Geostophic approximation (QG) with noises of different energy.

6.2.3. Free surface flows reconstruction and tracking

Participants: Dominique Heitz, Etienne Mémin.

We investigated the combined use of a Kinect depth sensor and of a stochastic data assimilation method to recover free-surface flows. More generally, we proposed a particle filter method to reconstruct the complete state of free-surface flows from a sequence of depth images only. The data assimilation scheme introduced accounts for model and observations errors. We evaluated the developed approach on two numerical test cases: a collapse of a water column as a toy-example and a flow in an suddenly expanding flume as a more realistic flow. The robustness of the method to simulated depth data quality and also to initial conditions was considered. We illustrated the interest of using two observations instead of one observation into the correction step. Then, the performance of the Kinect sensor to capture temporal sequences of depth observations was investigated. Finally, the efficiency of the algorithm was qualified for a wave in a real rectangular flat bottom tank. It was shown that for basic initial conditions, the particle filter rapidly and remarkably reconstructed velocity and height of the free surface flow based on noisy measures of the elevation

6.2.4. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: Pranav Chandramouli, Dominique Heitz, Etienne Mémin, Cordelia Robinson.

In this axis of work we are exploring the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach will be experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we wish also to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models or in longer-term goal to learn empirical subgrid models directly from image data. An accurate modeling of this term is essential for Large Eddies Simulation as it models all the non resolved motion scales and their interactions with the large scales.

We have pursued the first investigations on a 4DVar assimilation technique, integrating PIV data and Direct Numerical Simulation (DNS), to reconstruct two-dimensional turbulent flows. The problem we are dealing with consists in recovering a flow obeying Navier-Stokes equations, given some noisy and possibly incomplete PIV measurements of the flow. By modifying the initial and inflow conditions of the system, the proposed method reconstructs the flow on the basis of a DNS model and noisy measurements. The technique has been evaluated in the wake of a circular cylinder. It denoises the measurements and increases the spatiotemporal resolution of PIV time series. These results have been recently published in the Journal of Computational Physics [6]. Along the same line of studies the 3D case is ongoing. The goal consists here to reconstruct a 3D flow from a set of simultaneous time resolved 2D images of planar sections of the 3D volume. This work has been mainly conducted within the PhD of Cordelia Robinson. The development of the variational assimilation code has been initiated within a collaboration with A. Gronskis, S. Laizé (lecturer, Imperial College, UK) and Eric Lamballais (institut P' Poitiers). A High Reynolds number simulation of the wake behind a cylinder has been recently performed within this collaboration. The 4DVar assimilation technique based on the numerical code Incompact3D is now implemented. We are currently trying to reconstruct a 3D turbulent flow from dual

plane velocity observations. The control of subgrid parameterizations will be the main objective of the PhD of Pranav Chandramouli that is just starting.

6.2.5. Ensemble variational data assimilation of large scale fluid flow dynamics with uncertainty

Participant: Etienne Mémin.

This study is focused on the coupling of a large scale representation of the flow dynamics built from the location uncertainty principle with image data of finer resolution. The velocity field at large scales is described as a regular smooth component whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have assessed the performance of an ensemble variational assimilation technique with direct image data observation. Preliminary encouraging results have been obtained for simulation under uncertainty of 1D and 2D shallow water models.

6.2.6. Reduced-order models for flows representation from image data

Participants: Cédric Herzet, Etienne Mémin, Valentin Resseguier.

During the PhD thesis of Valentin Ressguier we proposed a new decomposition of the fluid velocity in terms of a large-scale continuous component with respect to time and a small-scale non continuous random component. Within this general framework, an uncertainty based representation of the Reynolds transport theorem and Navier-Stokes equations can be derived, based on physical conservation laws. This physically relevant stochastic model has been applied in the context of the POD-Galerkin method. The pertinence of this reduced order model has been successfully assessed on several wake flows. This study has been published in two conference papers and one journal article.

On the other hand, we also investigated the problem of reduced-model construction from partial observations. In this line of search, our contribution was twofold. We first proposed a Bayesian framework for the construction of reduced-order models from image data. Our framework enables to account for any prior information on the system to reduce and takes the uncertainties on the parameters of the model into account. Interestingly, the proposed approach reduces to some well-known model-reduction techniques when the observations are not partial (i.e., the observation operator can be inverted). Second, we provided a theoretical analysis of our methodology in a simplified context (namely, the observations are supposed to be noiseless linear combinations of the state of the system). This result provides worst-case guarantees on the reconstruction performance which can be achieved by a reduced model built from the data. These contributions have been accepted for presentation in two international conferences in 2016.

6.3. Analysis and modeling of turbulent flows

6.3.1. Turbulence similarity theory for the modeling of Ocean Atmosphere interface

Participants: Roger Lewandowski, Etienne Mémin, Benoit Pinier.

The Ocean Atmosphere interface plays a major role in climate dynamics. This interaction takes place in a thin turbulent layer. To date no sastifying universal models for the coupling of atmospheric and oceanic models exists. In practice this coupling is realized through empirically derived interaction bulks. In this study, corresponding to the PhD thesis of Benoit Pinier, we aim at exploring similarity theory to identify universal mean profile of velocity and temperature within the mixture layer. The goal of this work consists in exhibiting eddy viscosity models within the primitive equations. We will also explore the links between those eddy viscocity models and the subgrid tensor derived from the uncertainty framework studied in the Fluminance group. In that prospect, we have started to study the impact of the introduction of a random modeling of the friction velocity on the classical wall law expression.

6.3.2. Hot-wire anemometry at low velocities

Participant: Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. Based on a market study the french patent application has been extended this year to a Patent Cooperation Treaty (PCT) application.

6.3.3. Numerical and experimental image and flow database

Participants: Pranav Chandramouli, Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The first challenge was to enlarge a database mainly based on two-dimensional flows, with three-dimensional turbulent flows. Synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with time resolved Particle Image Velocimetry measurements in wake and mixing layers flows. This database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images. The second challenge was to carried out orthogonal dual plane time resolved stereoscopic PIV measurements in turbulent flows. The diagnostic employed two orthogonal and synchronized stereoscopic PIV measurements to provide the three velocity components in planes perpendicular and parallel to the streamwise flow direction. These temporally resolved planar slices observations will be used in 4DVar assimilation technique, integrating Direct Numerical Simulation (DNS) and Large Eddies Simulation (LES), to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Pranav Chandramouli. The third challenge was to carried out a time resolved tomoPIV experiments in a turbulent wake flow. These temporally resolved volumic observations will be used to assess the algorithms developped in the PhD of Ioana Barbu and in the postdoc of Kai Berger. Then this data will be used in 4DVar assimilation technique to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Cordelia Robinson.

6.4. Visual servoing approach for fluid flow control

6.4.1. Closed-loop control of a spatially developing shear layer

Participant: Christophe Collewet.

This study is led within a strong collaboration with Diemer Ando-Ondo and Johan Carlier of the Acta team (Irstea Rennes). It 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. Closed loop control is achieved to maintain the shear-layer in a desired state of interest for industrial applications, and thus to reject upstream perturbations. The industrial and scientific contexts advocates first for the use of image sensor to measure the flow velocity fields and second for applying the control on the upstream boundary condition. The optimal control was performed using a linear control law designed from a reduced linearized state space model of the Navier-Stokes equations. A steady desired state was first considered leading to a linear time-invariant system. The resulting feedback control law was validated on a powerful and realistic numerical Navier-Stokes 3D solver, which will be useful to anticipate the control of the shear layer in a dedicated wind tunnel. Two conference papers on this work have been submitted to the "16th European Control Conference" and "8th AIAA Flow Control Conference". We are now considering the case of an unsteady desired state to control the large roller vortices developing in the shear layer and that are the main contributor to entrainment and mixing processes.

LEMON Team

7. New Results

7.1. Hydrodynamics of the Tunquen lagoon, Chile

Participant: Antoine Rousseau.

In this internship co-advised with Céline Acary-Robert (Inria Chile), Loïc Dagnas developed a numerical hydrodynamic model for a specific lagoon of the chilean costline. This kind of lagoon is characterized by an intermittent connection to the sea and a regular fresh water input coming from the Andean mountains. The hydrodynamic model consists in a two-dimensional shallow water model, including tracer equations for the time evolution of temperature and salinity. The hydrodynamic circulation of the lagoon has been simulated taking into account various external forcings such as water exchanges with the atmosphere, wind effects and external pumping.

7.2. Upscaled modeling of Vaccares lake in Camargue

Participants: Carole Delenne, Antoine Rousseau, Vincent Guinot.

Sélim Cornet developed a numerical model for the hydrodynamics of Vaccares system in Camargue. The data and reference simulations (made with TELEMAC-2D) were provided by Tour du Valat (contact O. Boutron). Sélim's work consisted in the implementation and validation of the porosity shallow water model developped by Vincent GUINOT , in order to obtain accurate but inexpensive simulations of the Vaccares hydrosystem.

7.3. Numerical simulation of coastal flood made by Joanna storm (France, 2008)

Participant: Fabien Marche.

In collaboration with BRGM, a numerical platform based on Fabien Marche's numerical too WaveBox was developed in order to simulate coastal urban submersions associated with intense storms, see [22]. A nudging strategy is implemented with:

- a barotropic model at the regional scale,
- a spectral wave model with embeded meshes accounting for water level evolution and output from the large scale model,
- a free surface Shallow Water model (SURF2D, now called WaveBox) used at very high resolution for the submersion process.

7.4. Analysis of the inclusion of vorticity on fully nonlinear and weakly dispersive long wave models

Participant: Fabien Marche.

We study in [11] the propagation of long waves in the presence of vorticity. In the irrotational framework, the Green-Naghdi equations (also called Serre or fully nonlinear Boussinesq equations) are the standard model for the propagation of such waves. These equations couple the surface elevation to the vertically averaged horizontal velocity and are therefore independent of the vertical variable. In the presence of vorticity, the dependence on the vertical variable cannot be removed from the vorticity equation but it was however shown in [9] that the motion of the waves could be described using an extended Green-Naghdi system. In this paper we propose an analysis of these equations, and show that they can be used to get some new insight into wave-current interactions. We show in particular that solitary waves may have a drastically different behavior in the presence of vorticity and show the existence of solitary waves of maximal amplitude with a peak at their crest, whose angle depends on the vorticity. We also propose a robust and simple numerical scheme validated on several examples. Finally, we give some examples of wave-current interactions with a non trivial vorticity field and topography effects.

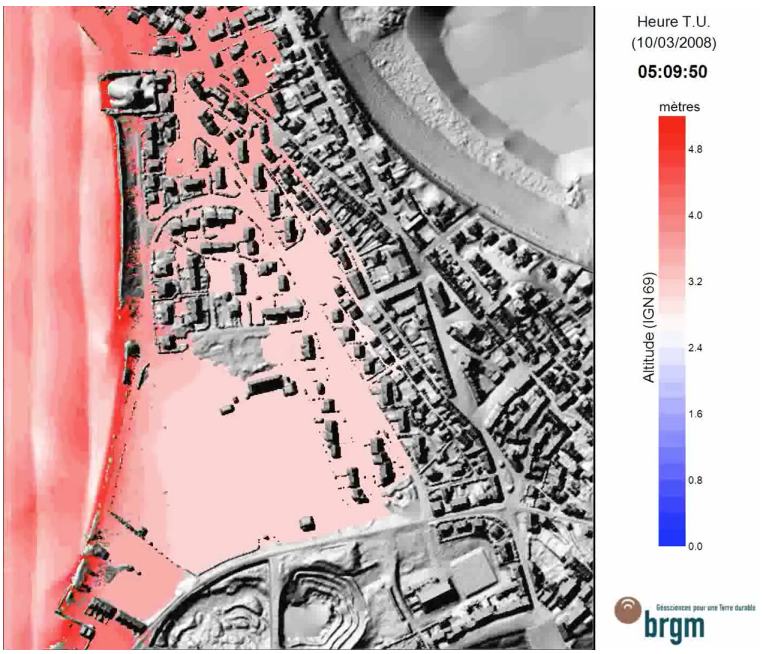


Figure 4. Simulation of Joanna storm (France, 2008). Snapshot of the movie made by BRGM, available on their Youtube page.

7.5. Multiscale aspects for confinement of coastal lagoons

Participant: Antoine Rousseau.

In [3] we expand a previous definition of paralic confinement (see [41]), and make it usable from the modeling slant, before implementing it in numerical tools. More specifically, we here deal with the multiscale aspect of the confinement. If a paralic environment is separated into two (or more) connected areas, we will show that is possible to split the confinement problem into two related problems, one for each area. We also focus on the importance of the interface length between the two subdomains.

7.6. Interface conditions for ocean models

Participant: Antoine Rousseau.

In [5] we are interested in the search of interface conditions to couple hydrostatic and nonhydrostatic ocean models. To this aim, we consider simplified systems and use a time discretization to handle linear equations. We recall the links between the two models (with the particular role of the aspect ratio $\delta = H/L$) and introduce an iterative method based on the Schwarz algorithm (widely used in domain decomposition methods). The convergence of this method depends strongly on the choice of interface conditions: this is why we look for exact absorbing conditions and their approximations in order to provide tractable and efficient coupling algorithms.
In [4] we present a study of optimized Schwarz domain decomposition methods for Navier-Stokes equations. Once discretized in time, optimal transparent boundary conditions are derived for the resulting Stokes equations, and a series of local approximations for these nonlocal conditions are proposed. Their convergence properties are studied, and numerical simulations are conducted on the test case of the driven cavity. It is shown that conditions involving one or two degrees of freedom can improve the convergence properties of the original algorithm.

7.7. Use of remote sensing data for hydraulic modelling

Participant: Carole Delenne.

Wetlands provide a vital resource to ecosystem services and associated rural livelihoods; but their extent, geomorphological heterogeneity and flat topography make the representation of their hydrological functioning complex. The main objective of this area of research is to assess the relevance of remote sensing data for the monitoring and hydraulic modelling of different hydrosystems. In [14], a semi automated method exploiting 526 MODIS 8-day 500 m resolution images was developed to study the spatial and temporal dynamics of the annual flood across the Niger Inner Delta over the period 2000?2011. The flooded area is detected using band ratio indexes. Results were evaluated against classified Landsat images, previous studies and field stage data for a range of hydrological units: river stretches, lakes, floodplains and irrigated areas. Depending on the study area, its extent, and the objective to be reached, different kinds of remote sensing data may be interesting: RADAR, multispectral, high/low spatial/temporal resolution, etc. Several paths for research are currently considered to upgrade the use of remote-sensing data in hydrodynamic modelling:

- use of detected flooded area for model validation and for the calibration of parameters such as friction coefficient.
- topography assessment using the detection of the flooded area of a given wetland at different times.
- characterization of statistical properties of the geometry of the urban medium (useful for large-scale models): statistical, subgrid-scale properties of the topography, and information regarding the flow connectivity properties of the urban medium.

7.8. Lumped hydrological models with infinite characteristic time transfer functions

Participant: Vincent Guinot.

Karst and mountainous catchments usually exhibit rainfall-runoff transfer functions involving multiple time scales. In most existing conceptual, hydrological models of such catchments, multiple time scale response is achieved by introducing several reservoirs and non-linear transfer functions. In [9], multiple time scales are introduced by proposing a transfer function with an infinite characteristic time. The heavy-tailed transfer function behaves asymptotically as an inverse power of time. In the limit of long time scales, the governing equation for the system obeys a fractional differential equation. With a single reservoir, the proposed approach is shown to perform satisfactorily compared to other models of similar or more complex structure. The fractional differential equation is shown to be useless for usual time scales and should not be used in practice.

7.9. Upscaled models for urban floods

Participant: Vincent Guinot.

Shallow water models with porosity have arisen over the last two decades as a promising alternative to refined flow models for the simulation of urban floods. Several porosity-based models have been proposed in the literature. In [10], the integral porosity formalism developed at the University of California Irvine is validated against scale model experiments. A sudden dike breaching near an idealized city layout is simulated in the scale model. Comparison with numerical simulations shows the superiority of the integral porosity model over the single porosity model in reproducing the effects of urban layout anisotropy on flood wave propagation properties. This research has initiated a collaboration between the LEMON team and UC Irvine for the development of a new porosity formalism.

7.10. Models for dispersion in porous media

Participants: Carole Delenne, Vincent Guinot.

Solute dispersion in porous media is usually modelled using Fick's law or fractional variations of the solute dispersion equation. The Fickian model, however, is known to exhibit a number of drawbacks, such as poor scaling properties. This is also true for its fractional counterparts, that perform with limited success when compared to experimental data sets. In [13], a high-quality experimental device is built in the form of periodic heterogeneities of length 15 cm. Placing up to 10 periods in series allows the scaling properties of the dispersion model to be analyzed. Besides providing a high quality experimental database, the results in [13] indicate that (i) previously identified scaling trends for the dispersion coefficient may easily be explained by experiment variability, (ii) there exists a linear transport model that allows the experimental behaviour to be reproduces at all scales, (iii) this model is not the advection-dispersion model (even fractional).

7.11. Invasion in growth-fragmentation-death models

Participant: Fabien Campillo.

In collaboration with Nicolas Champagnat and Coralie Fritsch (Inria Nancy), we present in [20] two approaches to study invasion in growth-fragmentation-death models: one based on a stochastic individual based model and one based on an integro-differential model. The invasion of the population is described by the survival probability for the first model and by an eigenproblem for the second one. We study these two notions of invasion fitness, giving different characterizations of the growth of the population, and we make links between these two complementary points of view. We apply our work in the context of adaptive dynamics in a chemostat model.

7.12. Stochastic growth model with extinction

Participant: Fabien Campillo.

In collaboration with Marc Joannides and Irène Larramendy-Valverde (IMAG / Université de Montpellier), we consider in [6] a stochastic logistic growth model given by a stochastic differential equation featuring both birth and death rates in the drift and diffusion coefficients. Our aim is to infer these rates, based on discrete observations with possible extinction. Since extinction occurs eventually for the model, the density of the diffusion process is not absolutely continuous with respect to the Lebesgue measure; we established the associated Fokker-Planck equation together with appropriate numerical schemes. This formulation allows to design variants of the standard methods that can handle extinction.

MAGIQUE-3D Project-Team

6. New Results

6.1. Seismic Imaging and Inverse Problems

6.1.1. hp-adaptive simulation and inversion of magnetotelluric measurements

Participants: Hélène Barucq, Julen Alvarez Aramberri, David Pardo.

The magnetotelluric (MT) method is a passive exploration technique that aims at estimating the resistivity distribution of the Earth's subsurface, and therefore at providing an image of it. This process is divided into two different steps. The first one consists in recording the data. In a second step, recorded measurements are analyzed by employing numerical methods. In this work, we provide a rigorous mathematical setting in the context of the Finite Element Method (FEM) that helps to understand the MT problem and its inversion process. In order to recover a map of the subsurface based on 2D MT measurements, we employ for the first time in MTs a multigoal oriented self adaptive hp-Finite Element Method (FEM). We accurately solve both the full formulation as well as a secondary field formulation where the primary field is given by the solution of a 1D layered medium. To truncate the computational domain, we design a Perfectly Matched Layer (PML) that automatically adapts to high-contrast material properties that appear within the subsurface and on the air-ground interface. For the inversion process, we develop a first step of a Dimensionally Adaptive Method (DAM) by considering the dimension of the problem as a variable in the inversion. Additionally, this dissertation supplies a rigorous numerical analysis for the forward and inverse problems. Regarding the forward modelization, we perform a frequency sensitivity analysis, we study the effect of the source, the convergence of the hp-adaptivity, or the effect of the PML in the computation of the electromagnetic fields and impedance. As far as the inversion is concerned, we study the impact of the selected variable for the inversion process, the different information that each mode provides, and the gains of the DAM approach.

6.1.2. Ultrasonic imaging of complex media

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé, Julien Diaz, Sébastien Tordeux, Ha Howard Faucher.

In 2015 we have begun a collaborating project with I2M (Physics Acoustics Department of Bordeaux 1 University). We aim at modeling and simulating efficiently the propagation of acoustic waves and later elastodynamic waves in highly heterogeneous media, the final goal is to use topological gradient imaging techniques. Classical techniques as finite elements can be too costly, we propose to design more efficient numerical techniques that exploit the fact that the wavelength is big with respect to the heterogeneities. For instance, we will use numerical upscaling, multiscale homogenization or asymptotic methods. A funding has been obtained for a PhD and a post doctoral position, that have both started in 2015. Our first step is to design a laboratory experiment and a simulation code in order to challenge the limits of the newly derived models and quantify their validity.

6.1.3. Impedance transmission conditions for the electric potential across a highly conductive casing

Participants: Hélène Barucq, Aralar Erdozain, David Pardo, Victor Péron.

In this study we present Impedance Transmission Conditions (ITCs) for the electric potential in the framework of borehole through-casing resistivity measurements. Such ITCs substitute the part of the domain corresponding to a highly conductive casing. The naturally small thickness of the casing makes it ideal for exhibiting ITCs. We numerically observe the delivered order of accuracy.

6.1.4. An efficient truncated SVD of large matrices based on the low-rank approximation for inverse geophysical problems

Participant: Sébastien Tordeux.

We have proposed a new algorithm to compute a truncated singular value decomposition of the Born matrix based on a low-rank arithmetic. Theoretical background to the low-rank SVD method has been investigated: the Born matrix of an acoustic problem can be approximated by a low-rank approximation derived thanks to a kernel independent multipole expansion. The new algorithm to compute T-SVD approximation consists of four steps, and they are described in detail. The largest singular values and their left and right singular vectors can be approximated numerically without performing any operation with the full matrix. The low-rank approximation is computed due to a dynamic panel strategy of cross approximation technique.

6.1.5. Handling clusters with a task-based runtime system: application to Geophysics

Participants: Emmanuel Agullo, Hélène Barucq, Lionel Boillot, George Bosilca, Julien Diaz.

The extreme complexity of hardware platforms makes them harder and harder to program. To fully exploit such machines, the High Performance Computing community often uses a MPI + X (X being pthreads, OpenMP, Cuda ...) programming models. We propose to use an alternative solution consisting of programming at a higher level of abstractions by describing a scientific, high performance computing application as a sequence of tasks whose execution is delegated to a runtime system. We compared MPI-based version and task-based version on Geophysics simulations, especially on the DIVA code of Total. Our previous studies demonstrated the task-based paradigm superiority on shared memory architectures (CPU or MIC), we are now working on distributed and heterogeneous architectures (CPUs+MICs) and, according to our preliminary results, the performances are still better than the MPI-version.

This work has been presented to the conferences PRACEdays [60], Rice Oil&Gas [43] and PASC [37].

6.2. Mathematical modeling of multi-physics involving wave equations

6.2.1. Elasto-acoustic coupling

Participants: Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz, Simon Ettouati, Conrad Hillairet, Elvira Shishenina.

In the framework of her Master thesis, Elvira Shishenina developed a Discontinuous Galerkin Method for the elastoacoustic coupling in time domain. The proposed solution methodology in general and can be applied to any kind of fluxes. We have implemented and validated in Elasticus a centered flux version and an upwind flux version in two dimensions. The time discretization is achieved thanks to Runge Kutta schemes of second and fourth orders.

In frequency domain, Conrad Hillairet developed a 3D elasto-coupling IPDG scheme, in the framework of his Master thesis. It has been implemented and validated in Hou10ni. Moreover, the code is able to handle p-adaptivity and we have proposed a strategy in order to determine the order of the cell as a function of the size of the cell and of the physical parameters. The results of this work have been presented to the Siam Conference on Geosciences in Stanford [39] and to the XXIV Congress on Differential Equations and Applications in Cadiz [32].

Finally, we have considered elastoacoustic coupling with curved interfaces and we have proposed a solution methodology based on Finite Element techniques, which allows for a flexible coupling between the fluid and the solid domain by using non-conforming meshes and curved elements. Since characteristic waves travel at different speeds through different media, specific levels of granularity for the mesh discretization are required on each domain, making impractical a possible conforming coupling in between. Advantageously, physical domains may be independently discretized in our framework due to the non-conforming feature. Consequently, an important increase in computational efficiency may be achieved compared to other implementations based on non-conforming techniques, namely by reducing the total number of degrees of freedom. Differently from other non-conforming approaches proposed so far, our technique is relatively simpler and requires only a

geometrical adjustment at the coupling interface at a preprocessing stage, so that no extra computations are necessary during the time evolution of the simulation. On the other hand, as an advantage of using curvilinear elements, the geometry of the coupling interface between the two media of interest is faithfully represented up to the order of the scheme used. In other words, higher order schemes are in consonance with higher order approximations of the geometry. Concerning the time discretization, we analyzed both explicit and implicit schemes. These schemes are energy conserving and, for the explicit case, the stability is guaranteed by a CFL condition.

This work, which has been achieved in collaboration with Angel Rodriguez Rozas, former post-doc of the team, was published in Journal of Computational Physics [27].

6.2.2. Atmospheric boundary conditions for helioseismology

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé, Victor Péron.

The sun does not have a clear boundary like a solid ball, but it has an atmosphere which can be modeled as an exponential decay of the density. We have studied the replacement of this atmosphere by an equivalent boundary condition in order to avoid meshing the atmosphere. When we assume that the exponential decay is large enough, asymptotic modeling can be performed with respect to this large parameter. Equivalent boundary conditions have been obtained for order 1, 2 and 3, and they substantially improve Dirichlet condition (order 0) for low frequencies. However for high frequencies, these conditions are no longer relevant. We have developed a first-order absorbing boundary condition adapted to an exponential decay of the density, this last condition provides good results for the tested range of frequency. These conditions have been used by the team of Laurent Gizon (Max Planck institute) to retrieve experimental dispersion curves, so called “power spectrum”.

6.2.3. Absorbing Boundary Conditions for 3D elastic TTI modeling

Participants: Hélène Barucq, Lionel Boillot, Julien Diaz.

We propose stable low-order Absorbing Boundary Conditions (ABC) for elastic TTI modeling. Their derivation is justified in elliptic TTI media but it turns out that they are directly usable to non-elliptic TTI configurations. Numerical experiments are performed by using a new elastic tensor source formula which generates P-waves only in an elliptic TTI medium. Numerical results have been performed in 3D to illustrate the performance of the ABCs.

This work has been presented to the conferences PANACM [38] and SEG [33].

6.2.4. The airfoil equation on near disjoint intervals : Approximate models and polynomial solutions

Participants: Leandro Farina, Marcos Ferreira, Victor Péron.

In [26], the airfoil equation is considered over two disjoint intervals. Assuming the distance between the intervals is small an approximate solution is found and relationships between this approximation and the solution of the classical airfoil equation are obtained. Numerical results show the convergence of the approximation to the solution of the original problem. Polynomial solutions for an approximate model are obtained and a spectral method for the generalized airfoil equation on near disjoint intervals is proposed.

6.2.5. Finite element subproblem method

Participants: Patrick Dular, Christophe Geuzaine, Laurent Krähenbühl, Victor Péron.

In [25], progressive refinements of inductors are done with a subproblem method, from their wire or filament representations with Biot-Savart models up to their volume finite-element models, from statics to dynamics. The reaction fields of additional magnetic and/or conducting regions are also considered. Accuracy improvements are efficiently obtained for local fields and global quantities, i.e., inductances, resistances, Joule losses, and forces.

6.2.6. Asymptotic study for Stokes-Brinkman model with Jump embedded transmission conditions

Participants: Philippe Angot, Gilles Carbou, Victor Péron.

In [18], one considers the coupling of a Brinkman model and Stokes equations with jump embedded transmission conditions. Assuming that the viscosity in the porous region is very small, we derive a Wentzel-Kramers-Brillouin (WKB) expansion in power series of the square root of this small parameter for the velocity and the pressure which are solution of the transmission problem. This WKB expansion is justified rigorously by proving uniform errors estimates.

6.2.7. On the solution of the Laplace equation in 3-D domains with cracks and elliptical edges

Participants: Victor Péron, Samuel Shannon, Zohar Yosibash.

An explicit asymptotic solution to the elasticity system in a three-dimensional domain in the vicinity of an elliptical crack front, or for an elliptical sharp V-notch is still unavailable. Towards its derivation we first consider in [30] the explicit asymptotic solutions of the Laplace equation in the vicinity of an elliptical singular edge in a three-dimensional domain. Both homogeneous Dirichlet and Neumann boundary conditions on the surfaces intersecting at the elliptical edge are considered. The dual singular solution is also provided to be used in a future study to extract the edges flux intensity functions by the quasi-dual function method. We show that just as for the circular edge case, the solution in the vicinity of an elliptical edge is composed of three series, with eigenfunctions being functions of two coordinates.

In [29] the singular solution of the Laplace equation with a straight-crack is represented by a series of eigenpairs, shadows and their associated edge flux intensity functions (EFIFs). We address the computation of the EFIFs associated with the integer eigenvalues by the quasi dual function method (QDFM).The QDFM is based on the dual eigenpairs and shadows, and we show that the dual shadows associated with the integer eigenvalues contain logarithmic terms. These are then used with the QDFM to extract EFIFs from p-version finite element solutions. Numerical examples are provided.

6.3. Supercomputing for Helmholtz problems

6.3.1. High order methods for Helmholtz problems in highly heterogeneous media

Participants: Théophile Chaumont-Frelet, Henri Calandra, Hélène Barucq, Christian Gout.

The numerical solution of Helmholtz problems set in highly heterogeneous media is a tricky task. Classical high order discretizations fail to handle such propagation media, because they are not able to capture any of the scales of the velocity parameter. Indeed, they are build upon coarse meshes and therefore, if the velocity parameter is taken to be constant in each cell (through averaging, or local homogenization strategy), scale information is (at least partially) lost. We propose to overcome this difficulty by introducing a multiscale medium approximation strategy. The velocity parameter is not assumed to be constant on each cell, but on a submesh of each cell. If the submeshes are designed properly, the medium approximation method is equivalent to a quadrature formula, adapted to the medium. In particular, we show that this methodology has roughly the same computational cost as the classical finite element method. This new solution methodology has been presented in a paper under revision. We have performed a mathematical analysis of the multiscale medium approximation techniques to higher order discretization. First, we show that the heterogeneous Helmholtz problem is well-posed and derive stability estimates with respect to the right hand side, and with respect to variations of the velocity parameter, justifying the use of medium approximation. Those results are obtained assuming the velocity parameter is monotonous and that the propagation medium is closed by first order absorbing boundary conditions. However, these hypothesis are not mandatory to discretize the problem. Second, we turn to the analysis of finite element schemes with subcell variations of the velocity. In particular, we show that even if the solution can be rough inside each cell because of velocity jumps, we are able to extend the asymptotic error estimates obtained in [93] to heterogeneous media with non-matching mesh in case of elements of order $1 \leq p \leq 3$. Third, we investigate numerically the stability of the scheme when the

frequency is increasing to figure out optimal meshing conditions. We show that in simple media, the optimal homogeneous pre-asymptotic error estimates are still valid. However, in more complex cases, it looks like this condition is not sufficient anymore. Apart from showing that the homogeneous results are not always applicable to the heterogeneous Helmholtz equation, we are not able to give a clear answer to the question. Finally, we are able to conclude that high order methods are actually interesting: in our examples, $p = 4$ discretizations always yield a smaller linear system than lower order discretizations for the same precision.

6.3.2. *Hybridizable Discontinuous Galerkin method for the elastic Helmholtz equations*

Participants: Marie Bonnasse-Gahot, Henri Calandra, Julien Diaz, Stéphane Lanteri.

In the framework of the PhD thesis of Marie Bonnasse-Gahot, we have proposed an hybridizable discontinuous Galerkin method for solving the anisotropic elastodynamics wave equations in harmonic domain, in two and three dimensions. The method was implemented in Hou10ni and in the platform of Total. We have analyzed the performance of the proposed method in 2D on simple test case and compared it to classical DG methods. We have shown that the HDG method provides a more accurate solution for less computational cost provided that the order is high enough. We have illustrated the usefulness of the p-adaptivity in 2D, which allows to reach the accuracy of a global method of degree p for the costs of a global method of degree $p-1$ or $p-2$. This feature is already implemented in the 3D code. We now have to determine an accuracy criteria for assigning an order to a given cell, similar to the criteria we proposed in 2D.
For the numerical analysis of the scheme, we have shown that the HDG method could be rewritten as an upwind fluxes DG method and one of our perspectives is to use this equivalence in order to perform a dispersion analysis following the work of Ainsworth, Monk and Muniz [64].
We have shown that HDG could be used for 2D simulation on geophysical benchmark, and we will now implement the method in a Reverse Time Migration software, the ultimate goal being to couple HDG method with a full waveform inversion solver. In order to tackle more realistic test cases in 3D, it will be mandatory to improve the linear solver and we are now considering the use of an hybrid solver such as Maphys developed by the Inria team-project HIEPACS.
The results of this work have been presented at the "SIAM Conference on Geosciences" [48] and at the "Oil and Gas HPC Workshop" [49].

6.4. Hybrid time discretizations of high-order

6.4.1. *High-order symmetric multistep schemes for wave equation*

Participants: Juliette Chabassier, Marc Duruflé, Guillaume Marty.

We have studied high-order symmetric multistep schemes for the second-order formulation $y'' = f(t, y)$ during the internship of Guillaume Marty. The stability condition (CFL) can be optimized for explicit schemes since they have free parameters. However, this optimization procedure is not easy since the optimum is reached for forbidden values (values for which the high-order accuracy is no longer obtained). We have proposed acceptable values of free parameters for schemes of order 4, 6 and 8. These schemes have been tested for the wave equation, they suffer from a lack of robustness with respect to rounding numerical errors. The stability of implicit schemes has also been explored. For fourth-order schemes, a family of energy-conserving schemes has been obtained. However, we have not found unconditionally stable high-order schemes, which is well-known for the first-order formulation as Dahlquist's barrier. It seems that for the second-order formulation, this barrier holds and only second-order accurate schemes are unconditionally stable. Implicit high-order schemes have a maximum CFL of $\sqrt{6}$, the same CFL as the standard θ-scheme with $\theta = \frac{1}{12}$. As a result, the implicit version of these schemes does not have a practical interest.

6.4.2. *High order conservative explicit and implicit schemes for wave equations.*

Participants: Juliette Chabassier, Sébastien Imperiale.

In 2015 we have studied the space/time convergence of a family of high order conservative explicit and implicit schemes for wave equations. An original proof of convergence has been proposed and provides an understanding of the lack of convergence of some schemes when the time step approaches its greatest admissible value for stability (CFL condition). An article is being written and will be submitted soon.

6.4.3. Multi-level explicit local time-stepping methods for second-order wave equations

Participants: Julien Diaz, Marcus Grote.

Local mesh refinement severely impedes the efficiency of explicit time-stepping methods for numerical wave propagation. Local time-stepping (LTS) methods overcome the bottleneck due to a few small elements by allowing smaller time-steps precisely where those elements are located. Yet when the region of local mesh refinement itself contains a sub-region of even smaller elements, any local time-step again will be overly restricted. To remedy the repeated bottleneck caused by hierarchical mesh refinement, multi-level local time-stepping methods are proposed, which permit the use of the appropriate time-step at every level of mesh refinement. Based on the LTS methods from Diaz and Grote [82], these multi-level LTS methods are explicit, yield arbitrarily high accuracy and conserve the energy.

The method was published in Computer Methods in Applied Mechanics and Engineering [24].

SAGE Project-Team

7. New Results

7.1. Numerical algorithms

7.1.1. Introduction to computational linear algebra

Participant: Jocelyne Erhel.

Publications: [23]

Abstract: The book "Introduction to Computational Linear Algebra" presents classroom-tested material on computational linear algebra and its application to numerical solutions of partial and ordinary differential equations. The book is designed for senior undergraduate students in mathematics and engineering as well as first-year graduate students in engineering and computational science.

The text first introduces BLAS operations of types 1, 2, and 3 adapted to a scientific computer environment, specifically MATLAB. It next covers the basic mathematical tools needed in numerical linear algebra and discusses classical material on Gauss decompositions as well as LU and Cholesky's factorizations of matrices. The text then shows how to solve linear least squares problems, provides a detailed numerical treatment of the algebraic eigenvalue problem, and discusses (indirect) iterative methods to solve a system of linear equations. The final chapter illustrates how to solve discretized sparse systems of linear equations. Each chapter ends with exercises and computer projects.

7.1.2. Hybrid algebraic sparse linear solvers

Participants: Jocelyne Erhel, David Imberti.

Grants and projects: EXA2CT 9.2.1 , EoCoE 9.2.2 , C2S@EXA 9.1.2

Publications: in preparation.

Abstract: Sparse linear systems arise in computational science and engineering. The goal is to reduce the memory requirements and the computational cost, by means of high performance computing algorithms. Krylov methods combined with Domain Decomposition are very efficient for both fast convergence and fast computations.

7.1.3. Hastings-Metropolis Algorithm on Markov Chains for Small-Probability Estimation

Participant: Lionel Lenôtre.

Grants: H2MNO4 9.1.1

Publications: [12]

Abstract: Shielding studies in neutron transport, with Monte Carlo codes, yield challenging problems of small-probability estimation. The particularity of these studies is that the small probability to estimate is formulated in terms of the distribution of a Markov chain, instead of that of a random vector in more classical cases. Thus, it is not straightforward to adapt classical statistical methods, for estimating small probabilities involving random vectors, to these neutron-transport problems. A recent interacting-particle method for small-probability estimation, relying on the Hastings-Metropolis algorithm, is presented. It is shown how to adapt the Hastings-Metropolis algorithm when dealing with Markov chains. A convergence result is also shown. Then, the practical implementation of the resulting method for small-probability estimation is treated in details, for a Monte Carlo shielding study. Finally, it is shown, for this study, that the proposed interacting-particle method considerably outperforms a simple Monte Carlo method, when the probability to estimate is small.

7.1.4. A Strategy for the Parallel Implementations of Stochastic Lagrangian Methods

Participant: Lionel Lenôtre.

Grants and projects: H2MNO4 9.1.1

Software: PALMTREE 6.5

Publications: [32]

Abstract: We present some investigations on the parallelization of a stochastic Lagrangian simulation. For the self sufficiency of this work, we start by recalling the stochastic methods used to solve Parabolic Partial Differential Equations with a few physical remarks. Then, we exhibit different object-oriented ideas for such methods. In order to clearly illustrate these ideas, we give an overview of the library PALMTREE that we developed. After these considerations, we discuss the importance of the management of random numbers and argue for the choice of a particular strategy. To support our point, we show some numerical experiments of this approach, and display a speedup curve of PALMTREE. Then, we discuss the problem in managing the parallelization scheme. Finally, we analyze the parallelization of hybrid simulation for a system of Partial Differential Equations. We use some works done in hydrogeology to demonstrate the power of such a concept to avoid numerical diffusion in the solution of Fokker-Planck Equations and investigate the problem of parallelizing scheme under the constraint entailed by domain decomposition. We conclude with a presentation of the latest design that was created for PALMTREE and give a sketch of the possible work to get a powerful parallelized scheme.

7.1.5. About a generation of a log-normal correlated field

Participants: Jocelyne Erhel, Géraldine Pichot.

Grants: HYDRINV 9.3.3 , H2MN04 9.1.1

Software: GENFIELD 6.1

Publications: [18].

Abstract: Uncertainty quantification often requires the generation of large realizations of stationary Gaussian random field over a regular grid.

We compare the classical methods used to simulate the field defined by its covariance function, namely the Discrete Spectral method, the Circulant Embedding approach, and the Discrete Karhunen-Loève approximation. We design and implement a parallel algorithm related to the Discrete Spectral method.

7.2. Numerical models and simulations applied to heat transfer

7.2.1. Small scale modeling of porous media

Participants: Édouard Canot, Salwa Mansour.

Grants: ECOS Sud Chili (ARPHYMAT project) 9.3.2

Software: GLiMuH 6.2

Publications: [13]

Conferences: [20]

Abstract: This study is devoted to the heat transfer between two spherical grains separated by a small gap; dry air is located around the grains and a liquid water meniscus is supposed to be present between them. This problem can be seen as a micro-scale cell of an assembly of solid grains, for which we are looking for the effective thermal conductivity. For a fixed contact angle and according to the volume of the liquid meniscus, two different shapes are possible for the meniscus, giving a “contacting” state (when the liquid makes a true bridge between the two spheres) and a “non-contacting” one (when the liquid is split in two different drops, separated by a thin air layer); the transition between these two states occurs at different times when increasing or decreasing the liquid volume, thus leading to a hysteresis behavior when computing the thermal flux across the domain. We consider also another process where humidity varies, for example during an evaporation or condensation process; in this situation, the shape of the menisci changes a lot, because some liquid bridges may break, and this can strongly affect the effective thermal conductivity. Then, the reorganization of the liquid menisci is predicted, especially their surface area variation; it is an important parameter for a global model of the evaporation phenomenon in wet porous media.

7.2.2. Inverse problem for determining the thermo-physical properties of a porous media

Participants: Édouard Canot, Salwa Mansour.

Grants: HYDRINV 9.3.3

Software: TPIP (6.7)

Publications: [15], [27]

Conferences: [22]

Abstract: This study concerns the inverse problem which consists of the estimation of thermophysical properties of the soil knowing the temperature at few selected points of the domain. In order to solve this inverse problem, we used the least square criterion where we try to minimize the error function between real measures and simulated ones. The coupled system composed of the energy equation together with the three sensitivity boundary initial problems resulting from differentiating the basic energy equation with respect to the soil properties must be solved. To overcome the stiffness of our problem (due to the use of Apparent Heat Capacity method), the high nonlinearity of the coupled system and the problem of large residuals we used the Damped Gauss Newton and Levenberg-Marquardt methods. To take into account uncertainties of the position of the sensors, some constraints have been added to the least square problem. Results are good when the number of sensors is sufficiently large.

7.2.3. Evaporation/Condensation in a wet granular medium: the EWGM model

Participants: Édouard Canot, Salwa Mansour.

Grants: ECOS Sud Chili (ARPHYMAT project) 9.3.2

Software: HeMaTis (6.4)

Publications: [26], [25]

Abstract: The physical model of the HeMaTiS code (6.4) has been completed by a new variant dedicated to the unsaturated case. The pendular regime concerns the special case where a very few quantity of liquid water is contained in a granular medium. The new model involves seven variables and can be considered as a two-phase two-component one; it contains both air and water, this latter component being liquid or gas. Generally, the diffusive transport of humidity in soils is extremely slow, we numerically show that humidity is convected quickly when the medium is subjected to a strong temperature gradient. The key feature of the thermal process is the simultaneous evaporation and condensation of water near a discontinuity of the liquid layout.

7.3. Models and simulations for skew diffusion

7.3.1. Simulating Diffusion Processes in Discontinuous Media: Benchmark Tests

Participant: Géraldine Pichot.

Grants: H2MN04 9.1.1

Software: SBM 6.6

Publications: submitted.

Abstract: We present several benchmark tests for Monte Carlo methods for simulating diffusion in one-dimensional discontinuous media, such as the ones arising the geophysics and many other domains. These benchmarks tests are developed according to their physical, statistical, analytic and numerical relevance. We then perform a systematic study on four numerical methods.

7.3.2. One-dimensional skew diffusions: explicit expressions of densities and resolvent kernel

Participants: Lionel Lenôtre, Géraldine Pichot.

Grants: H2MN04 9.1.1

Publications: [31]

Abstract: The study of skew diffusion is of primary concern for their implication in the modeling and simulation of diffusion phenomenons in media with interfaces. First, we provide results on one-dimensional processes with discontinuous coefficients and their connections with the Feller theory of generators as well as the one of stochastic differential equations involving local time. Second, in view of developing new simulation techniques, we give a method to compute the density and the resolvent kernel of skew diffusions. Explicit closed-form are given for some particular cases.

7.3.3. Algorithms for the simulation of Feller processes

Participant: Lionel Lenôtre.

Grants and projects: H2MNO4 9.1.1 .

Publications: [34].

Abstract: Two new numerical schemes are created for Skew Diffusions processes. Both algorithms rely on a more generic numerical scheme that can be used for any kind of Feller processes. The proof of convergence for this generic numerical scheme is performed.

7.3.4. Theoretical results on multidimensional Skew Diffusions

Participant: Lionel Lenôtre.

Grants and projects: H2MNO4 9.1.1 .

Publications: [33].

Abstract: Some significant results on the distribution of the marginal processes of multidimensional Skew Diffusions are found together with new formula. In addition, totally analytical proofs of some results and algorithms given by A. Lejay are given.

7.4. Models and simulations for flow and transport in porous fractured media

7.4.1. An adaptive sparse grid method for elliptic PDEs with stochastic coefficients

Participant: Jocelyne Erhel.

Grants and projects: HYDRINV 9.3.3 , H2MN04 9.1.1

Publications: [14].

Abstract: The stochastic collocation method based on the anisotropic sparse grid has become a significant tool to solve partial differential equations with stochastic inputs. The aim is to seek a vector of weights and a convenient level of interpolation for the method. The classical approach uses an a posteriori approach on the solution, which causes an additional prohibitive cost.

In this work, we discuss an adaptive approach of this method to calculate the statistics of the solution. It is based on an adaptive approximation of the *inverse* diffusion parameter. We construct an efficient error indicator which is an upper bound of the error on the solution. In the case of unbounded variables, we use an appropriate error estimation to compute suitable weights for the method. Numerical examples are presented to confirm the efficiency of the approach, and to show that the cost is considerably reduced without loss of accuracy.

7.4.2. A global reactive transport model applied to the MoMaS benchmark

Participant: Jocelyne Erhel.

Grants and projects: H2MN04 9.1.1

Software: GRT3D 6.3

Publications: [19].

Abstract: Reactive transport models are very useful for groundwater studies such as water quality, safety analysis of waste disposal, remediation, and so on. The MoMaS group defined a benchmark with several test cases. We present results obtained with a global method and show through these results the efficiency of our numerical model.

7.4.3. About some numerical models for geochemistry

Participant: Jocelyne Erhel.

Grants and projects: H2MN04 9.1.1

Publications: [16], [17].

Abstract: Reactive transport models are very useful to study the fate of contaminants in grounwater. These models couple transport equations with geochemistry equations. In this talk, we focus on precipitation and dissolution chemical reactions, because they induce numerical difficulties.

We consider a set of solute species and minerals, with precipitation occuring when a saturation threshold is reached. A challenge is to detect which minerals are dissolved and which minerals are precipitated. This depends on the total quantities of chemical species. We propose an analytical approach to build a phase diagram, which provides the interfaces between the different possible cases. We illustrate our method with three examples arising from brine media and acid mine drainage.

7.4.4. Power-averaging method to characterize and upscale permeability in DFNs

Participants: Jean-Raynald de Dreuzy, Géraldine Pichot.

Publications: [21].

Abstract: In a lot of geological environments, permeability is dominated by the existence of fractures and by their degree of interconnections. Flow properties depend mainly on the statistical properties of the fracture population (length, apertures, orientation), on the network topology, as well as on some detailed properties within fracture planes. Based on an extensive analysis of 2D and 3D DFNs as well as on reference connectivity structures, we investigate the relation between the local fracture structures and the effective permeability. Defined as the relative weight between the two extreme harmonic and arithmetic means, the power-law averaging exponent gives a compact way to compare fracture network hydraulics. It may further lead to some comprehensive upscaling rules.

SERENA Team

7. New Results

7.1. Guaranteed bounds for Laplace eigenvalues and eigenvectors

In [22], we have derived a posteriori error estimates for the Laplace eigenvalue problem. Guaranteed, fully computable, and optimally convergent upper and lower bounds for the first eigenvalue are given. They are valid under explicit, a posteriori conditions on the computational mesh and on the approximate solution. Guaranteed, fully computable, and polynomial-degree robust bounds for the energy error in the approximation of the first eigenvector are derived as well, under the same conditions. Remarkably, all the constants in our theory can be fully estimated, and no convexity/regularity assumption on the computational domain/exact eigenvector(s) is needed. This general result can still be improved when an elliptic regularity assumption is satisfied (with known constants), typically for convex two-dimensional domains. The application of our framework to conforming finite element approximations of arbitrary polynomial degree is provided, along with a numerical illustration on a set of test problems.

STEEP Project-Team

6. New Results

6.1. Methods for the calibration of LUTI models

The setting up of a LUTI model requires, like most numerical models, at least one phase of parameter estimation. This is concisely referred to here as calibration, although the calibration of a LUTI model also entails other aspects such as the definition of spatial zones, of economic sectors, etc. The TRANUS LUTI model plus software, like many other existing models, come along with a relatively simple calibration methodology. Most LUTI models indeed perform parameter estimation in a piecewise fashion, by sequentially estimating subsets of parameters. While this reduces the mathematical and computational complexity of calibration, neglecting the interactions across different modules and their parameters, may result in a significant loss of a model?s quality. A second issue is that TRANUS, like several other LUTI softwares, employs rudimentary numerical routines for parameter estimation. We aim at reducing these weaknesses.

In 2014, we had obtained first results along these lines: parameter estimation of the so-called shadow prices (specific parameters of the TRANUS model) was posed as optimization problem and several solution procedures were developed which were based on "unwinding" the dynamics of the model, making the problem amenable to standard numerical optimisation techniques.

The work continued throughout 2015, along different directions. First, the calibration was extended to handle several different parameter types simultaneously (shadow prices as well as the so-called substitution parameters, which are notoriously difficult to estimate) [7]. Such a simultaneous estimation of different parameter sets seems to be rare in LUTI practice.

Second, we proposed a methodology for assessing properties (convergence, accuracy) of our (and other) LUTI calibration methods [6]. This consists in generating synthetic data, starting from a model calibrated on observed data, such that the synthetic data are completely consistent, i.e. there are a set of model parameters that exactly reproduce these data (which is not the case with the observed data). The ground truth model parameters are then easily used to assess calibration parameters. Such a methodology, akin to twin experiments in data assimilation, seems to be novel for LUTI research.

Third, LUTI models are usually calibrated on a base year or period, and used in a prospective manner (via simulated "predictions" for future periods). As with any numerical model, it is wise to make sure that a calibrated LUTI model does not overfit the observations used for calibration; otherwise, its "predictions" may be grossly erroneous. Potential overfitting does not seem to have been deeply studied in the LUTI literature. We have made an initial investigation by calibrating different versions of a TRANUS model, varying the number of shadow prices used as parameters in the model (there is, by default, one shadow price per combination of geographical zone of the study area and economic sector) [6]. For instance, after an initial calibration using all shadow prices, we then dropped the two third smallest of them and re-calibrated the model using the remaining third. The goodness-of-fit to observations was worse by only 3%. In line with well-known principles of model selection (Occam's razor), this may suggest that it is preferable to use the model with fewer parameters when doing predictions. This is still work in progress; showing its relevance is planned to be studied by a similar methodology as above, using simulated twin experiments.

This work is done in collaboration with Arthur Vidard from the AIRSEA Inria project-team and Brian Morton from the University of North Carolina at Chapel Hill.

6.2. Estimation of Sobol' indices combining nested designs and replication method

Sensitivity analysis studies how the uncertainty on an output of a mathematical model can be attributed to sources of uncertainty among the inputs. Global sensitivity analysis of complex and expensive mathematical

models is a common practice to identify influent inputs and detect the potential interactions between them. Among the large number of available approaches, the variance-based method introduced by Sobol' allows to calculate sensitivity indices called Sobol' indices. Each index gives an estimation of the influence of an individual input or a group of inputs. These indices give an estimation of how the output uncertainty can be apportioned to the uncertainty in the inputs. One can distinguish first-order indices that estimate the main effect from each input or group of inputs from higher-order indices that estimate the corresponding order of interactions between inputs. This estimation procedure requires a significant number of model runs, number that has a polynomial growth rate with respect to the input space dimension. This cost can be prohibitive for time consuming models and only a few number of runs is not enough to retrieve accurate informations about the model inputs.

The use of replicated designs to estimate first-order Sobol' indices has the major advantage of reducing drastically the estimation cost as the number of runs becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. However the replication method still requires a large number of model evaluations. By rendering this method iterative, the required number of evaluations can be controlled. The estimation procedure is therefore stopped when the convergence of estimates is considered reached. The key feature of this approach is the construction of nested designs. For the estimation of first-order indices, we exploit a nested Latin Hypercube already introduced in the litterature. For the estimation of closed second-order indices, two methods are proposed to construct a nested orthogonal array. One of the two leads to a partition of the coordinate space over a Galois field.

This work has been done in collaboration with Laurent Gilquin and Clementine Prieur (members of Moise Team), and belongs to the work program of CiTIES project. The proposed procedure will be soon applied to study the sensitivity of TRANUS model.

6.3. Environmental pressures associated with material flows

This work is the follow-up of a previous study dedicated to material flow analysis of the French cereal supply chain at various spatial levels [12]. The goal was twofold:

- trace the flows to their initial geographic origin or final destination,
- couple material flows with a series of environmental pressures associated to them.

For the first goal, we used an Absorbing Markov Chains model where transient states represent raw or semi-products and absorbing states correspond to final consumption products. For the second goal, we used pressure ratios for environmental pressures most relevant to cereals, namely energy use, GHG emissions, land use, use of pesticides and blue water footprint. The model is based on physical supply and use tables and distinguishes between 21 industries, 22 products, 38 regions of various spatial resolution (22 French regions, 10 countries, 6 continents) and 4 modes of transport. Illustrative examples were taken in order to demonstrate the versatility of the results produced, for instance: What is the fate/supply area of a region's production/consumption? What are the production and consumption footprint of a region? These results are designed to be a first step towards scenario analysis for decision-aiding that would also include socioeconomic indicators [13]

6.4. Material flows of the French forest-wood supply chain

The methodology developed in Courtonne et al. [12] on the case of the cereal supply chain was adapted to the French forest-wood supply chain in collaboration with the Laboratoire d'Economie Forestière. Supply chain flows were estimated both at the national and regional scale for wood harvest, addition to stock, production, imports and exports of construction wood, industrial wood and energy wood. These results can be a basis to analyze potential value losses throughout the supply chain, for instance exports of raw materials instead of local transformation. They can also be use to study the competitive use of wood for energy, industry and construction/furnitures, which is a question of growing importance in the context of energy transition.

6.5. Land Use/Land Cover Change (LUCC) Modelling and Ecosytem Services

The ESNET project (EcoSystem services NETworks) is a collaboration lead by LECA (Laboratoire d'ECologie Alpine, UJF) that aims at characterizing the ecosystem services of the Grenoble urban region (about 2/3 of the Isere département) at the 2040 horizon under various constraints of urban policy planning, changes in agricultural and forest management, and climate change impact on ecosystems.

The cartographic effort of the project has been hosted at Inria, and has produced in 2014 three very detailed maps of land use and land cover at the 15m resolution over the whole study area, in 1998, 2003 and 2009, respectively. An extensive analysis of the patterns of landscape change has been performed from these data, with special emphasis on urban sprawl and the associated loss of arable land. This work has been submitted for publication very recently.

A second related piece of work has been produced, both from this cartographic source and more specific remote sensing data. The objective was to characterize in detail the cultural successions and patterns of the study area, in order to produce fine scale maps of associated ecosystem services. This work has just been submitted for publication at the time of writing.

Finally, the scenarios of future land use and land cover that have been elaborated for this project have all been projected at the 2040 horizon at the 15m scale with a well-known LUCC modelling environment (Dinamica) for urban changes, and from in-project models for the other types of land use and cover. A third article bearing on these scenarios and their LUCC modelling is in preparation.

As an aside of this land use/cover modelling effort, the STEEP team has been involved in two of the most detailed ecosystem service models developed for the project: one for the analysis of crop production and associated nitrogen cycle assessment — with the final aim to constrain both the production services and water quality issue related to nitrogen loading — and one on "recreational services". Our involvement in these models was directly related to the acquired expertise in land use modelling.

In the process of this modelling exercise, the STEEP team has acquired an in-depth knowledge and expertise of LUCC models. As a consequence, various theoretical flaws have been identified in the theoretical foundations of such models. An important by-product of the ESNET project is therefore a series of articles in preparation in the team, whose aim is to address and correct these flaws in a very general way; it is hoped that LUCC theory will be put on a more serious theoretical footing as a result of this methodological work, which should be submitted for publication in 2016 for the most part. Another but more limited methodological contribution bears on the development of error models for landscape metrics, another important methodological blind-spot in the specialized literature.

6.6. A benchmarking tool to assess the compatibility of the INDCs with the 2°C long-term target

Climate negotiations related to global warming are another important issue of sustainable development. In this framework that is place at international scale we have developed a benchmarking tool which allows to assess the compatibility of the Intended Nationally-Determined Contributions (INDCs) given by all states for the Conference COP21, with the 2°C long-term target. This benchmarking tool has been designed via an adaptation of REDEM model and algorithm we developed in 2014 with EDDEN laboratory. This tool has been used by the "*Groupe Interdisciplinaire sur les Contributions Nationales*" (GICN) which has been mandated by french ministry of Sustainable Development to prepare the climate change conference COP21 at Paris.

TONUS Team

7. New Results

7.1. Particle-in-cell simulations for highly oscillatory Vlasov-Poisson systems

Participants: Edwin Chacon Golcher, Sever Adrian Hirstoaga [correspondent], Mathieu Lutz.

The aim of the following works is to study the dynamics of charged particles under the influence of a strong magnetic field by numerically solving in an efficient way the Vlasov-Poisson and guiding center models.

First, we work on the development of the time-stepping method introduced in [7], [8] in two directions: improve the accuracy of the algorithm and adapt the algorithm for general configuration of magnetic field.

Second, by using appropriate data structures, we implement an efficient (from the memory access point of view) Particle-In-Cell method which enables simulations with a large number of particles. Thus, we present in [13] numerical results for classical one-dimensional Landau damping and two-dimensional Kelvin-Helmholtz test cases. The implementation also relies on a standard hybrid MPI/OpenMP parallelization. Code performance is assessed by the observed speedup and attained memory bandwidth. A convergence result is also illustrated by comparing the numerical solution of a four-dimensional Vlasov-Poisson system against the one for the guiding center model.

7.2. Eulerian simulations of parallel transport in the SOL

Participants: David Coulette, Sever Adrian Hirstoaga [correspondent], Giovanni Manfredi.

We continue to investigate kinetic models for simulating the heat load on the divertor plates during transient events as edge-localised modes (ELMs). Our previous work [36] deals with Vlasov-Poisson equations for two particle species for the dynamics of their transport parallel to the magnetic field. We started to improve this model by adding an equation for the evolution in time of the perpendicular temperatures. These equations take also into account the collisions between species which may play a role over long times. The first numerical results are encouraging, showing different features with respect to the older (simpler) model when computing total particles and energy fluxes on the divertor plates.

7.3. Quasi-neutrality equation in a polar mesh

Participants: Christophe Steiner [correspondent], Michel Mehrenberger, Nicolas Crouseilles, Philippe Helluy.

In this work [21], we are concerned with the numerical resolution of the quasi-neutrality equation arising in plasma physics. A classic method is based on a Padé approximation. Two other methods are proposed in this paper: a high order Padé approximation and a direct method in the space configuration which consists in integrating on the gyrocircles using an interpolation operator. Numerical comparisons are performed with analytical solutions and considering the 4D drift-kinetic model with one Larmor radius. This is a preliminary study; further study in GYSELA is envisioned.

7.4. The Semi-Lagrangian method on curvilinear grids

Participants: Aurore Back, Adnane Hamiaz, Michel Mehrenberger [correspondent], Pierre Navaro, Hocine Sellama, Eric Sonnendrücker.

We study the semi-Lagrangian method on curvilinear grids [18], [9]. The classical backward semi-Lagrangian method preserves constant states but is not mass conservative. Natural reconstruction of the field permits nevertheless to have at least first order in time conservation of mass, even if the spatial error is large. Interpolation is performed with classical cubic splines and also cubic Hermite interpolation with arbitrary reconstruction order of the derivatives. High odd order reconstruction of the derivatives is shown to be a good ersatz of cubic splines which do not behave very well as time step tends to zero. A conservative semi-Lagrangian scheme is then described; here conservation of mass is automatically satisfied and constant states are shown to be preserved up to first order in time.

Semi-Lagrangian guiding center simulations are performed on sinusoidal perturbations of cartesian grids, and on deformed polar grids with different boundary conditions. Key ingredients are: the use of a B-spline finite element solver for the Poisson equation and the classical backward semi-Lagrangian method (BSL) for the advection. We are able to reproduce standard Kelvin-Helmholtz and diocotron instability tests on such grids. When the perturbation leads to a strong distorted mesh, we observe that the solution differs if one takes standard numerical parameters that are used in the cartesian reference case. We can recover good results together with correct mass conservation, by diminishing the time step.

7.5. Solving the Guiding-Center model on a regular hexagonal mesh

Participants: Michel Mehrenberger [correspondent], Laura Mendoza, Charles Prouveur, Eric Sonnendrücker.

This work [11] introduces a Semi-Lagrangian solver for the Vlasov-Poisson equations on a uniform hexagonal mesh. The latter is composed of equilateral triangles, thus it doesn't contain any singularities, unlike polar meshes. We focus on the guiding-center model, for which we need to develop a Poisson solver for the hexagonal mesh in addition to the Vlasov solver. For the interpolation step of the Semi-Lagrangian scheme, a comparison is made between the use of box-splines and of Hermite finite elements. The code will be adapted to more complex models and geometries in the future.

7.6. High-order Hamiltonian splitting for Vlasov-Poisson equations

Participants: Fernando Casas, Nicolas Crouseilles, Erwan Faou, Michel Mehrenberger [correspondent].

In this work [12], we consider the Vlasov-Poisson equation in a Hamiltonian framework and derive new time splitting methods based on the decomposition of the Hamiltonian functional between the kinetic and electric energy. Assuming smoothness of the solutions, we study the order conditions of such methods. It appears that these conditions are of Runge-Kutta-Nyström type. In the one dimensional case, the order conditions can be further simplified, and efficient methods of order 6 with a reduced number of stages can be constructed. In the general case, high-order methods can also be constructed using explicit computations of commutators. Numerical results are performed and show the benefit of using high-order splitting schemes in that context. Complete and self-contained proofs of convergence results and rigorous error estimates are also given.

7.7. Velocity space transformations: collisional case

Participants: Emmanuel Franck, Philippe Helluy [correspondent], Laurent Navoret.

The method of "velocity space transformations" allows to obtain an interesting discretization of the Kinetic equations like Vlasov-Poisson or Vlasov Maxwell equations as has been proved in the works of P. Helluy, L. Navoret and N. Pham. During this year, we have begun to extend this method to the collisional case using the entropy variable to write a general collisional operator. To treat all the regimes (small or large collisional regime), asymptotic preserving schemes (stability and convergence independent of the collisional frequency) have been designed. However, this method admits some numerical difficulties if we use the physical entropy to construct the collisional operator. Now we propose to use modified entropy, which has good numerical properties and gives limit regime close to the real one in the low Mach context. If this new approach gives interesting results, we will study the adaptivity of the velocity discrete basis which would allow to treat the collisional and non-collisional regimes with the same method.

7.8. Preconditioning and implicit solvers

Participants: Emmanuel Franck [correspondent], Philippe Helluy, Matthias Hoelzl, Ahmed Ratnani, Malcolm Roberts, Eric Sonnendrücker, Stefano Serra-Capizzano.

The Viscous-resistive MHD model used to simulate the instabilities is a multi-scale models with fast waves. In this context, it is not possible to use full explicit time schemes. However the classical implicit schemes are not usable directly since the matrices are ill-conditioned. For this reason it is necessary to use a preconditioning method. During this year we have studied a method called "physic based preconditioning" for the wave equations which consists to approximate the solution by suitable smaller and simpler systems. The results are very good. After this, we have extended this method to the Linearized Euler equation. During this new study, we have found additional difficulties which appear in some regimes. Two methods to treat this problem will be tested in 2016. We have also implemented a version of this preconditioning for the reduced MHD models of JOREK. The first results are positive. To finish, we have begun a collaboration with S. Serra-Capizzano to study at the theoretical level the physic based preconditioning and propose new preconditioning for each sub-systems of the Physic-Based PC efficient in all the physics regimes and for an arbitrary order.
We have also developed an implicit solver for the transport equation based on the upwind nature of the DG numerical flux. This solver will be used for solving Vlasov models or fluid models thanks to the Lattice-Boltzmann methodology. We have obtained recently a SPPEXA support (http://www.sppexa.de) in a joint french-german-japanese project.

7.9. Finite element for full-MHD problems

Participants: Emmanuel Franck [correspondent], Eric Sonnendrücker.

This work have begun at the end of 2015. It is organized around a PhD: Mustafa Gaja supervised by E. Sonnendrücker, A. Ratnani and E. Franck at the Max-Planck Institute of Plasma Physic. The aim of this work is to design and study compatible finite element method (finite element method which preserve the DeRham sequence and the inclusion between the functional space) for B-Splines. This method will allow to discretize efficiently the Maxwell equations, the MHD model and some operators as curl-curl or grad-div vectorial operators which appear in the physic-based PC. For now, we have begun to study the finite element discretization of vectorial operators which appears in the linearized Euler equations and in the physic-based PC associated.

7.10. Lagrangian averaged gyrokinetic-waterbag continuum

Participant: Nicolas Besse [correspondent].

In this paper [26], we first present the derivation of the anisotropic Lagrangian averaged gyrowaterbag continuum (LAGWBC-α) equations. The gyrowaterbag (nickname for gyrokinetic-waterbag) continuum can be viewed as a special class of exact weak solution of the gyrokinetic-Vlasov equation, allowing to reduce this latter into an infinite dimensional set of hydrodynamic equations while keeping its kinetic features such as Landau damping. In order to obtain the LAGWBC-α equations from the gyrowaterbag continuum we use an Eulerian variational principle and Lagrangian averaging techniques introduced by Holm, Marsden, Ratiu and Shkoller for the mean motion of ideal incompressible flows, extended to barotropic compressible flows by Bhat and co-workers and some supplementary approximations for the electrical potential fluctuations. Regarding to the original gyrowaterbag continuum, the LAGWBC-α equations show some additional properties and several advantages from the mathematical and physical viewpoints, which make this model a good candidate for describing accurately gyrokinetic turbulence in magnetically confined plasma. In the second part of this paper we prove local-in-time well-posedness of an approximate version of the anisotropic LAGWBC-α equations, that we call the “isotropic” LAGWBC-α equations, by using quasilinear PDE type methods and elliptic regularity estimates for several operators.

7.11. Hamiltonian structure, fluid representation, stability for the Vlasov-Dirac-Benney equation

Participants: Claude Bardos, Nicolas Besse [correspondent].

This contribution [23] is an element of a research program devoted to the analysis of a variant of the Vlasov−Poisson equation that we dubbed the Vlasov−Dirac−Benney equation or in short V−D−B equation. As such it contains both new results and efforts to synthesize previous observations. One of main links between the different issues is the use of the energy of the system. In some cases, such energy becomes a convex functional and allows to extend to the present problem the methods used in the study of conservation laws. Such use of the energy is closely related to the Hamiltonian structure of the problem.

7.12. Semi-classical limit of an infinite dimensional system of nonlinear Schrödinger equations

Participants: Claude Bardos, Nicolas Besse [correspondent].

In this paper [24], we study the semi-classical limit of an infinite dimensional system of coupled nonlinear Schrödinger equations towards exact weak solutions of the Vlasov-Dirac-Benney equation, for initial data with analytical regularity in space. After specifying the right analytic extension of the problem and solutions, the proof relies on a suitable version of the Cauchy-Kowalewski Theorem and energy estimates in Hardy type spaces with convenient analytic norms.

7.13. Aligned interpolation for gyrokinetic Tokamak simulations

Participants: Guillaume Latu, Michel Mehrenberger [correspondent], Maurizio Ottaviani, Eric Sonnendrücker.

This work is devoted to study the aligned interpolation method in semi-Lagrangian codes. The scheme is presented and algorithms used implementing the scheme are given. A theoretical justification of the method is given with convergence estimates in the simplified context of 2D constant advection, assuming stability of the scheme. The stability is here studied numerically, letting the formal proof as an open problem. The solution is successfully applied in the gyrokinetic context: first in a simplified case in cylindrical geometry and then in toroidal geometry. In the first case, the solutions provided by simulations based on the scheme are in accordance with linear dispersion analysis; in the second case, numerical simulations produced by the Gysela code are presented, simulation based on the standard scheme are compared to those based on the new aligned scheme. This work will lead to a project of paper, which will be submitted in 2016.

BIOCORE Project-Team

7. New Results

7.1. Mathematical methods and methodological approach to biology

7.1.1. Mathematical analysis of biological models

7.1.1.1. Mathematical study of semi-discrete models

Participants: Jean-Luc Gouzé, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Elsa Rousseau, Nicolas Bajeux.

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [96]. We used such models and analyzed their properties in several practical situations that are developed in Section 7.2.3 , some of them requiring such a modeling to describe external perturbations of natural systems, and others to take seasonality into account. External perturbations of interacting populations occur when some individuals are introduced or removed from a natural system, which occurs frequently in pest control applications, either through the direct removal of pests, or through the introduction of biological control agents [71],[27]. Seasonality is an important property of most agricultural systems in temperate environments since the year is divided into a cropping season and a 'winter' season, where the crop is absent, as in our analysis of the sustainable management of crop resistance to pathogens [25] or in the dynamics of plant pathogens [50].

7.1.1.2. Model reduction and sensitivity analysis

Participants: Suzanne Touzeau, Jean-Luc Gouzé, Stefano Casagranda, Victor Bernal Arzola.

Analysis and reduction of biochemical models. Dynamic models representing complex biological systems with numerous interactions can reach high dimensions and include complex nonlinearities. A model reduction method based on process weighing and pruning was developed and implemented on various models (ERK signaling pathway, circadian rythms in Drosophila) [41]. A global sensitivity analysis was performed to check the method robustness against parameter uncertainty and variability. This work is part of Stefano Casagranda's ongoing PhD thesis and is also a collaboration with Bayer (Sophia-Antipolis).

Parameter identification in compartmental systems. In collaboration with F. Dayan (R&D Manager, Dassault Systèmes), we worked on practical problems of identifiability of parameters in linear pharmacokinetic models. This was the subject of the internship of V. A. Bernal [58].

7.1.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Ismail Belgacem, Olivier Bernard, Stefano Casagranda, Francis Mairet, Sofia Almeida.

7.1.2.1. Continuous models analysis

Piecewise quadratic systems for studying growth rate in bacteria. These new systems (first introduced in [82]) use an expression for growth rate that may depend on any number of variables and have several quadratic modes. Relative to the "classical" piecewise affine systems, this new formulation allows the existence of sliding motion as well as oscillatory behaviour for solutions at the thresholds where the vector fields are opposing [21].

Transcription and translation models in bacteria. We study detailed models of transcription and translation for genes in a bacterium, in particular the model of gene expression of RNA polymerase. We also study other models of the global cellular machinery. This is part of the PhD theses of Ismael Belgacem [11] and Stefano Casagranda, and done in collaboration with Inria IBIS project-team, in particular with D. Ropers.

Design of a bistable switch to control cellular uptake. In a joint work with Diego Oyarzún (Imperial College), we analyse the construction of a synthetic bistable system using an unbranched metabolic chain with a global enzyme regulator, as an application of [109]. Bistability can be achieved by choosing an appropriate pattern of regulation. Robustness conditions are given in terms of the promoter dynamic ranges to guarantee a bistable uptake flux [35].

A reduced model for the mammalian cell cycle. We focused on identifying and analyzing the main mechanisms behind the cell cycle and proposed a mathematical model to describe them. This reduced model successfully reproduces oscillatory behaviors including the progress towards a mitosis phase, and then mitosis itself, characterized by an increase in cyclin B. The model was the topic of a poster at the Signalife Workshop [68]. This is a collaboration with F. Delaunay (Ibv Nice) in the framework of Labex Signalife.

7.1.2.2. Hybrid models analysis

Attractor computation using interconnected Boolean networks. During the visit of Daniel Figueiredo, we have worked on an extension of the method proposed in [83]. The idea is to not only use the attractors but also an appropriate set of strongly connected components in the computation of the asymptotic graph [115]. Numerical simulations show a great improvement in the problem of discarding spurious attractors.

Periodic orbits in non monotonic negative feedback circuits. We study the occurrence of periodic solutions in an n-dimensional class of negative feedback systems defined by smooth vector fields with a window of not necessarily monotonic activity. By circumscribing the smooth system by two piecewise linear ones, we show there exists an invariant toroidal region which contains a periodic orbit of the original smooth system [37].

7.1.2.3. Estimation and control

Optimal allocation of resources in a bacterium. We study by techniques of optimal control the optimal allocation between metabolism and gene expression during growth of bacteria [85], in collaboration with Inria IBIS project-team.

Control of a model of synthesis of a virulence factor. In collaboration with J.-A. Sepulchre (INLN Nice), we model the production of a virulence factor by a bacterium in a continuous stirred tank reactor. The production of this enzyme is genetically regulated, and degrades a polymeric external substrate into monomers. A nonlinear control is built [48].

7.2. Fields of applications

7.2.1. Bioenergy

7.2.1.1. Modelling microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Ghjuvan Grimaud, Quentin Béchet, David Demory, Hubert Bonnefond, Jean-Philippe Steyer, Francis Mairet.

Experimental developments

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae [23] and support model development. These experiments have been carried out in the Lagrangian simulator, under constant or periodic light and temperature, varying the total amount of light dose in the day. The response in terms of storage carbon (triglycerides and carbohydrates) has been measured and correlated to the environment fluctuations.

Other experiments were carried out to reproduce the light signal percept by a cell in a raceway pond [84], derived from hydrodynamical studies [92]. An electronic platform was developed to reproduce this high frequency light signal. The experiments show that the microalgae adapt their pigments to the average light that they have received [23]. Experiments with coloured light demonstrated that the growth rate results from the absorbed light, whatever its wavelength.

A new methodology to measure cell viability has been set up. This approach is very promising to distinguish between net and gross growth rate [20]. It was used in the models to assess the impact of temperature on growth and mortality. The mortality turns out to increase exponentially with temperature. The effect of a short term temperature stress was also tested to understand the consequences of a temperature peak in a cultivation system. Finally, it was shown that microalgae can bear with temperature peaks above $T_{\max}$ if they do not last too long [57].

On top of this, we set up a new experimental platform to carry out pilot experiments with solar light. The platform includes four raceways and the equipment to inoculate and harvest the microalgae [60]. We tested the impact of coloured film mimicking possible photovoltaic material. The collected data were used to calibrate models integrating the light spectrum [64].

These works have been carried out in collaboration with A. Talec, S. Rabouille, and E. Pruvost (CNRS/UPMC -Oceanographic Laboratory of Villefranche-sur-Mer LOV).

In collaboration with the IFREMER-PBA team (Nantes) we contributed to a study on the efficiency of dyes (BODIPY and Nile red) to quantify lipid content in microalgae [38].

Metabolism of carbon storage and lipid production

A macroscopic model for lipid production by oleaginous microalgae [7] has been previously proposed. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. A review of the microalgal metabolism reconstruction [15] together with the associated metabolic models has been carried out [14]. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*. It predicts carbohydrate and lipid accumulation, under conditions of light/dark cycles and/or nitrogen deprivation [72], [1]. A model was developed to represent heterotrophic growth on a mixture of acetate and butyrate [39]. A metabolic model was set up, on the basis of the DRUM framework [1], in order to simulate autotophic, heterotropic and mixotrophic growth, and to determine how to reduce substrate inhibition.

Modelling the coupling between hydrodynamics and biology

In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on multilayer Saint Venant system) with microalgae growth was developed [79]. This model is supported by the work of ANGE aiming at improving the multi-layer Saint-Venant approach to more finely represent the hydrodynamics of the raceway [54].

Modelling the photosynthesis response to fast fluctuating light

The impact of hydrodynamics on the light perceived by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [90]. The light signals that a cell experiences at the Lagrangian scale, depending on the fluid velocity, were then estimated. A Droop-Han model was used to assess the impact of light fluctuation on photosynthesis. A new model accounting for photoacclimation was also proposed [34]. Single cell trajectories were simulated, and the effect on photosynthesis efficiency was assessed using models of photosynthesis [91]. These results were compared to experimental measurements where the high frequency light was reproduced [84].

Modeling microalgae production processes

The integration of different models developed within BIOCORE [54], [19], [7] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems, on the basis of the dynamical model developed to describe microalgal growth under light and nitrogen limitations. The strength of this model is that it takes into account the strong interactions between the biological phenomena (effects of light and nitrogen on growth, photoacclimation ...), temperature effect [78], [111] and the radiative transfer in the culture (light attenuation due to the microalgae).
Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [107]. This model is used to predict lipid production in the perspective of large scale biofuel production [54]. It was also used to assess the microalgal production potential in France,

when taking into account the actual meteorology on a 2.5 degree grid, for 2012, the use of lands, slope, proximity of nutrients and CO_2 [93].
In the framework of the ANR project Purple Sun, we developed a thermic model of a raceway pond within a greenhouse in order to estimate the culture temperature. We also included in the microalgae model the effect of light wavelength. This model has been calibrated on experimental data from LOV and has been used to support lighting strategy in order to optimize microalgal productivity (a patent on this process has been submitted).

Nitrogen fixation by diazotrophs

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [87] quantifying te main fluxes of carbon and nitrogen in the cell. The model was calibrated and validated with the data of three experiments carried out with different duration of the light period and daily dose. Extension of the model were studied to include the effect of temperature [61].

This work is done in collaboration with Sophie Rabouille (CNRS-Oceanographic Laboratory of Villefranche-sur-Mer LOV).

Modelling thermal adaptation in microalgae

An extended statistical analysis was carried out on a database representing the temperature response of more than 200 microalgal species. First the model proposed by [78] turned out to properly reproduce the temperature response. A model was then extracted to predict the observed link between the cardinal temperatures. This lead to the reduction of the parameter number down to 2, with still a good prediction capability.

We have used Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. For a constant temperature, we have shown that the optimal temperature trait tends to equal the environment temperature. We then studied the case where temperature is periodically fluctuating [88]. We now use this method at the scale of the global ocean, validating our approach with experimental data sets from 194 species [42], [49].

7.2.1.2. Control and Optimization of microalgae production

On-line monitoring

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [86]. Several developments were carried out in this direction to improve the design and performances of interval observers, and accounting for a specific structure (*i.e.* triangular) or property (*i.e.* Input to State Stable), [104]. Interval observers were designed for the estimation of the microalgae growth and lipid production within a production process [101][54] and validated experimentally [100][29].

Optimization of the bioenergy production systems

Based on simple microalgae models, analytical optimization strategies were proposed. We assessed strategies for optimal operation in continuous mode using the detailed model for raceways [106], [107]. We first solved numerically an optimal control problem in which the input flow rate of the raceway is calculated such that the productivity in microalgae biomass is maximized on a finite time horizon. Then, we re-analysed the optimization problem and derived a simplified strategy to reach biomass productivities very near to the maximal productivities obtained with the optimal control. These approaches were extended to outdoor cultivation, considering a possible variable culture depth. The optimal strategy for both depth and dilution rate was proposed in order to better manage the process inertia and finally avoid over warming periods. This work was done during the doctoral stay of Riccardo de Luca (Univ. Padova).

We also propose a nonlinear adaptive controller for light-limited microalgae culture, which regulates the light absorption factor (defined by the ratio between the incident light and the light at the bottom of the reactor). We show by numerical simulation that this adaptive controller can be used to obtain near optimal productivity under day-night cycles [31].

Interactions between species

Large scale culture of microalgae for bioenergy involves a large biodiversity. Control of such systems requires to consider the interactions between the different species. Such systems involve bacteria and microalgae, and the competition between these organisms can have several equilibrium points, which can be studied with Monod, Contois and Droop type models [28].

In the framework of the ANR Facteur 4 project, we propose to drive this competition exploring different strategies in order to select species of interest.

We had formerly proposed an adaptive controller which regulates the light at the bottom of the reactor [102]. When applied for a culture with n species, the control law allows the selection of the strain with the maximum growth rate for a given range of light intensity. This is of particular interest for optimizing biomass production as species adapted to high light levels (with low photoinhibition) can be selected. We have also proposed a strategy based on light stresses in order to penalize the strains with a high pigment content and finally select microalgae with a low Chlorophyll content [12]. This characteristic is of particular interest for maximizing biomass production in dense culture. The strategy has been carried out at the LOV and eventually the productivity of *Tisochrysis lutea* was improved by 75% [62]. A patent on this strategy is under submission. Strategies to improve the temperature response have also been proposed. First we modelled the adaptive dynamics for a population submitted to a variable temperature [88]. This was used at the LOV to design experiments with periodic temperature stresses during 200 days aiming at enlarging the thermal niche of *Tisochrysis lutea*. It resulted in an increase by 2 degrees of the thermal niche [12].
Finally, in a more theoretical framework, we studied how to select as fast as possible a given species in a chemostat with two species at the initial instant. Using the Pontryagin maximum principle, we have shown that the optimal strategy is to maintain the substrate concentration to the value maximizing the difference between the growth rates of two species [73]. We now try to extend this result for n species with mutations.

7.2.2. Biological depollution

7.2.2.1. Control and optimization of bioprocesses for depollution

Participants: Olivier Bernard, Francis Mairet, Jean-Luc Gouzé.

We have considered the problem of global stabilization of an unstable bioreactor model (e.g. for anaerobic digestion), when the measurements are discrete and in finite number ("quantized"). These measurements define regions in the state space, wherein a constant dilution rate is applied We show that this quantized control may lead to global stabilization: trajectories have to follow some transitions between the regions, until the final region where they converge toward the reference equilibrium [30].

Although bioprocesses involve an important biodiversity, the design of bioprocess control laws are generally based on single-species models. In [98], we have proposed to define and study the multispecies robustness of bioprocess control laws: given a control law designed for one species, what happens when two or more species are present? We have illustrated our approach with a control law which regulates substrate concentration using measurement of growth activity. Depending on the properties of the additional species, the control law can lead to the correct objective, but also to an undesired monospecies equilibrium point, coexistence, or even a failure point. Finally, we have shown that, for this case, the robustness can be improved by a saturation of the control.

Moreno [105] have proposed an optimal strategy for fed-batch bioreactor with substrate inhibition. Thanks to the Pontryagin maximum principle and the Hamilton-Jacobi equation, we have shown that the same strategy is still optimal when mortality is included in the model [75]. We have also studied the problem when the singular arc is non-necessary admissible everywhere (i.e. the singular control can take values outside the admissible control set). We have pointed out the existence of a frame point on the singular arc above which any singular trajectory is not globally optimal. Moreover, we have provided an explicit way for computing numerically the switching curves and the frame point [17].

7.2.2.2. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Francis Mairet.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Phycover project is aiming at evaluating the potential of this process [113], [112].

In a first stage, we developed models for anaerobic digestion of microalgae. Two approaches were used: first, a dynamic model has been developed trying to keep a low level of complexity so that it can be mathematically tractable for optimisation [97] . On the other hand, we have tested the ability of ADM1 [114] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [99] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

We have proposed and analysed a three dimensional model which represent the coupling of a culture of microalgae limited by light and an anaerobic digester. We first prove the existence and attraction of periodic solutions. Applying Pontryagin's Maximum Principle, we have characterized optimal controls, including the computation of singular controls, in order to maximize methane production. Finally, we have determined numerically optimal trajectories by direct and indirect methods [74].

Finally, we have studied the coupling between three ecosystems: an anaerobic digester, a wastewater treatment pond (with microalgae and nitrifiers) and a microalgal pond. Different possible coupling configurations were tested in simulation. A numerical optimization was carried out to identify, depending on the choice of the objective function (energy production, pollution removal) the optimal arrangement between the three processes. The optimal volume for each process was then determined. This work has been carried out in the framework of the Phycover ANR project and was the subject of the internship of Ignacio Lopez (Universidad de Chile).

7.2.2.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE (Laboratory of Environmental Biotechnology, Narbonne) and Pierre Collet (IFPEN).

In the sequel of the pioneering life cycle assessment (LCA) work of [94], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable.
The improvements due to technological breakthrough (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricity is produced on the plant [36]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage. This motivated the design of the purple sun ANR-project where electricity is produced by semi transparent photovoltaic panels [77] under which microalgae are growing.

Finally, some work are aiming at normalising LCA for microalgae and proposing guidelines to make the LCA more easily comparable [22].

These works have been carried out in collaboration with E. Latrille and B. Sialve (INRA-LBE).

7.2.3. Design of ecologically friendly plant production systems

7.2.3.1. Controlling plant pests

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Nicolas Bajeux.

Optimization of biological control agent introductions

The question of how many and how frequently natural enemies should be introduced into crops to most efficiently fight a pest species is an important issue of integrated pest management. The topic of optimization of natural enemies introductions has been investigated for several years [6] [108], unveiling the crucial influence of within-predator density dependent processes. Since parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, which are prevalent among predatory biocontrol agents, the current modeling effort consists in studying the impact of positive predator-predator interactions on the optimal introduction strategies (PhD of Nicolas Bajeux, [70], [71]).

The influence of the spatial structure of the environment on biological control efficacy has also been investigated; first results indicate that spatial structure tends to influence it in quite a same way as intra-specific competition does [27].

Connected research on the influence of space on the establishment of biological control agents is also being pursued both through computer simulations and laboratory experiments on parasitoids of the genus *Trichogramma*. This is the topic of the PhD thesis of Thibaut Morel Journel (UMR ISA) [13]; in particular, we showed how landscape connectivity or spatial heterogeneity shape establishment dynamics in spatially structured environments [33], [51], [40]. This research linked to invasion biology also led some of us to contribute with opinion or review contributions to a special issue on biological invasions, in connexion with the GdR Invabio [26], [32].

7.2.3.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Elsa Rousseau, Mélanie Bonneault.

Sustainable management of plant resistance

Because plants can get sick, 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. It is therefore necessary to find ways to protect the durability of such resistances, which are a natural exhaustible resource. We looked at landscape scale spatial deployment strategies of resistant crops able to maximize crop yield [25], allowing for the modification of the spatial arrangement of resistant crops over cropping seasons, showing dramatic increases in crop yield in particular epidemic situations [25].

Experiments were also conducted in INRA Avignon, followed by high-throughput sequencing (HTS) to identify the dynamics of virus strains competing within host plants. Different plant genotypes were chosen for their contrasted effects on genetic drift and selection they induce on virus populations. Those two evolutionary forces can play a substantial role on the durability of plant resistance. Therefore we fitted a mechanistic-statistical model to these HTS data in order to disentangle the relative role of genetic drift and selection during within-host virus evolution [53], [69], [43], [44]. A stochastic model was also produced to simulate the effect of drift on the virus epidemiological dynamics and on the durability of qualitative resistances [59]. This is the topic of Elsa Rousseau's PhD thesis, and is done in collaboration with Frédéric Fabre (INRA Bordeaux) and Benoît Moury (INRA Avignon).

We also developed an epidemiological model describing the dynamics of root-knot nematodes in a protected vegetable cropping system, to design optimal management strategies of crop resistance [110]. The model was fitted to experimental and field data. Preliminary results show that alternating susceptible and resistant crops not only increased the resistance durability, but reduced the disease intensity over time [63].

Finally we developed an epidemiological model including non-conventional gene-for-gene interactions in crops, based on the phoma stem canker of oilseed rape, to assess the durability of crop resistance in the field and design efficient deployment strategies [65]. This ongoing work is part of the K-Masstec project, which also incorporates experimental and field studies in collaboration with BIOGER (INRA Grignon).

Eco-evolutionary dynamics of plant pathogens in seasonal environments

Understanding better pathogen evolution also requires to understand how closely related plant parasites may coexist. Such coexistence is widespread and is hardly explained through resource specialization. We showed that, in agricultural systems in temperate environments, the seasonal character of agrosystems is an important force promoting evolutionary diversification of plant pathogens [89]. The plant parasites reproduction mode may also strongly interact with seasonality. In this context, we investigated the special case of oak powdery mildew, an oak disease which is actually caused by a complex of two different species, combining original plant epidemic data with the semi-discrete seasonal plant epidemic model we introduced a few years ago [50] [95]

This work has been done in collaboration with Frédéric Hamelin (Agrocampus Ouest) during Anne Bisson's internship, Marie Laure Desprez Loustau and Frederic Fabre (INRA Bordeaux).

7.2.3.2.1. Optimality/games in population dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Optimal foraging and residence times variations

A continued collaboration with Vincent Calcagno (UMR ISA) has yielded paper where we reanalyzed the so-called Marginal Value Theorem (MVT), first published in 1976 [80], [81]. Ongoing work aims at pointing how this latter theorem has been misused in some biological literature.

We also investigated the problem in foraging theory of evaluating the expected harvest of an animal when conspecifics may arrive on the same patch of resource in a stochastic fashion, specificaly according to a Poisson process or a Bernoulli process. A joint article with Frédéric Hamelin (Agrocampus Ouest) has been submitted for publication.

With Marc Deschamps, similar questions were studied in theoretical economy in the context of a Cournot competition on a single market. Again, an article has been submitted for publication.

The handicap paradox

We have investigated the question of "how could evolution have reached a state characterized by the handicap paradox?" with the tools of adaptive dynamics. We have reached the conclusion that, if one accepts adaptive dynamics as a model of evolution, and our model of sexual selection, the handicap paradox equilibrium is indeed the limit state of evolution [18].

This work was conducted with Frédéric Hamelin (Agrocampus Ouest).

CARMEN Team

7. New Results

7.1. Inverse Problem

Electrocardiograms simulated by our group with a highly realistic and detailed forward model were used for several inverse-modeling studies [34], [33], [38], [35].

- Stability analysis of the POD reduced order method for solving the bidomain model in cardiac electrophysiology: In this work we show the numerical stability of the Proper Orthogonal Decomposition (POD) reduced order method used in cardiac electrophysiology applications. The difficulty of proving the stability comes from the fact that we are interested in the bidomain model, which is a system of degenerate parabolic equations coupled to a system of ODEs representing the cell membrane electrical activity. The proof of the stability of this method is based an a priori estimate controlling the gap between the reduced order solution and the Galerkin finite element one. We present some numerical simulations confirming the theoretical results. We also combine the POD method with a time splitting scheme allowing a faster solution of the bidomain problem and show numerical results. Finally, we conduct numerical simulation in 2D illustrating the stability of the POD method in its sensitivity to the ionic model parameters. We also perform 3D simulation using a massively parallel code. We show the computational gain using the POD reduced order model. We also show that this method has a better scalability than the full finite element method.
- In silico assessment of drugs effects on human embryonic stem cells derived cardiomyocytes electrical activity: Computational modeling and simulation is extensively used to investigate diseases in cardiac electrophysiological activity and also drug effects, side effects and interactions. Human embryonic stem cell-derived cardiomyocytes (hESC-CMs) have been recently considered as a promising tool in regenerative medicine: their major role in repairing damaged tissue is due to pluripotency and ability to differentiate. These pluripotent cells are also used in early stages of drugs development. Pharmaceutical companies use the MultiElectrode Array (MEA) device in order to perform many in vitro experiments on hESC-CMs. The goal of our study is to derive a mathematical model and to simulate these in vitro experiments.
- Sensitivity of the Electrocardiography Inverse Solution to the Torso Conductivity Uncertainties: Electrocardiography imaging (ECGI) is a new non invasive technology used for heart diagnosis. It allows to construct the electrical potential on the heart surface only from measurement on the body surface and some geometrical informations of the torso. The purpose of this work is twofold: First, we propose a new formulation to calculate the distribution of the electric potential on the heart, from measurements on the torso surface. Second, we study the influence of the errors and uncertainties on the conductivity parameters, on the ECGI solution. We use an optimal control formulation for the mathematical formulation of the problem with a stochastic diffusion equation as a constraint. The descretization is done using stochastic Galerkin method allowing to separate random and deterministic variables. The optimal control problem is solved using a conjugate gradient method where the gradient of the cost function is computed with an ad-joint technique. The efficiency of this approach to solve the inverse problem and the usability to quantify the effect of conductivity uncertainties in the torso are demonstrated through a number of numerical simulations on a 2D geometrical model. Our results show that adding $\pm 50\%$ uncertainties in the fat conductivity does not alter the inverse solution, whereas adding $\pm 50\%$ uncertainties in the lung conductivity affects the reconstructed heart potential by almost 50%.
- Inverse Localization of Ischemia in a 3D Realistic Geometry: A Level Set Approach: The reconstruction of cardiac ischemic regions from body surface potential measurements (BSPMs) is usually performed at a single time instant which corresponds to the plateau or resting phase of the cardiac

action potential. Using a different approach, we previously proposed a level set formulation that incorporates the knowledge of the cardiac excitation process in the inverse procedure, thus exploiting the spatio-temporal correlations contained in the BSPMs. In this study, we extend our inverse level-set formulation for the reconstruction of ischemic regions to 3D realistic geometries, and analyze its performance in different noisy scenarios. Our method is benchmarked against zero-order Tikhonov regularization. The inverse reconstruction of the ischemic region is evaluated using the correlation coefficient (CC), the sensitive error ratio (SN), and the specificity error ratio (SP). Our algorithm outperforms zero-order Tikhonov regularization, specially in highly noisy scenarios.

- Inverse problem in electrocardography via the factorization method of boundary value problems: We present a new mathematical approach for solving the inverse problem in electrocardiography. This approach is based on the factorization of boundary value problems method. In this paper we derive the mathematical equations and test this method on synthetical data generated on realistic heart and torso geometries using the state-of-the-art bidomain model in the heart coupled to the Laplace equation in the torso. We measure the accuracy of the inverse solution using spatial Relative Error (RE) and Correlation Coefficient (CC).
- In the inverse problem en electrocardiology, the goal is to recover electrophysiological activity in the heart without measuring directly on its surface (without using catheter in- terventions). Note that today the inverse computation is frequently used by solving the quasi-static model. This model doesn't take into account the heart dynamic in time and may result in considerable errors in the reconstruction of the solution on the heart. In [1] we study a 3D numerical inverse problem constrained by the bidomain equations in electro- cardiology. The state equations consisting in a coupled reaction-diusion system modelling the propagation of the intracelullar and extracellular electrical potentials, and ionic cur- rents, are extended to further consider the eect of an external bathing medium. Thus, we demonstrate that the novel concept of applying electrophysiological data might be useful to improve noninvasive reconstruction of electrical heart activity. Finally, we present numerical experiments representing the eect of the heart dynamic on the inverse solutions. Moreover in [2], we study the stability result for the conductivities of the approximate bidomain model. The proof is based on the combination of a Carleman estimate obtained in [3] and certain weight energy estimates for parabolic systems.
- The static inverse ECG problem needs to solve the well known ill posed Cauchy problem for the Laplace equation. A new approach investigated in the team uses the method of factorization of bounadary value problems. This method, studied for itself, provides in this context the computation of Dirichlet-Neumann operators as solution of a Riccati equation. Results have been presented at IEEE international symposium on biomedical imaging, New York april 16-19, 2015. Further investigations will be lead using more precise numerical methods to solve the Riccati equation. The non-linearity and time dependence of the coupling resistance between cardiac cells (gap junctions) is stdied in the Liryc institute and thought to be of importance in the unerstanding of cardiac arrhythmias. The internship of Nhan Le Thanh was a first step to investigate their numerical simulation.

An Example of calculus on a torso is shown in figure 1

7.2. Cardiac Electromechanics

In [1] we study a coupled elliptic-parabolic system modeling the interaction between the propagation of electric potential and subsequent deformation of the cardiac tissue. The problem consists in a reaction-diusion system governing the dynamics of ionic quantities, intra and extra-cellular potentials, and the linearized elasticity equations are adopted to describe the motion of an incompressible material. The coupling between muscle contrac- tion, biochemical reactions and electric activity is introduced with a so-called active strain decomposition framework, where the material gradient of deformation is split into an active (electrophysiology-dependent) part and an elastic (passive) one. In this paper we prove exis- tence of weak solutions to the underlying coupled reaction-diusion system and uniqueness of regular solutions. We close with a numerical example illustrating the convergence of the method and some features of the model.

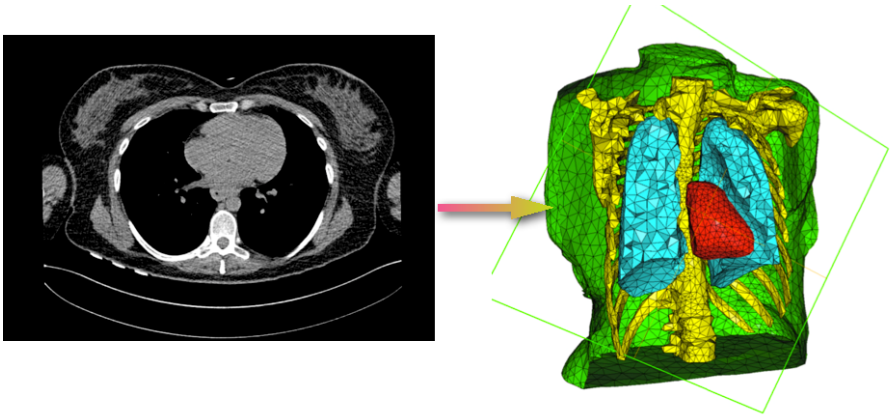


Figure 1. Mesh of the torso obtained with an IRM picture; this problem is link with the IDAM project

7.3. Cardiac Electrophysiology at the Microscopic Scale

We focused on establishing a microscopic model for cardiac electrophysiology simulations and proving the existence of a solution. We started with writing a mathematical proof allowing from well known physical equations and properties of the cardiac tissue to establish the model. Then, we worked on a variational formulation of the problem, and describing a weak solution of it. The idea is to compute energy estimates and to bound them so that we can extract a convergent sequence of functions in the appropriate Sobolev space. With my PhD advisor, we started to write an article about these two proofs. We also worked on CEPS code to implement some functionalities that will fit my requirements in a near future regarding the simulations we have to design. The main difficulty we identified is, provided we get a well defined geometry and mesh of cardiac cells, to implement the ionic flux between cells. First simulation of a simple "two-cells communication" problem will probably, if the results meet experimental observations, lead to another article. We also attended Imaged Based Biomedical Modelling 2015, a summer course organized by SCI institute (University of Utah), which was designed to give attendees guidelines about vizualisation and modelling, especially on cardio electrophysiology.

7.4. High order numerical scheme for ionic models

C. Douanla lonti worked on time numerical schemes like Admas-Bashforth in order to have a high degree of convergence between an exact solution and the approximated solution. This method is a generalisation of Rush-Larshen scheme adapted for electrophysiology cardiac.

DRACULA Project-Team

6. New Results

6.1. Implication of the autologous immune system in BCR-ABL transcript variations in chronic myelogenous leukemia patients treated with Imatinib

Imatinib (IM) and other tyrosine kinase inhibitors (TKI) have improved treatment of chronic myelogenous leukemia (CML); however, most patients are not cured. Deeper mechanistic understanding may improve TKI combination therapies to better control the residual leukemic cell population. In analyzing our patients' data, we found that many patients who otherwise responded well to IM therapy still showed variations in their BCR-ABL transcripts. To investigate this phenomenon, we applied a mathematical model (see [14]) that integrates CML and an autologous immune response to the patients' data. We define an immune window, or a range of leukemic loads for which the autologous immune system induces an improved response. Our modeling results in [14], suggest that, at diagnosis, a patient's leukemic load is able to partially or fully suppress the autologous immune response developed in a majority of patients, towards the CML clone(s). IM therapy drives the leukemic population into the "immune window", allowing the patient's autologous immune cells to expand and eventually mount an efficient recognition of the residual leukemic burden. This response drives the leukemic load below this immune window, allowing the leukemic population to partially recover until another weaker immune response is initiated. Thus, the autologous immune response may explain the oscillations in BCR-ABL transcripts regularly observed in patients on IM.

6.2. Predicting pathogen-specific CD8 T cell immune responses from a modeling approach

The primary CD8 T cell immune response constitutes a major mechanism to fight an infection by intra-cellular pathogens. We aim at assessing whether pathogen-specific dynamical parameters of the CD8 T cell response can be identified, based on measurements of CD8 T cell counts, using a modeling approach. We generated experimental data consisting in CD8 T cell counts kinetics during the response to three different live intra-cellular pathogens: two viruses (influenza, vaccinia) injected intranasally, and one bacteria (Listeria monocytogenes) injected intravenously. All pathogens harbor the same antigen (NP68), but differ in their interaction with the host. In parallel, we developed in [16] a mathematical model describing the evolution of CD8 T cell counts and pathogen amount during an immune response. This model is characterized by 9 parameters and includes relevant feedback controls. The model outputs were compared with the three data series and an exhaustive estimation of the parameter values was performed. By focusing on the ability of the model to fit experimental data and to produce a CD8 T cell population mainly composed of memory cells at the end of the response, critical parameters were identified. We show that a small number of parameters $(2-4)$ define the main features of the CD8 T cell immune response and are characteristic of a given pathogen. Among these parameters, two are related to the effector CD8 T cell mediated control of cell and pathogen death. The parameter associated with memory cell death is shown to play no relevant role during the main phases of the CD8 T cell response, yet it becomes essential when looking at the predictions of the model several months after the infection.

6.3. Dynamics of cell generation and turnover in the human heart

The contribution of cell generation to physiological heart growth and maintenance in humans has been difficult to establish and has remained controversial. We report in [8] that the full complement of cardiomyocytes is established perinataly and remains stable over the human lifespan, whereas the numbers of both endothelial and mesenchymal cells increase substantially from birth to early adulthood. Analysis of the integration of

nuclear bomb test-derived $14C$ revealed a high turnover rate of endothelial cells throughout life ($> 15\%$ per year) and more limited renewal of mesenchymal cells ($< 4\%$ per year in adulthood). Cardiomyocyte exchange is highest in early childhood and decreases gradually throughout life to $< 1\%$ per year in adulthood, with similar turnover rates in the major subdivisions of the myocardium. We provide an integrated model of cell generation and turnover in the human heart.

6.4. Travelling waves of cell differentiation

The paper [7] is devoted to modelling of cell differentiation in an initially homogeneous cell population. The mechanism which provides coexistence of two cell lineages in the initially homogeneous cell population is suggested. If cell differentiation is initiated locally in space in the population of undifferentiated cells, it can propagate as a travelling wave converting undifferentiated cells into differentiated ones. We suggest a model of this process which takes into account intracellular regulation, extracellular regulation and different cell types. They include undifferentiated cells and two types of differentiated cells. When a cell differentiates, its choice between two types of differentiated cells is determined by the concentrations of intracellular proteins. Differentiated cells can either stimulate differentiation into their own cell lineage or into another cell lineage. In the case of the positive feedback, only one lineage of differentiated cells will finally appear. In the case of negative feedback, both of them can coexist. In this case a periodic spatial pattern emerges behind the wave.

6.5. Pattern regeneration based on cell memory

In [24], we present a new model of the cellular dynamics that enable regeneration of complex biological morphologies. Biological cell structures are considered as an ensemble of mathematical points on the plane. Each cell produces a signal which propagates in space and is received by other cells. The total signal received by each cell forms a signal distribution defined on the cell structure. This distribution characterizes the geometry of the cell structure. If a part of this structure is removed, the remaining cells have two signals. They keep the value of the signal which they had before the amputation (memory), and they receive a new signal produced after the amputation. Regeneration of the cell structure is stimulated by the difference between the old and the new signals. It is stopped when the two signals coincide. The algorithm of regeneration contains certain rules which are essential for its functioning, being the first quantitative model of cellular memory that implements regeneration of complex patterns to a specific target morphology. Correct regeneration depends on the form and the size of the cell structure, as well as on some parameters of regeneration.

6.6. Target morphology and cell memory

Despite the growing body of work on molecular components required for regenerative repair, we still lack a deep understanding of the ability of some animal species to regenerate their appropriate complex anatomical structure following damage. A key question is how regenerating systems know when to stop growth and remodeling – what mechanisms implement recognition of correct morphology that signals a stop condition? In [11], we review two conceptual models of pattern regeneration that implement a kind of pattern memory. In the first one, all cells communicate with each other and keep the value of the total signal received from the other cells. If a part of the pattern is amputated, the signal distribution changes. The difference from the original signal distribution stimulates cell proliferation and leads to pattern regeneration, in effect implementing an error minimization process that uses signaling memory to achieve pattern correction. In the second model, we consider a more complex pattern organization with different cell types. Each tissue contains a central (coordinator) cell that controls the tissue and communicates with the other central cells. Each of them keeps memory about the signals received from other central cells. The values of these signals depend on the mutual cell location, and the memory allows regeneration of the structure when it is modified. The purpose of these models is to suggest possible mechanisms of pattern regeneration operating on the basis of cell memory which are compatible with diverse molecular implementation mechanisms within specific organisms.

6.7. Transplanted bone marrow-derived cells contribute to human adipogenesis

Because human white adipocytes display a high turnover throughout adulthood, a continuous supply of precursor cells is required to maintain adipogenesis. Bone marrow (BM)-derived progenitor cells may contribute to mammalian adipogenesis; however, results in animal models are conflicting. In [22], we demonstrate in 65 subjects who underwent allogeneic BM or peripheral blood stem cell (PBSC) transplantation that, over the entire lifespan, BM/PBSC-derived progenitor cells contribute 10% to the subcutaneous adipocyte population. While this is independent of gender, age, and different transplantation-related parameters, body fat mass exerts a strong influence, with up to 2.5-fold increased donor cell contribution in obese individuals. Exome and whole-genome sequencing of single adipocytes suggests that BM/PBSC-derived progenitors contribute to adipose tissue via both differentiation and cell fusion. Thus, at least in the setting of transplantation, BM serves as a reservoir for adipocyte progenitors, particularly in obese subjects.

6.8. Modelling of platelet–fibrin clot formation in flow

The paper [23] is devoted to mathematical modelling of clot growth in blood flow. Great complexity of the hemostatic system dictates the need of usage of the mathematical models to understand its functioning in the normal and especially in pathological situations. In this work we investigate the interaction of blood flow, platelet aggregation and plasma coagulation. We develop a hybrid DPD–PDE model where dissipative particle dynamics (DPD) is used to model plasma flow and platelets, while the regulatory network of plasma coagulation is described by a system of partial differential equations. Modelling results confirm the potency of the scenario of clot growth where at the first stage of clot formation platelets form an aggregate due to weak inter-platelet connections and then due to their activation. This enables the formation of the fibrin net in the centre of the platelet aggregate where the flow velocity is significantly reduced. The fibrin net reinforces the clot and allows its further growth. When the clot becomes sufficiently large, it stops growing due to the narrowed vessel and the increase of flow shear rate at the surface of the clot. Its outer part is detached by the flow revealing the inner part covered by fibrin. This fibrin cap does not allow new platelets to attach at the high shear rate, and the clot stops growing. Dependence of the final clot size on wall shear rate and on other parameters is studied.

6.9. Conceptual model of morphogenesis and regeneration

The paper [24] is devoted to computer modelling of the development and regeneration of multicellular biological structures. Some species (e.g. planaria and salamanders) are able to regenerate parts of their body after amputation damage, but the global rules governing cooperative cell behaviour during morphogenesis are not known. Here, we consider a simplified model organism, which consists of tissues formed around special cells that can be interpreted as stemcells. We assume that stem cells communicate with each other by a set of signals, and that the values of these signals depend on the distance between cells. Thus the signal distribution characterizes location of stem cells. If the signal distribution is changed, then the difference between the initial and the current signal distribution affects the behaviour of stem cells—e.g. as a result of an amputation of a part of tissue the signal distribution changes which stimulates stem cells to migrate to new locations, appropriate for regeneration of the proper pattern. Moreover, as stem cells divide and form tissues around them, they control the form and the size of regenerating tissues. This two-level organization of the model organism, with global regulation of stem cells and local regulation of tissues, allows its reproducible development and regeneration.

6.10. Delay differential-difference system for hematopoietic stem cell dynamics

We investigate in [2] and [3] a mathematical model of hematopoietic stem cell dynamics. We take two cell populations into account, quiescent and proliferating one, and we note the difference between dividing cells that enter directly to the quiescent phase and dividing cells that return to the proliferating phase to divide again. The resulting mathematical model is a system of two age-structured partial differential equations. By

integrating this system over age and using the characteristics method, we reduce it to a delay differential-difference system, and we investigate the existence and stability of the steady states. We give sufficient conditions for boundedness and unboundedness properties for the solutions of this system. By constructing a Lyapunov function, the trivial steady state, describing cell's dying out, is proven to be globally asymptotically stable when it is the only equilibrium. The stability analysis of the unique positive steady state, the most biologically meaningful one, and the existence of a Hopf bifurcation allow the determination of a stability area, which is related to a delay-dependent characteristic equation. Numerical simulations illustrate our results on the asymptotic behavior of the steady states and show very rich dynamics of this model. This study may be helpful in understanding the uncontrolled proliferation of blood cells in some hematological disorders.

6.11. Discrete limit and monotonicity properties of the Floquet eigenvalue in an age structured cell division cycle model

We consider in [19] a cell population described by an age-structured partial differential equation with time periodic coefficients. We assume that division only occurs after a minimal age (majority) and within certain time intervals. We study the asymptotic behavior of the dominant Floquet eigenvalue, or Perron-Frobenius eigenvalue, representing the growth rate, as a function of the majority age, when the division rate tends to infinity (divisions become instantaneous). We show that the dominant Floquet eigenvalue converges to a staircase function with an infinite number of steps, determined by a discrete dynamical system. As an intermediate result, we give a structural condition which guarantees that the dominant Floquet eigenvalue is a nondecreasing function of the division rate. We also give a counter example showing that the latter monotonicity property does not hold in general.

6.12. Optimal linear stability condition for scalar differential equations with distributed delay

Linear scalar differential equations with distributed delays appear in the study of the local stability of nonlinear differential equations with feedback, which are common in biology and physics. Negative feedback loops tend to promote oscillations around steady states, and their stability depends on the particular shape of the delay distribution. Since in applications the mean delay is often the only reliable information available about the distribution, it is desirable to find conditions for stability that are independent from the shape of the distribution. We show in [9] that for a given mean delay, the linear equation with distributed delay is asymptotically stable if the associated differential equation with a discrete delay is asymptotically stable. We illustrate this criterion on a compartment model of hematopoietic cell dynamics to obtain sufficient conditions for stability.

6.13. A mathematical model of leptin resistance

Obesity is often associated with leptin resistance, which leads to a physiological system with high leptin concentration but unable to respond to leptin signals and to regulate food intake. We propose in [20] a mathematical model of the leptin-leptin receptors system, based on the assumption that leptin is a regulator of its own receptor activity, and investigate its qualitative behavior. Based on current knowledge and previous models developed for body weight dynamics in rodents, the model includes the dynamics of leptin, leptin receptors and the regulation of food intake and body weight. It displays two stable equilibria, one representing a healthy state and the other one an obese and leptin resistant state. We show that a constant leptin injection can lead to leptin resistance and that a temporal variation in some parameter values influencing food intake can induce a change of equilibrium and a pathway to leptin resistance and obesity.

M3DISIM Team

7. New Results

7.1. Modeling

7.1.1. Model-based analysis of continually measured signals of aortic pressure and flow

Participants: Radomir Chabiniok, Dominique Chapelle [correspondant], Arthur Le Gall, Philippe Moireau, Fabrice Vallée.

We have started an application of reduced-order cardiac modeling in identifying relevant functional properties and state of heart from clinical records obtained during long-term (minutes-hours) monitoring of patients. Those are obtained either from anesthetized or intensive care patients by Fabrice Vallée, medical doctor in the department of anesthesia and intensive care at Lariboisière Hospital, Paris, who has joined the M3DISIM team in November 2015. The collaboration was initiated already in February 2015, and together with Fabrice we supervised the master's internship of Arthur Le Gall (medical doctor in his last year of specialization residency training). The internship took place at Lariboisière Hospital and in our lab at Inria Saclay (50:50%, period of April-September 2015). First published results are expected in 2016, when also a master's internship of a second student of Fabrice Vallée is scheduled. In addition, we intend to start a PhD on this topic in late 2016.

7.1.2. Thermodynamical framework for modeling chemical-mechanical coupling in muscle contraction – Formulation and validation

Participants: Matthieu Caruel, Dominique Chapelle [correspondant], Philippe Moireau.

Muscle contraction occurs at the nanoscale of a hierarchical multi-scale structure with the attachment of so-called cross-bridges within sarcomeres, namely, the creation of chemical bonds between myosin heads and specific sites on actin filaments. A cross-bridge in itself can be seen as a special chemical entity having internal mechanical variables – or degrees of freedom – pertaining to the actual geometric configuration, which implies that the free energy of the cross-bridge – whether in an attached or unattached state – must be made dependent on these internal variables (T.L. Hill, Free Energy Transduction And Biochemical Cycle Kinetics, Dover, 2004). This provides a thermodynamical basis for modeling the complex interplay of chemical and mechanical phenomena at the sarcomere level. Within this framework we propose a muscle model with two mechanical variables associated with a cross-bridge. For the action of individual cross-bridges occurring at the nanometer scale, the energy provided by the Langevin thermostat cannot be neglected, and we therefore propose to endow the internal mechanical variables with stochastic dynamics. Important motivations for this modeling choice include the ability to represent (i) the so-called power-stroke phenomenon and (ii) short-time responses of a muscle, e.g. to load steps. Our approach allows for systematic treatment of the model energetics, and in particular one goal of the proposed description is to investigate the potential benefit in mechanical efficiency with systems including – in addition to chemically-induced transformations – thermally-induced conformational changes such as the power-stroke.

7.1.3. Biophysical modeling of seismocardiograms measurements

Participants: Alexandre Laurin, Sébastien Imperiale [correspondant], Philippe Moireau, Dominique Chapelle.

We are developing models of various levels of complexity to represent seismocardiograms (SCG) that record mechanical thoracic vibrations induced by the beating heart. Our model combines a complete heartbeat model with a mechanical model of the thorax. The coupling is ensured by a unilateral contact modeling the non-penetration between the beating heart and the thoracic chest. In parallel, we are fine-tuning signal processing algorithms to identify the relevant characteristics of SCG and creating an iPhone application that is capable of acquiring the signal with its standard sensors. The application is also developed to integrate a simplified version of the cardio-thoracic model.

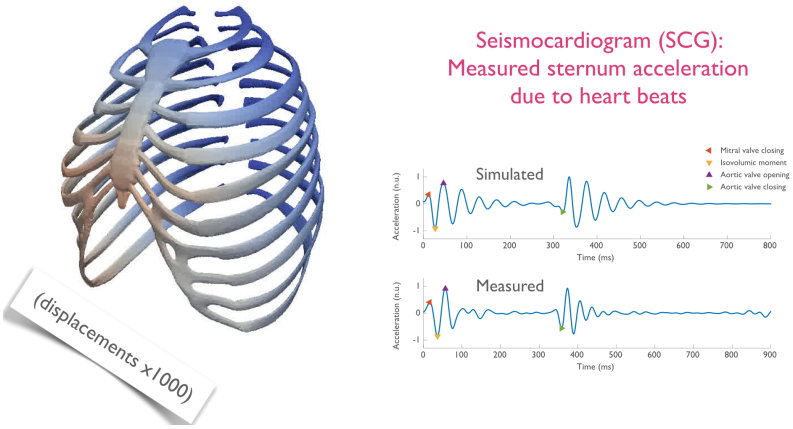


Figure 1. Model-based synthetic SCG compared to a measured SCG.

7.2. Numerical Analysis

7.2.1. Dirichlet-to-Neumann operator for diffraction problems in stratified anisotropic acoustic waveguides

Participants: Antoine Tonnoir [correspondant], Sonia Fliss [Poems team], Anne-Sophie Bonnet-Ben Dhia [Poems team].

In this work, we are interested in the construction of a Dirichlet-to-Neumann operator for the diffraction problem in stratified anisotropic acoustic waveguides. The key idea consists in using an adapted change of coordinates that allows to recover the completeness and the orthogonality of the modes on "deformed" cross-sections of the waveguide. Thus, we can properly define the diffraction problem and construct transparent boundary conditions to reformulate this problem in a bounded domain. Using classical arguments we easily prove the well-posedness. The method has also been implemented in a `C++` code and has been validated.

7.2.2. Fourth-order energy-preserving locally implicit time discretization for linear wave equations

Participants: Sébastien Imperiale [correspondant], Juliette Chabassier [MAGIQUE-3D team].

In collaboration with Juliette Chabassier, we have constructed a family of fourth-order implicit-explicit time schemes for linear wave equations. Our application is the simulation of elastic waves propagation in a locally stiff medium. The domain of propagation is decomposed into several regions where different fourth-order time discretization are used, chosen among a family of implicit (for the stiff regions) or explicit fourth-order schemes. The coupling is based on a Lagrangian formulation on the boundaries between several non-conforming meshes of the regions. A global discrete energy is shown to be preserved and leads to global fourth-order consistency in time. Numerical results in 1D and 2D illustrate the good behavior of the schemes and their potential for the efficient simulation of realistic highly heterogeneous media for which using an explicit scheme everywhere can be extremely penalizing. Accuracy up to fourth-order reduces the numerical dispersion inherent to implicit methods used with a large time step, and makes this family of schemes attractive compared to second-order accurate methods.

7.2.3. Numerical methods for poromechanics: Applications to cardiac perfusion

Participants: Bruno Burtschell, Dominique Chapelle [correspondant], Philippe Moireau.

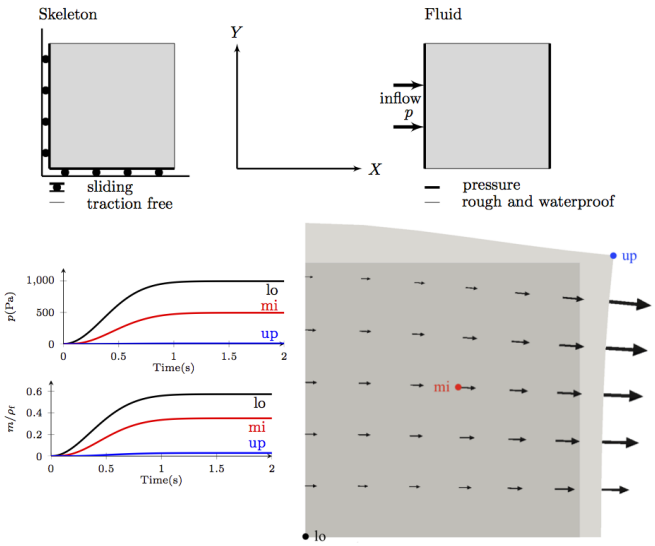


Figure 2. Swelling of a porous sample. On the left, fluid pressure and mass at three points in time during the swelling test. On the right, the fluid velocity field vector at the steady state in the deformed configuration.

We have previously formulated a rather general modeling framework of poromechanics – formulations that combine solid and fluid components to represent the behavior of a porous medium – to take into account large deformations and rapid fluid flows, see [6]. This allows to consider, in particular, the application of blood perfusion within the cardiac tissue, which features these specific complex phenomena, out of the scope of classical poromechanical models. One of our major objectives now, within the PhD of Bruno Burtschell, is to propose and analyse some associated relevant numerical schemes.

Some existing algorithms of fluid-structure interaction, with which our poromechanics formulations feature deep similarities, have been implemented – in FreeFEM++, both in axisymmetric configuration and in 3D – and compared. Their numerical and theoretical analysis – consistency, convergence – has been performed. Then, the adaptation of these algorithms to our poromechanics formulations enabled us to propose a time discretisation well-fitted to our framework, and to present its energy stability analysis. Spatial discretizations issues have also been specifically addressed, based on a complete analysis performed on a linearized problem, in order to guarantee pressure stability – via the selection of adequate inf-sup-compatible discretization spaces – including when the solid constituent is nearly or fully incompressible. Implementation and detailed numerical validations of these schemes have ben performed. Integration into FELISCE ("HappyHeart" module) in 3D, and into a reduced model of cardiac cycle to take into account myocardium perfusion, are ongoing work.

7.3. Model-Data Interaction

7.3.1. Displacement Reconstructions in Ultrasound Elastography

Participant: Sébastien Imperiale.

In collaboration with Guillaume Bal (Columbia University, New York, USA), we have considered the reconstruction of internal elastic displacements from ultrasound measurements, which finds applications in the medical imaging modality called elastography. By appropriate interferometry and windowed Fourier transforms of the ultrasound measurements, we have proposed a reconstruction procedure of the vectorial

structure of spatially varying elastic displacements in biological tissues. This provides a modeling and generalization of scalar reconstruction procedures routinely used in elastography. The proposed algorithm has been justified using a single scattering approximation and local asymptotic analysis. Its validity has been assessed by numerical simulations.

7.3.2. Recursive joint state and parameter estimation

Participants: Atte Aalto, Philippe Moireau [correspondant].

We propose a method for estimating the parameters of a linear dynamical system from noisy measurements over a given, finite time, interval. For this purpose we develop a recursive modification of the joint state and parameter estimation method proposed in [7]. As the time interval is fixed, any errors in the initial state of the system may cause a significant error in the parameter estimate. Therefore, the parameter estimator is complemented by the so called back and forth nudging (BFN) method for estimating the system's initial state. The proposed strategy can also be regarded as a hybrid least squares optimization method for minimizing the quadratic discrepancy between the measured and simulated outputs over the set of all possible initial states and system parameters.

The optimality of the BFN method with colocated feedback has been considered as well. We have shown that in the case when the system's dynamics are governed by a skew-adjoint generator, the initial state estimate given by the BFN method converges to the minimizer of the quadratic output discrepancy – provided that the observer gains are chosen suitably. If the system's generator is essentially skew-adjoint and dissipative, a certain modification of the feedback operator is required in order to obtain such convergence.

7.3.3. Convergence of discrete-time Kalman filter estimate to continuous-time estimate

Participant: Atte Aalto [correspondant].

The Kalman(-Bucy) filter gives the optimal (minimum variance) solution to the state estimation problem for linear systems with Gaussian initial state, and white input and output noise processes. The implementation of the discrete-time Kalman filter is straightforward as it is readily formulated in an algorithmic manner. Thus, it may be tempting to use the discrete-time filter on the time-sampled continuous-time system. We study the convergence of the state estimate obtained from the discrete-time Kalman filter to the continuous-time estimate as the temporal discretization is refined. The convergence follows from the martingale convergence theorem, but surprisingly, no results exist on the rate of convergence. We derive convergence rate estimates for the discrete-time estimate under a number of different sets of assumptions starting from finite-dimensional systems and infinite-dimensional systems with bounded output operators and then proceeding to systems with unbounded output operators and systems with analytic semigroups. The proofs are based on applying the discrete-time Kalman filter on a dense, numerable subset of the time interval of interest, and bounding the change in the state estimate as the new data points are being added. These bounds, in turn, are based on smoothness estimates of the noise-free output.

7.3.4. Observers for the wave equation in unbounded domains

Participants: Sébastien Imperiale, Philippe Moireau [correspondant], Antoine Tonnoir, Sonia Fliss [Poems team], Karim Ramdani [Sphinx team].

We are interested in the reconstruction of initial data for the wave equation problem in unbounded domains using an observer strategy. A major advantage of this method for problems set in bounded domains is the exponential convergence of the algorithm of reconstruction. In our case, the specificity is the unboundedness of the domain which requires to bound it with artificial boundaries for numerical computations. To avoid spurious reflections due to these artificial boundaries, we consider transparent boundary conditions. The difficulty then is to adapt the classical observers technique to this case. Indeed, after enough time, the outgoing waves have left the computational domain and the related information is in some sense "lost".

First results have been obtained for the 1D case: the theoretical proof of the (exponential) convergence of the algorithm has been done, and the method has been numerically validated. We are currently working on the extension to the 2D case, which raises new difficulties. In particular, the construction of the transparent boundary condition is not obvious and implies a non-local operator in both time and space. Due to this non-local operator, the theoretical analysis of the convergence of the method is then much more difficult.

7.3.5. A Luenberger observer for reaction-diffusion models with front position data

Participants: Dominique Chapelle, Annabelle Collin, Philippe Moireau [correspondant].

We propose a Luenberger observer for reaction-diffusion models with propagating front features, and for data associated with the location of the front over time. Such models are considered in various application fields, such as electrophysiology, wild-land fire propagation and tumor growth modeling. Drawing our inspiration from image processing methods by considering a data similarity measure of Mumford-Shah type, we start by proposing an observer for the eikonal-curvature equation that can be derived from the reaction-diffusion model by an asymptotic expansion. We then carry over this observer to the underlying reaction-diffusion equation by an "inverse asymptotic analysis", and we show that the associated correction in the dynamics has a stabilizing effect for the linearized estimation error. We also discuss the extension to joint state-parameter estimation by using the earlier-proposed ROUKF strategy. We published a first work [17] where the observer feedback is derived from the shape-derivative of the data similarity measure. Then, in [21], in order to improve the observer formulation, we followed a strategy of increasing importance in shape optimization or "level-set"-based image segmentation by complementing the required shape derivatives, used to modify the shape contours, by a topological derivative that represents the sensitivity of the similarity measure when removing a small part of the domain. Both results are illustrated with test problems pertaining to electrophysiology modeling, including with a realistic model of cardiac atria. Our numerical trials show that state estimation is directly very effective with the proposed Luenberger observer.

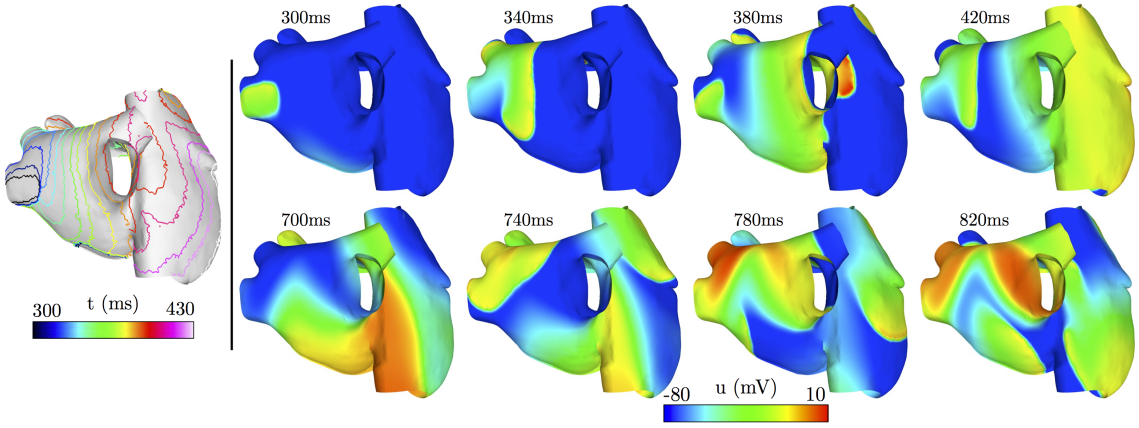


Figure 3. Atrial fibrillation: synthetic front data with noise (left, note that only the first times of passage after the onset of fibrillation are displayed, for the sake of clarity), and corresponding observer solutions (right)

7.3.6. Identification of weakly coupled multiphysics problems. Application to the inverse problem of electrocardiography

Participants: Cesare Corrado [Reo team], Jean-Frédéric Gerbeau [Reo team], Philippe Moireau [correspondant].

This work addresses the inverse problem of electrocardiography from a new perspective, by combining electrical and mechanical measurements. Our strategy relies on the definition of a model of the electromechanical contraction which is registered on ECG data, but also on measured mechanical displacements of the heart tissue typically extracted from medical images. In this respect, we establish in this work the convergence of a sequential estimator which combines for such coupled problems various state-of-the-art sequential data assimilation methods in a unified consistent and efficient framework. Indeed, we aggregate a Luenberger observer for the mechanical state and a Reduced-Order Unscented Kalman Filter applied on the parameters to be identified and a POD projection of the electrical state. Then, using synthetic data we show the benefits of our approach for the estimation of the electrical state of the ventricles along the heart beat, compared with more classical strategies that only consider an electrophysiological model with ECG measurements. Our numerical results actually show that the mechanical measurements improve the identifiability of the electrical problem, allowing to reconstruct the electrical state of the coupled system more precisely. Therefore, this work is intended to be a first proof of concept, with theoretical justifications and numerical investigations, of the advantage of using available multi-modal observations for the estimation and identification of an electromechanical model of the heart.

7.3.7. Data assimilation of cine-MR images by a biophysical model

Participants: Radomir Chabiniok, Dominique Chapelle [correspondant], Alexandra Groth, Philippe Moireau, Juergen Weese.

Within the European project VP2HF, we participated in extending the image segmentation tool developed by Philips Hamburg (Alexandra Groth, Jürgen Weese) to process clinically routine cine-MR images for creating anatomical models of heart. Secondly, together with A. Groth and J. Weese we defined a discrepancy operator – between a biomechanical heart model and cine-MR images – that does not require segmenting MR images prior to data assimilation. Initial results of the state estimation using this discrepancy operator were presented at the 2nd VP2HF evaluation meeting (December 2015), and extending these results into a journal paper is a joint objective of the M3DISIM team and of Philips Hamburg.

MAMBA Project-Team

7. New Results

7.1. Cancer

Participants: Luís Lopes Neves de Almeida, Rebecca Chisholm, Jean Clairambault, François Delhommeau [Haematology department, St Antoine Hospital, Paris], Dirk Drasdo, Ján Eliaš, Alexandre Escargueil [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Ghassen Haddad [ENIT, Tunis], Shalla Hanson [Department of mathematics, Duke University, Durham, NC], Pierre Hirsch [Haematology department, St Antoine Hospital, Paris], Groups Invade, Lungsysii, Tim Johann, Group Klingmueller [German Cancer Center, Heidelberg], Michal Kowalczyk [Univ. Santiago de Chile], Annette Larsen [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Tommaso Lorenzi, Alexander Lorz, Benoît Perthame, Andrada Quillas Maran, Fernando Quirós [Univ. Autónoma de Madrid], Michèle Sabbah [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Min Tang [Jiaotong University, Shanghai], Emmanuel Trélat [LJLL, UPMC], Paul Van Liedekerke, Nicolas Vauchelet, Irène Vignon-Clementel [REO], Yi Yin.

7.1.1. Drug resistance

We have continued to develop our phenotypically based models of drug-induced drug resistance in cancer cell populations, representing their Darwinian or Lamarckian evolution under drug pressure by integro-differential equations. In one of them [23], a 1D space variable has been added to the phenotypic structure variable to account for drug diffusion in tumour spheroids. In another one, focusing on both Darwinian selection and Lamarckian-like (non-genetic) instruction, published in *Cancer Research* [41], where deterministic and agent-based modelling are processed in parallel, we have added advection and diffusion terms to the initial integro-differential model and considered a physiologically based 2-dimensional phenotypic structure variable. This model, designed to take account of previously published biological observations on (reversible) drug tolerance persistence in a cultured population of non-small cell lung cancer (NSCLC) cells [90], reproduces the observations and we propose to assess the model by testing biologically based hypotheses. This work, also presented in various conferences ([34], [35], [31]) is conducted in close collaboration with the INSERM-UPMC team "Cancer biology and therapeutics" (A. Larsen, A. Escargueil, M. Sabbah) at St Antoine Hospital. It has also led our postdoctoral fellows Rebecca Chisholm and Tommaso Lorenzi to prolong their work on the *Cancer Research* paper by publishing two more articles [21], [48], one of which is a joint work with Alexander Lorz. This work is currently continued from the point of view of optimal control in Camille Pouchol's PhD thesis.

7.1.2. Evolution and cancer, therapy optimisation

Guided by our goal to understand and overcome drug resistance in cancer cell populations[41], we are considering cancer as an evolutionary phenomenon at two time scales: a large time scale (billions of years) of evolution of the genomes, from unicellular organisms to organised multicellularity (viewing cancer as more an archeoplasm than a neoplasm, an evolution backwards, following Davies and Lineweaver, Phys Biol 2011, and others [78], [66], [92], [79]) with shortcomings due to malfunctions in the processes of control of cell differentiation, and a short time scale (duration of a human life) of evolution in the "epigenetic landscape" of a given genome (as advocated by Sui Huang and Angela Pisco, e.g. recently in Nature, Br J Cancer and elsewhere [76], [77], [85], [86], [94]). It leads us to propose theoretical frameworks for innovative cancer therapeutics from this evolutionary biology viewpoint, taking into account the major clinical issue of drug resistance in cancer cell populations, as presented in [31] and exposed to a medical audience at the symposium "Réseau Cancer des Points Cardinaux" (http://www.frog-oncogeriatrie.com/fichiers/evnmt_41.pdf).

7.1.3. Interactions between tumour cell populations and their cellular micro-environments

A phenotype-structured model of the interactions between a brest cancer cell population (MCF7 cultured cells, collaboration with M. Sabbah, St Antoine Hospital) and its adipocyte stroma support cell population has been developed (T. Lorenzi, C. Pouchol, J. Clairambault) in the framework of Camille Pouchol's Inria internship ([56]). It has led to hiring C. Pouchol as a PhD student at UPMC (on a university grant "Interfaces pour le Vivant") on the same subject with perspectives in optimal therapeutic control, under the supervision of J. Clairambault, M. Sabbah and E. Trélat, see below "Supervision".

7.1.4. Combining chemo- and immunotherapies

Both from the point of view of interactions with the tumour micro-environment and of innovative anticancer therapies, it is necessary to take into account the immune response in cancer. This recently developed activity, (illustrated by presentations in session 70 in ICNAAM 2016 [32]) has led to the involvement in 2015 of Shalla Hanson as a PhD student in co-tutela between Duke University, NC and UPMC, see below "Supervision".

7.1.5. Hele-Shaw model of tumour growth

The mathematical analysis of macroscopic models of tumor growth with one type of cancer cells has been continued. On the one hand, in [47], the concept of viscosity solutions has been implemented for the case with active motion. On the other hand, the regularity of the free boundary is proved in [51] using methods developed for the standard Hele-Shaw equation and a new formulation.

7.1.6. The p53 protein spatio-temporal dynamics

Our previously developed spatio-temporal models for an intracellular dynamical response of the p53 protein to DNA damage, have been exploited further, and several testable biological hypotheses have been proposed in [33]. Among them, we suggest ideas that link spatio-temporal location of the p53 protein with a specific cell fate of a single cell in [33], [2] and, based on our new oscillator relying on both positive and negative regulation of p53 by Mdm2 (in tight cooperation with MdmX), we provide molecular insights into an excitability of the p53 network, i.e., we propose a molecular explanation for a full pulsatile response of p53 independently of input (ATM) signalling, challenging thus different fates of ATM downstream targets in the regulation of p53 in response to different stimuli, such as γ- and UV-radiation.

Our mathematical models, all included in J. Eliaš's PhD thesis [2] (defended on the 1st September 2015), contribute to understanding the variability of p53 in response to single and double strand breaks and reveal some new aspects of the core p53-Mdm2 protein feedback.

7.1.7. Lung and breast cancer

We developed an image analysis software and designed image analysis pipelines which we used to quantify the invasion pattern of non-small cell lung cancer (NSCLC) cells in multicellular spheroid *in vitro* experiments [24]. Based on the analyses, we demonstrated that the concomitant over-expression of FIR (far upstream element binding protein interacting repressor) and its splice variants drives NSCLC migration and dissemination.

We developed an agent-based, centre-based model of cell migration in cancer invasion based upon experimental observations of cell shape and cell behaviour in multicellular spheroid experiments of breast and lung cancer cells. In these experiments, cells deform from a sphere into an oblong shape upon migration, and adopt a spherical shape again whenever they turn back to such spheroids. This was implemented. Moreover, we developed a 3D model for the extracellular matrix (ECM) in which the matrix is modelled by an irregular network of springs with nodes represented as elastic objects. Migrating cells anchor in the network to move, leading to network deformation. We implemented a number of different biological mechanisms of cell migration and cell-ECM interaction. We find that a relatively simple model is sufficient to explain all phenomena of a single invading cell (Palm et. al., in preparation).

The combination of image analysis and of the abovementioned refined invasion model should allow a quantitative model of multicellular invasion following the same line of research as for SK-MES-1 cells, where we inferred a multicellular spheroid growth model from image data within a pipeline of experiment, imaging, image analysis and modelling [17]. In that paper, we used spatial-temporal image data of cell nucleus distribution, cell proliferation, death, and ECM distribution for two growth conditions (oxygen and glucose) to calibrate a model which was then able to quantitatively correctly predict the growth kinetics of the tumor spheroids for two other growth conditions, one strongly glucose limited, another strongly oxygen-limited.

Finally, we developed an image analysis pipeline to estimate the number of cancer cells in a patient with non-small cell lung cancer (NSCLC) from non-invasive image modalities. The estimate bases upon cell counts from histological serial sections of the tumor which have been related to the D-value inferred from Diffusion Weighted (DW) MRI (Yi et. al., paper in preparation).

7.2. Aggregation Kinetics

Participants: Aurora Armiento, Tom Banks [CRSC, NCSU, Raleigh, USA], Thibault Bourgeron, José Antonio Carrillo [Imperial College, London, United Kingdom], Marie Doumic, Miguel Escobedo [Universidad del País Vasco, Bilbao, Spain], Sarah Eugène, Marc Hoffmann [Ceremade, Université Paris-Dauphine], François James [MAPMO, Université d'Orléans], Nathalie Krell [Université de Rennes 1], Carola Kruse, Frédéric Lagoutière [Département de mathématiques d'Orsay], Philippe Moireau [Inria Paris Saclay, M3DISIM project-team], Benoît Perthame, Stéphanie Prigent, Human Rezaei [VIM, INRA Jouy-en-Josas], Lydia Robert [Laboratoire Jean Perrin, UPMC], Philippe Robert [Inria Paris, RAP project-team], Maria Teresa Teixeira [IBCP, Paris], Nicolas Vauchelet, Min Tang [Jiaotong University, Shanghai], Zhou Xu [IBCP, Paris], Wei-Feng Xue [University of Kent, United Kingdom].

7.2.1. Heterogeneity as an intrinsic feature in biological dynamics

Combining deterministic and probabilistic approaches, we investigated in two different applications - namely senescence and protein aggregation - the impact of heterogeneity on dynamical features of the considered populations.

Yeast Senescence and Telomere replication In eukaryotes, the absence of telomerase results in telomere shortening, eventually leading to replicative senescence, an arrested state that prevents further cell divisions. While replicative senescence is mainly controlled by telomere length, the heterogeneity of its onset is not well understood. Insights on this key question may have consequences both for cancer and aging issues.

In collaboration with T. Teixeira and Z. Xue from IBCP, we proposed a mathematical model based on the molecular mechanisms of telomere replication and shortening to decipher the causes of this heterogeneity [7]. Using simulations fitted on experimental data obtained from individual lineages of senescent Saccharomyces cerevisiae cells, we decompose the sources of senescence heterogeneity into interclonal and intraclonal components, and show that the latter is based on the asymmetry of the telomere replication mechanism. We also evidence telomere rank-switching events with distinct frequencies in short-lived versus long-lived lineages, revealing that telomere shortening dynamics display important variations. Thus, the intrinsic heterogeneity of replicative senescence and its consequences find their roots in the asymmetric structure of telomeres.

These promising first results lead us to an ongoing collaboration, and hopefully will allow still more insight on complex mechanisms not yet modelled mathematically.

Variability in nucleated polymerisation

The kinetics of amyloid assembly show an exponential growth phase preceded by a lag phase, variable in duration as seen in bulk experiments and experiments that mimic the small volumes of cells. To investigate the origins and the properties of the observed variability in the lag phase of amyloid assembly currently not accounted for by deterministic nucleation dependent mechanisms, we formulated a new stochastic minimal model that is capable of describing the characteristics of amyloid growth curves despite its simplicity [44]. We then solved the stochastic differential equations of our model and gave a mathematical proof of a central limit theorem for the sample growth trajectories of the nucleated aggregation process. These results

give an asymptotic description for our simple model, from which closed-form analytical results capable of describing and predicting the variability of nucleated amyloid assembly were derived. We also demonstrated the application of our results to inform experiments in a convenient and clear way. Our model offers a new perspective and paves the way for a new and efficient approach on extracting vital information regarding the key initial events of amyloid formation.

7.2.2. *Inverse Problems and Data Assimilation Applied to Protein Aggregation and other settings*

As mathematical models become more complex with multiple states and many parameters to be estimated using experimental data, there is a need for critical analysis in model validation related to the reliability of parameter estimates obtained in model fitting. This leads to a fundamental question: how much information with respect to model validation can be expected in a given data set or collection of data sets?

In the biological context of amyloid formation, the question is to quantify to which extent a given model may be appropriately fitted and selected for, given relatively sparse data. Estimating reaction rates and size distributions of protein polymers is an important step towards understanding the mechanisms of protein misfolding and aggregation, a key feature for amyloid diseases. Specifically, experimental measurements often consist in the time-dynamics of a moment of the population (*i.e.*, for instance the total polymerised mass, as in Thioflavine T measurements, or the second moment measured by Static Light Scattering).

In a first study [4], in collaboration with H.T. Banks and H. Rezaei, we illustrated the use of tools (asymptotic theories of standard error quantification using appropriate statistical models, bootstrapping, model comparison techniques) in addition to sensitivity that may be employed to determine the information content in data sets. We do this in the context of recent models [87] for nucleated polymerisation in proteins, about which very little is known regarding the underlying mechanisms; thus the methodology we developed may be of great help to experimentalists.

In another study [39], related to a different biological setting (the frog olfactive tract), we use a method based on the Mellin transform, as in [64], to solve a spectral inverse problem arising from the modeling of the transduction of an odor into an electrical signal. The problem is to find the spatial distribution of CNG ion channels along the cilium of a frog, which allow a depolarizing influx of sodium ions, which initiate the electrical signal. This problem comes down to solving a Fredholm integral equation. We prove observability and continuity inequalities by estimating the Mellin transform of the kernel of this integral equation. We perform numerical computations using experimental data.

To get more insight into the estimation of reaction rates and size distributions of protein polymers, we are now developing an approach based on a data assimilation strategy. In this purpose, A. Armiento's Ph.D is focused on setting this framework problem when the experimental measurements consist in the time-dynamics of a moment of the population (*i.e.* for instance the total polymerised mass, as in Thioflavine T measurements, or the second moment measured by Static Light Scattering). In [37] we proposed a general methodology, and we solved the problem theoretically and numerically in the case of a depolymerising system. We then applied our method to experimental data of degrading oligomers, and conclude that smaller aggregates of ovPrP protein should be more stable than larger ones. This has an important biological implication, since it is commonly admitted that small oligomers constitute the most cytotoxic species during prion misfolding process.

7.2.3. *Time asymptotics for growth-fragmentation equations*

The long-term dynamics of fragmentation and growth-fragmentation equations has been for long an important research field for BANG then MAMBA research team. Thanks to these common efforts, these equations are now well understood. However, there remain some interesting open questions. In particular, if the generic long-time behaviour for the linear equation is known - given by a (generally exponential) trend towards a steady exponential growth described by the positive eigenvector linked to the dominant eigenvalue, see [84] for most recent results - critical cases are not yet fully understood.

With Miguel Escobedo, we focused on an important critical case, when the fragmentation is constant and the growth rate is either null or linear [43]. Using the Mellin transform of the equation, we determine the long time behaviour of the solutions and the speed of convergence, which may be either exponential or at most polynomial according to the subdomain of $(t, x) \in \mathbb{R}_+^2$ which is considered. Our results show in particular the strong dependence of this asymptotic behaviour with respect to the initial data, in contrast to the generic results. Following our study, J. Bertoin and A. Watson proposed a complementary probabilistic analysis of related models [60]. These results exemplify the continuing need for further analysis of these interesting equations.

7.2.4. Cell aggregation by chemotaxis

We follow our investigation on the kinetic model describing the chemotactic motion of bacteria. When taxis dominates the unbiased movements, the kinetic system is approximated by the aggregation equation. The study of such equation is challenging since blow-up in finite-time of solutions occurs. We have defined the notion of measure-valued solution [8] and we have proposed and studied a numerical scheme to simulate these solutions[18].

In another approach, more accuracy can be obtained with the kinetic model by adding an internal variable describing the methylation level of the internal receptors of bacteria. In [55] we have investigated the link between these kinetic models with an internal variable and the one without internal variable.

7.3. Liver modeling

Participants: Noémie Boissier, Dirk Drasdo, Géraldine Cellière, Adrian Friebel, Group Heinzle [Univ. Saarbruecken, Germany], Group Hengstler [IfADo, Germany], Stefan Hoehme, Tim Johann, Irène Reo [Vignon-Clementel], Paul Van Liedekerke, Eric Vibert [Hopital Paul Brousse], Group Zerial [Max-Planck Inst. for Molecular Genetics, Dresden, Germany], Groups Iflow, Notox, Vln.

7.3.1. Ammonia detoxification after drug-induced damage

Overdosing acetaminophen (APAP) is the main reason for acute liver failure in the US and UK. Overdose of APAP destroys the hepatocytes located in the center of each liver lobule (pericentral damage), the repetitive functional and anatomical tissue units of liver. Human has about a million of such lobules. As a consequence, the blood is not sufficiently detoxified from ammonia, which is toxic to the body and can lead to encephalopathy. In France about 1000 cases of ammonia intoxication each year. In recent papers we demonstrated by an integrated model that the widely accepted key reactions scheme of ammonia detoxification is insufficient to explain ammonia detoxification after pericentral lobule damage and predicted a missing ammonia sink [73]]. This finding has triggered new experiments leading to the identification of a widely ignored but fundamentally important ammonia sink mechanism We could show by a testing a number of different mechanisms within novel models that this sink mechanism was the only one able to explain the data [15]. The reaction turned out to have the potential to be therapeutically used by injection of a molecular cocktail triggering it. In the animal model death could be prevented using this cocktail hence providing a possible therapy approach for patients suffering from hyperammonemia. [15]. In a follow-up work, further models have been studied and classified by statistical methods to quantify model selection (Cellière et al., in preparation).

7.3.2. Concepts of modeling of liver across all scales: multiscale liver modeling

Based upon developped multiscale concepts [12], we developped a multi-level spatial temporal multiscale models of APAP (paracetamol, acetaminophen) toxicity and ammonia metabolism. In on of these models we integrated molecular pathways of APAP drug toxicity (PD); in another one, we represented the ammonia detoxification pathway into each individual hepatocyte of an agent-based model that describes the precise liver lobule architecture (compare with [73]). This allows us to study the impact of space and architecture on the drug toxicity and drug detoxification. We find in certain cases important differences between models that do represent architecture and those that do not (Cellière et al., in preparation).

7.3.3. Predicting in vivo drug toxicity from in vitro data

APAP (paracetamol, acetaminophen) *in vitro* experiments have been used to calibrate a model of APAP drug toxicity with *in vitro* data, and modify this model to predict *in vivo* toxicity. This procedure is aimed at as a general pathway among cosmetic and pharmaceutical companies to eliminate or at least reduce animal experiments and it should allow a better prediction of drug toxicity in human. Three critical differences between *in vitro* and *in vivo* settings were stepwise integrated in the model calibrated with *in vitro* toxicity data to study their impact on *in vivo* toxicity predictions: (1) The temporal drug exposure profile, (2) the temporal concentration profile of a class of key enzymes, CYP enzymes. Only in hepatocytes in which CYP enzymes are present, APAP is metabolised and downstream apoptosis can occur. (3) The liver architecture, that is responsible for critical differences in the spatial distribution of the drug. The results are in preparation for publication (Cellière et. al., in preparation).

7.3.4. Miscellaneous

In addition, regenerating lobules after partial hepatectomy were analysed by image analysis, and first simulations of blood and bile flow and molecular transport in those lobules simulated.

7.4. Miscellaneous

Participants: Noémie Boissier, Maria José Cáceres [Universidad de Granada], Julien Chevallier [Université de Nice], Géraldine Cellière, Marie Doumic, Dirk Drasdo, Adrian Friebel, Group Heinzle [Univ. Saarbruecken, Germany], Group Hengstler [IfADo, Germany], Stefan Hoehme, Tim Johann, Group Klingmueller [German Cancer Center, Heidelberg], Johannes Neitsch, Benoît Perthame, Patricia Reynaud [Université de Nice], Group Reo [Inria Paris - Rocquencourt], Paul Van Liedekerke, Eric Vibert [Hopital Paul Brousse], Yi Yin, Group Zerial [Max-Planck Inst. for Molecular Genetics, Dresden, Germany], Groups Iflow, Notox, Vln.

7.4.1. Network formation and neuroscience

Motivated by neurodevelopment and differentiation in developing tissues, a new explanation for sharp boundary formation is analysed in [25]; interestingly, this phenomenon relies on a limited diffusion of homeoproteins (collaboration with the Mycenae team).

Models for neural networks have been proposed which describe the probability to find a neuron for which a time s has elapsed since the last discharge. These are written under the form of a nonlinear age-structured equation where the total network activity modulates the firing rate. An inhomogeneous network of networks with variability on the refractory period is studied in [19].

We have also continued the analysis and numerical simulation of models for natural transportation networks formation based on an elliptic-parabolic system of partial differential equations. The model describes the pressure field using a Darcy's type equation and the dynamics of the conductance network under pressure force effects. Randomness in the material structure is represented by a linear diffusion term and conductance relaxation by an algebraic decay term [16]. Figure 1 below gives a numerical simulation of a network formed by such a model.

7.4.2. Microscopic approach of a time elapsed neural model

The spike trains are the main components of the information processing in the brain. To model spike trains several point processes have been investigated in the literature. More macroscopic approaches have also been studied, using partial differential equation models. With J. Chevallier, M. Cáceres and P. Reynaud-Bouret, we wanted to build a bridge between several point processes models (Poisson, Wold, Hawkes) that have been proved to statistically fit real spike trains data and age-structured partial differential equations as introduced by Pakdaman, Perthame and Salort. To do so, we focused on a seemingly simple one-neuron model, for which we stated the - nonlinear and strongly coupled - PDE model satisfied in average by its point measure when the process model is a Poisson, a Wold or a Hawkes process [10].

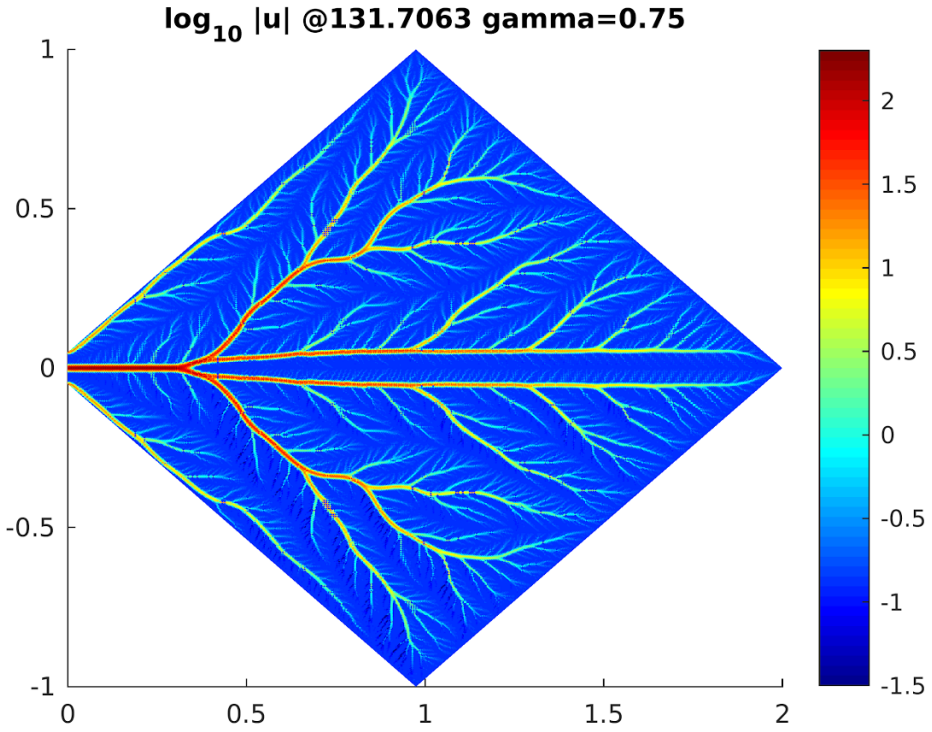


Figure 1. Network formation based on an elliptic-parabolic system of partial differential equations.

7.4.3. Uncertainty propagation

In [42], we study two intrusive methods for uncertainty propagation in scalar conservation laws based on their kinetic formulations. The first one is based on expansions on an orthogonal family of polynomials. The second method uses convolutions based on Jackson kernels. We prove that it satisfies BV bounds and converges to the entropy solution but with a spurious damping phenomenon. Therefore we introduce a second method, which is based on projection on layered Maxellians, and which arises as a minimisation of entropy. This new method satisfies the maximum principle by construction as well as partial entropy inequalities and thus provides an alternative to the standard method of moments which, in general, does not satisfy the maximum principle. Simple numerical simulations for the Burgers equation illustrate these theoretical results.

7.4.4. Simulation of tissue mechanics with agent-based models

In ref. [29] we study and discuss in how far mechanical effects of cells in tissue organisation and growth processes can be captured by agent-based models. We consider a wide range of agent-based models, i.e., lattice base models with one lattice site allowing for many cells or one cell at most, many lattice sites occupied by a single cells (so called Cellular Potts model, Lattice Gas Cellular Automaton approaches, center-based models and vertex models, in which the forces between cells are calculated as forces between the cell centers, as well as deformable cell models in which the cell surface is triangulated. We consider growth of monolayers and multicellular spheroids as reference problems. We also compare in this paper spatial resolution, the capability of the different approaches to represent the physics, cell shape, the computational efficiency and code access. In addition, models evaluating the mechanical effects of growing cell populations in elastic capsules were established and studied.

MODEMIC Project-Team

7. New Results

7.1. Mathematical models for microbial ecology

7.1.1. Differential equations models

Participants: Jérôme Harmand, Claude Lobry, Alain Rapaport, Yessmine Daoud, Sonia Hassam, Zeyneb Khedim, Alejandro Maximiliano Rojas.

Anaerobic digestion refers to the transformation of biodegradable material by micro-organisms in absence of oxygen (it can be found in waste-water treatments or industrial fermentation, and occurs naturally in soils). It receives an increasing consideration due to recent technological advances, but also because it is a source of renewable energy (bio-gas, fuel...). The anaerobic digestion is a complex set of bio-processes, for which there is a strong expectation of tractable models. We have proposed and studied new mathematical models that takes into account the following features:

- Microbial food chains are present in anaerobic digestion where the different reaction steps can be seen as such: the waste products of the organisms at one trophic level (i.e. one reaction step) are consumed by organisms at the next trophic level (i.e. the next reaction step). In [54] we study a model of a two-tiered microbial 'food chain' with feedback inhibition, which was recently presented as a reduced and simplified version of the anaerobic digestion model ADM1 of the International Water Association (IWA) (cf. [61]). It is known that in the absence of maintenance (or decay) the microbial 'food chain' is stable. In [61], using a purely numerical approach and ADM1 consensus parameter values, it was shown that the model remains stable when decay terms are added. In [54] we prove that introducing decay in the model preserves stability whatever its parameters values are and for a wide range of kinetics.
- In the thesis by Sonia Hassam [13], we have proposed a new procedure to easily and systematically obtain a simple model useful for control purposes of any process for which an ADM1 is available. The simplified model has two major characteristics : its states keep their physical meaning and it remains nonlinear. The technique is based on the state association technique proposed in [26].
- Zeyneb Khedim (University of Tlemcen, Algeria) has began her PhD in 2014. She is working on the modeling and control of anaerobic digestors. In particular, she works on the reduction of models using the state association approach proposed by Sonia Hassam but for substrates highly loaded in nitrogen such as algae. She has published this year a survey with Sonia Hassam [36].
- Yessmine Daoud (ENIT-LAMSIN, Tunis, Tunisia) continues her work on the analysis of a model of the literature to optimize anaerobic processes [35]. She is preparing a journal paper which should be submitted during 2016.

Formerly, the team has studied chemostat models where the bacterial compartment is split into "planktonic" and "attached" bacteria (such as in flocculation or biofilms formation), under the hypothesis that attachment and detachment are fast phenomena. Under certain mixing conditions, this last condition is no longer satisfied. We have studied on the non-reduced model the competition between a species that presents growth inhibition in planktonic form with a species that does not attach. This consideration leads to multiple positive equilibria but surprisingly it can also conduct to limit cycles [53] (paper under revision for Applied Math. Model.).

Spatial heterogeneity is often observed in non perfectly mixed bioprocesses or in populations in natural environments. The representation o spatial heterogeneity in population models with patches or interconnected models, rather than p.d.e., is one of the specialties of the team, that allows us to characterize non intuitive effects of spatialization :

- The very basic Rosenzweig-MacArthur model is subject to the "atto-fox" problem [2] when considered for homogeneous populations. Is it still true in case of heterogeneous populations? The idea is: the resource population being not small at the same time in different places is it possible that, thanks to dispersal, it will not disappear? One possible idealization of heterogeneous populations is to use reaction-diffusion equations. We do not take this direction for two reasons
- (i) Due to the presence of a limit cycle in the homogeneous system mathematics of such reaction diffusion are difficult.
- (ii) Idealization through reaction-diffusion is not the best one; patch-systems (or lattice differential equations in mathematical terms) are better in many cases.

Our ultimate objective is to provide mathematical results for systems with a large number of patches but, as a first step, in the paper [27] we consider two patches. It is proved that for some migration rates, stable periodic solutions avoiding "atto-fox" exist.

- The standard model for the dynamics of a fragmented density-dependent population is built from several local logistic models coupled by migrations.

First introduced in the 1970s and used in innumerable articles, this standard model applied to a two-patch situation has never been completely analyzed. The motivation for studying this problem came out from discussions at the Bernoulli semester organized in 2014 and 2015 by the team at the EPFL (see the 2014 activity report and Section 8.3.3.1). It addresses very fundamental issues in theoretical ecology. In the paper [15] written in collaboration with R. Arditi (U. Fribourg) an T. Sari (IRSTEA Montpellier), we complete this analysis and we delineate the conditions under which fragmentation associated to dispersal is either beneficial or detrimental to total population abundance. Therefore, this is a contribution to the SLOSS question. Importantly, we also show that, depending on the underlying mechanism, there is no unique way to generalize the logistic model to a patchy situation. In many cases, the standard model is not the correct generalization. We analyze several alternative models and compare their predictions. Finally, we emphasize the shortcomings of the logistic model when written in the r-K parameterization and we explain why Verhulst's original polynomial expression is to be preferred.

- We have carried on our former work on the role of particular interconnections patterns on the global stability of chemostat model with inhibition. While we focused formerly on the conditions for which a spatial structure ensures the global stability when the chemostat model is bi-stable in homogeneous environment, we have shown that at the opposite a spatial structure can make unstable the dynamics of the chemostat model with inhibition when it is stable in a homogeneous environment [30].
- In collaboration with Géosciences Rennes (Jean-Raynald de Dreuzy, Tristan Babey) and in the scope of the co-supervision of the PhD of Alejandro Rojas (also in the collaboration within the associated team with Chile), we have carried on the complete equivalence between several models used in Geosciences to characterize soil fractures : MINC (Multiple INteracting Continua), MRMT (Multi-Rate Mass Transfer) and SINC (Structured INteracting Continua). We have shown that the irreducibility of the network graph is not sufficient to obtain equivalence : a controllability assumption has also to be fulfilled [42]. Moreover, this kind of models has been used to fit experimental data of reconstituted soils at Inra Grignon and has shown the role of convection in the acquisition of pesticides by micro-organisms [46] (paper in preparation). This work will be continued in the framework of the new ANR project Soil$\mu 3D$ (see Section 8.2.1).

In resources/consumers models, heterogeneity can be also due to time varying inputs of resources (e.g. light in micro-algae populations). While, most of the literature studies periodic inputs, we have begun investigations of more general time varying inputs in chemostat like models, having in mind to characterize "pull-back attractors" (rather than forward attractors) [43].

7.1.2. Stochastic and hybrid discrete-continuous dynamical models

Participants: Bertrand Cloez, Claude Lobry.

7.1.3. Approximation of quasi-stationary distributions

The study of the long-time behavior of a stochastic process is one of the main questions of interest for modeling. In a standard Markov setting, this leads to the study of the convergence towards the invariant distribution. However, in many applications such as population dynamics for instance, the stochastic dynamics is killed in a finite (random) time so that the standard asymptotic regime is trivial. In this case, it can be interesting to focus on the behavior of the process conditionally to its non-extinction before a given time t Under appropriate assumptions, one can exhibit a convergence of this conditional distribution towards a law called *Quasi-Stationary Distribution*. Properties of this law is then fundamental. In [21], we study an algorithm to approximate this distribution and we provide proof of convergence as well as precise rates for convergence. This one is based on a reinforced random walk.

7.1.3.1. Lotka Volterra in fluctuating environment

In the paper [49], we consider two dimensional Lotka-Volterra systems in a fluctuating environment. Relying on recent results on stochastic persistence and piecewise deterministic Markov processes, we show that random switching between two environments that are both favorable to the same species can lead to the extinction of this species or coexistence of the two competing species. This work has been accepted in Journal of applied probabilities, provided major revisions. We submitted a new version with the new title: Lotka Volterra with randomly fluctuating environments or "how switching between beneficial environments can make survival harder".

7.2. Analysis and supervision of bioprocesses

7.2.1. Models development and identification

Participants: Yessmine Daoud, Jérôme Harmand, Nesrine Kalboussi, Guilherme Pimentel, Alain Rapaport.

Membrane bioreactors combine a filtration process (with a membrane) and a suspended growth rate bioreactor. This recent technology present many advantages compared to conventional ones, but is more sophisticated and requires refined control because of possible problems related to the risk of membrane fouling. After the PhD by Amine Charfi defensed in 2014 we continue to work on the modeling and control of membrane bioreactors.

- Within this framework, new results have been obtained and a new model including several fouling mechanisms has been proposed [22].
- In the scope of the PhD of Guilherme Pimentel (defensed in February 2015, [11]), we have proposed a simple three time scales model in view of the control of the cake formation [14], [33]. This model has been validated on real data from a pilot plant at Univ. Mons (Belgium).
- The PhD thesis by Nesrine Kalboussi (ENIT-LAMSIN, Tunis, Tunisia) has just begun. It is dedicated to the early detection and control of membrane fouling. At present time, Nesrine is working on the bibliography about modeling and control of membrane bioreactors.

In many bioprocesses models, the loss of nutrient used for the maintenance of bacteria is neglected compared to the important nutrient supply. In poor environment, such as natural one in the oceans, this is no longer verified. In our collaboration with the LOMIC lab (Banyuls-sur-Mer), we have shown that the consideration of a maintenance term in the chemostat model allows to fit the data observed in experimental chemostats, and moreover that the level of maintenance is correlated to the activities of bacteria under the presence of light [23]. This gives a possible explanation of the variable yield observed in the bacterial compartment of marine ecosystems.

7.2.2. Synthesis of control laws

Participants: Térence Bayen, Walid Bouhafs, Amel Ghouali, Jérôme Harmand, Claude Lobry, Guilherme Pimentel, Alain Rapaport, Victor Riqueleme.

We investigate two kinds of bioprocesses to be controlled, arising in industrial biotechnology (digesters, waste-water purification...) or in the bioremediation of natural environments (lakes, landfill...).

7.2.2.1. Industrial biotechnology

In the framework of the PhD of Guilherme Pimentel [11] and the pilot plant at Mons, a nonlinear predictive law based on the model exposed in 7.2.1 has been tested and validated for piloting the process[34].

Control of biological reactors are still of great interest, notably but not only with respect to anaerobic digestors that can be destabilized due to the accumulation of intermediate metabolites that can inhibate the growth of some bacteria.

- Amel Ghouali (Cotutelle Univ. Montpellier and Univ. of Tlemcen, Algeria) who has defensed her PhD in December has developed an optimal control strategy to optimize the production of biogas over a given period of time [12]. In particular, she has solved an original optimal control problem using the maximum principle of Pontryagin [25].
- Within the scope of the PhD thesis by Walid Bouhafs (ENIT-LAMSIN, Tunis, Tunisia), we have proposed a new optimal control strategy for systems in which two specific substrates are degraded by two different bacterial consortia, one being limited by the oxygen while the other is inhibited. Walid will defense his PhD in next February.

The minimal time criterion is of particular interest in biotechnology, as it leads to time-independent feedback controllers.

- The paper [19] is devoted to the study of the minimal time problem of a fed-batch reactor, under the presence of a saturation point on the singular locus (this typically occurs whenever the growth rate function is of Haldane type and when typically the maximum input flow rate is not high enough to maintain the substrate concentration constant). This brings non-intuitive issues for the optimal synthesis (existence of switching curve and point of prior saturation).
- In the work [47], we study the minimal time control to drive a chemostat model to a target point. Such a problem finds application typically in the case where the input substrate concentration changes yielding in a new steady state. Converging fast towards this new reference point is much desired in practice. One essential feature of the present work is that the system takes into account a recirculation of biomass (as it is more and more often the case in modern biotechnology). We depict an optimal synthesis and provide an optimal feedback control by using the Pontryagin Maximum Principle and geometric control theory for both Monod and Haldane kinetics.

7.2.2.2. Bioremediation of natural environments

In the scope of the associated team with Chile (see 8.3.1.1) and the co-supervision of the PhD of Victor Riquelme, we have carried on the study of optimal syntheses for the minimal time treatment of natural water reservoirs (such as lakes) [41]. We have proved that the minimal time strategy consists in a most-rapid approach to homogeneous concentrations, even though the optimal control problem is non convex. Moreover, we have shown that a large diffusion increases the treatment time when the resource is everywhere highly polluted, while it can at the opposite be beneficial when only part of the resource is polluted (paper under revision for SIAM J. Cont. & Optim.). This feature should serve the practitioners in the choice of pumps positioning in a originally clean water resource that is suddenly affected by a local pollution. Moreover, we have shown, in collaboration with A. Rousseau (EPI LEMON), how these analytic feedback laws obtained on a over-simplified representation of the spatial heterogeneity behave quite satisfactorily when simulated [17]. This year we have started to study to problem of treating two different pollutants, with a anaerobic/aerobic process in series.

Also in the scope of the associated team with Chile, we have characterized the optimal strategy to treat as fast as possible a landfill with the recirculation as a manipulated variable [40], [29], based on a model that we have proposed last year. In presence of singular arcs that are non-admissible (in the sense that the upper bound of the recirculation pump does not allow to stay on the singular arc), we have shown that a kind of *anticipation* law is necessary before operating optimally the switching. This analyses reveals several sub-domains for which the optimal policy requires different kind of measurements. Knowing in which sub-domain the initial stage

of landfill could be inform then the practitioners about which concentration (leachate or solubilized or both) should be ideally measured. This primarily work has led to the co-development of a software mock-up with Chilean partner (see 6.1), in order the study the consideration of spatial heterogeneity in landfill, with the approach exposed in 7.1.1 .
This year, again in the scope of the associated team with Chile and Inria Chile, we have begun a new investigation on modeling and control strategy for the regulation of a lagoon that communicates temporarily with the sea and whose water is exploited by pumping.

7.2.2.3. Theoretical development

The time crisis is an interesting criterion that measures the time spent by a system in a "bad" zone or in "danger". Typically, when a desired species is under a given (low) threshold, one can consider that this defines a crisis domain. For controlled system, the minimal time crisis has already been proposed in the literature [56]. Nevertheless, only sufficient conditions (i.e. characterization of the solutions of the associated Hamilton-Bellman-Jacobi equation) have been given, and no necessary conditions have been yet proposed, due to the lack of continuity of the integrand cost. We have proposed a regularization of this problem by a family of optimal control problems for which the usual necessary conditions can be derived, and studied the convergence [20]. Practically, this allows to use classical software, such as Bocop, to approximate the optimal solutions. In the internship of C. Romero (U. Chile), this technique has been successfully applied on the Lotka-Volterra model with a control on the predator, and a threshold on the prey.

Monc Team

7. New Results

7.1. Axis 1: Tumor modeling for patient-specific simulations

7.1.1. Lung metastasis

Patient specific simulation of tumor growth, response to the treatment and relapse of a lung metastasis: a clinical case [10], [1]
Team participants: Thierry Colin, Julien Jouganous, François Cornelis (Hôpital Pellegrin), Olivier Saut
Other participant: Jean Palussière (Bergonié Institute)

In this work, a parametrization strategy based on reduced order methods is presented for tumor growth PDE models. This is applied to a new simple spatial model for lung metastasis including angiogenesis. The goal is to help clinicians monitoring tumors and eventually predicting its evolution or response to a particular kind of treatment. To illustrate the whole approach, a clinical case including the natural history of the lesion, the response to a chemotherapy and the relapse before a radiofrequency ablation is presented.

Nenuphar
Team participants: Thierry Colin, Julien Jouganous, Marie Martin, Olivier Saut

This work concerns the development of *Nenuphar* which is a software devoting to the evaluation and the surveillance of the tumor aggressiveness.

7.1.2. Take into account the drug resistance

Modeling and analysis of tumor heterogeneity during treatments resistance: GIST liver metastases case
Team participants: Thierry Colin, François Cornelis, Guillaume Lefebvre, Clair Poignard, Olivier Saut

This works deals with tumor heterogeneity analysis and modeling during treatments resistances. A patient-dependent PDEs model, that takes into account two kinds of treatments, is presented. It qualitatively and quantitatively reproduces the different stage during the tumor growth undergoing treatments. In order to overcome a numerical instability linked to the type of modeling, a new numerical scheme is built. Then, an image synthesis method is developed to enable a better comparison between the numerical results and the clinical data. Finally, a robust criteria that quantifies the tumor heterogeneity from the clinical data and from the synthesis images, is built.

Mathematical study and asymptotic analysis of a model for tumour drug resistance [19]
Team participants: Thierry Colin, Thomas Michel, Clair Poignard

In this work we study a partial differential equations model for tumour growth taking into account drug resistance. It is well known that angiogenesis, the process of creation of new blood vessels from existing ones, is induced by tumour cells to get the amount of nutrients and oxygen needed to continue their proliferation when the tumour has reached a critical size. Angiogenesis is therefore a target for therapy. The model we study takes into account two kinds of treatments: a cytotoxic treatment and a treatment which is both cytotoxic and anti-angiogenic. It is based on mass-balance equations on cells densities coupled with a diffusion equation for the nutrients and oxygen concentration. In a first part we prove that the model is well-posed if the initial tumour is compactly supported in the domain, which is the case for tumour metastases. The proof states that the tumour remains compactly supported in a finite time. In the model, we also consider the presence of a necrotic compartment composed of dead cells. Since some tumours can present necrosis while other do not, we want a model which can reproduce these two different cases. The second part of this work is devoted to an asymptotic analysis which proves that the absence of necrosis is the limit case of our model when the necrosis is immediately evacuated.

7.1.3. Motility phenotype

TMOD-03 * Motility controls growth and progression patterns of glioblastoma multiforme [13]
Team participants: Olivier Saut, Thierry Colin
Other participants: Hassan Fathallah, Elizabeth Scribner

Purpose: Glioblastoma multiforme (GBM) is a malignant brain tumor with poor prognosis and high morbidity due to its invasiveness. Hypoxia-driven motility (HM) and concentration-driven motility (CM) are two mechanisms of GBM invasion in the brain. The use of anti-angiogenic drugs has uncovered new progression patterns of GBM associated with significant differences in overall survival times. Here, we test the hypotheses that the types and rates of GBM motility predict its progression pattern and the patients' survival times. Methods: We applied a mathematical model of GBM growth and invasion in humans to simulate a clinical trial and study the effects of the rate and mechanism of motility on the patterns of progression and on survival times. Results: The motility phenotype appears to determine the progression pattern as well as the survival time of a patient treated by anti-angiogenesis. Highly-dispersive tumors are associated with the longest survival times (p , 0.001) and with progression by Expanding FLAIR. Moderately-Dispersive tumors are associated with short survival times and with progression by Expanding FLAIR + Necrosis. Tumors with HMare associated with the shortest survival times and with progression by Expanding Necrosis. The survival times of the latter are similar to non-responders. This investigation also uncovered the HM-CM principle: the aggressive HM-dependent phenotype surfaces only when the rate of CM is low in both untreated and bevacizumab-treated GBM. Conclusions: Finding that the motility phenotype is a fundamental property that controls progression and survival times, has biological, clinical and therapeutic implications.

7.2. Axis 2: Bio-physical modeling for personalized therapies

7.2.1. Electropermeabilization

Non-Linear Steady-State Electrical Current Modeling for the Electropermeabilization of Biological Tissue [8]
Team participants: Clair Poignard, Michael Leguebe
Other participants: Marie Breton, Lluis M. Mir (Vectorology and Anticancer Therapies), Francois Buret, Riccardo Scorretti, Damien Voyer, Laurent Krähenbühl (Ampère Laboratory (Lyon) participants), Ronan Perrussel (LAPLACE - Laboratoire Plasma et Conversion d'Energie, Toulouse)

We propose a non-linear steady-state model of irreversible electropermeabilization in a biological tissue. The non-linear problem is solved using a modified fixed point iteration. The unknown parameters are experimentally estimated from the observation of the necrosis on a potato tissue for different applied voltages. A variability study of the parameters involved in the model is performed.

A second-order Cartesian method for the simulation of electropermeabilization cell models [12]
Team participants: Clair Poignard, Michael Leguèbe
Other participant: Lizl Weynans (Memphis team, Inria)

In this work, we present a new finite differences method to simulate electropermeabilization models, like the model of Neu and Krassowska or the recent model of Kavian et al. These models are based on the evolution of the electric potential in a cell embedded in a conducting medium. The main feature lies in the transmission of the voltage potential across the cell membrane: the jump of the potential is proportional to the normal flux thanks to the well-known Kirchoff law. An adapted scheme is thus necessary to accurately simulate the voltage potential in the whole cell, notably at the membrane separating the cell from the outer medium. We present a second-order finite differences scheme in the spirit of the method introduced by Cisternino and Weynans for elliptic problems with immersed interfaces. This is a Cartesian grid method based on the accurate discretization of the fluxes at the interface, through the use of additional interface unknowns. The main novelty of our present work lies in the fact that the jump of the potential is proportional to the flux, and therefore is not explicitly known. The original use of interface unknowns makes it possible to discretize the transmission conditions with enough accuracy to obtain a second-order spatial convergence. We prove the second-order spatial convergence in the stationary linear one-dimensional case, and the first-order temporal convergence for the dynamical non-linear model in one dimension. We then perform numerical experiments in two dimensions that corroborate these results.

Cell membrane permeabilization by 12-ns electric pulses: Not a purely dielectric, but a charge-dependent phenomenon [15]
Team participants: Clair Poignard, Michael Leguèbe
Other participants: Aude Silve (KIT - Karlsruhe Institute of Technology), Isabelle Leray, Lluis M. Mir (Université Paris Sud)

Electric pulses of a few nanoseconds in duration can induce reversible permeabilization of cell membrane and cell death. Whether these effects are caused by ionic or purely dielectric phenomena is still discussed. We address this question by studying the impact of conductivity of the pulsing buffer on the effect of pulses of 12 ns and 3.2 MV/m on the DC-3F mammalian cell line. When pulses were applied in a high-conductivity medium (1.5 S/m), cells experienced both reversible electropermeabilization and cell death. On the contrary, no effect was observed in the low-conductivity medium (0.1 S/m). Possible artifacts due to differences in viscosity, temperature increase or electrochemical reactions were excluded. The influence of conductivity reported here suggests that charges still play a role, even for 12-ns pulses. All theoretical models agree with this experimental observation, since all suggest that only high-conductivity medium can induce a transmembrane voltage high enough to induce pore creation, in turn. However, most models fail to describe why pulse accumulation is experimentally required to observe biological effects. They mostly show no increase of permeabilization with accumulation of pulses. Currently, only one model properly describes pulse accumulation by modeling diffusion of the altered membrane regions.

7.2.2. Cell protrusion

Free boundary problem for cell protrusion formations: theoretical and numerical aspects [20]
Team participants: Olivier Gallinato, Clair Poignard
Other participants: Masahito Ohta (Tokyo University of Sciences), Takashi Suzuki (Osaka University)

In this work, we derive a free boundary problem for cell protrusion formation in which the cell membrane is precisely described thanks to a level-set function, whose motion is due to specific signalling pathways. The model consists in Laplace equation with Dirichlet condition inside the cell coupled to Laplace equation with Neumann condition in the outer domain. The motion of the interface is due the gradient of the inner quantity. We prove the well-posedness of our free boundary problem under a sign condition on the datum similarly to the Taylor criterion in water waves. We also propose an accurate numerical scheme to solve the problem and we exhibit the main biological features that can be accounted for by the model. Even though simplistic from the modeling point of view, we claim that this work provides the theoretical and numerical grounds for single cell migration modeling. In particular, specific chemical reactions that occurred at the cell membrane could be precisely described in forthcoming works.

7.3. Axis 3: Quantitative cancer modeling for biological and preclinical studies

7.3.1. Modelling of metastasis development

Computational Modelling of Metastasis Development in Renal Cell Carcinoma [2]
Team participants: Etienne Baratchart, Sébastien Benzekry, Thierry Colin, Olivier Saut
Other participants: Andreas Bikfalvi, Lindsay S. Cooley, Raphäel Pineau, Wilfried Souleyreau (LAMC - Laboratoire Angiogenèse et Micro-environnement des Cancers), Emeline J Ribot (RMSB - Résonance magnétique des systèmes biologiques)

To improve our understanding of the biology of the metastatic colonization process, we conducted a modelling study based on multi-modal data from an orthotopic murine experimental system of metastatic renal cell carcinoma. The standard theory of metastatic colonization usually assumes that secondary tumours, once established at a distant site, grow independently from each other and from the primary tumour. Using a mathematical model describing the metastatic population dynamics under this assumption, we challenged the theory against our data that included: 1) dynamics of primary tumour cells in the kidney and metastatic cells in the lungs, retrieved by green fluorescent protein tracking, and 2) magnetic resonance images (MRI) informing on the number and size of macroscopic lesions. While the model could fit the primary tumour and total metastatic burden, the predicted size distribution was not in agreement with the MRI observations. Moreover, the model was incompatible with the growth rates of individual metastatic tumours. To explain the observed metastatic patterns, we hypothesised that metastatic foci derived from one or a few cells could aggregate, resulting in a similar total mass but a smaller number of metastases. This was indeed observed in our data and led us to investigate the effect of spatial interactions on the dynamics of the global metastatic burden. We derived a novel mathematical model for spatial tumour growth, where the intra-tumour increase in pressure is responsible for the slowdown of the growth rate. The model could fit the growth of lung metastasis visualized by magnetic resonance imaging. As a non-trivial outcome from this analysis, the model predicted that the net growth of two neighbouring tumour lesions that enter in contact is considerably impaired (of $31\% \pm 1.5\%$, mean $\pm$ standard deviation), as compared to the growth of two independent tumours. Together, our results have implications for theories of metastatic development and suggest that global dynamics of metastasis development is dependent on spatial interactions between metastatic lesions.

Modeling spontaneous metastasis following surgery: an in vivo-in silico approach [6]
Team participant: Sebastien Benzekry
Other participants: Amanda Tracz, Michalis Mastri, Ryan Corbelli, Dominique Barbolosis, John Ebos (Buffalo University)

Rapid improvements in the detection and tracking of early-stage tumor progression aim to guide decisions regarding cancer treatments as well as predict metastatic recurrence in patients following surgery. Mathematical models may have the potential to further assist in estimating metastatic risk, particularly when paired with in vivo tumor data that faithfully represent all stages of disease progression. Herein we describe mathematical analysis that uses data from mouse models of spontaneous metastasis developing after surgical removal of orthotopically implanted primary tumors. Both presurgical (primary tumor) and postsurgical (metastatic) growth was quantified using bioluminescence and was then used to generate a mathematical formalism based on general laws of the disease (i.e. dissemination and growth). The model was able to fit and predict pre-/post-surgical data at the level of the individual as well as the population. Our approach also enabled retrospective analysis of clinical data describing the probability of metastatic relapse as a function of primary tumor size. In these data-based models, inter-individual variability was quantified by a key parameter of intrinsic metastatic potential. Critically, our analysis identified a highly nonlinear relationship between primary tumor size and postsurgical survival, suggesting possible threshold limits for the utility of tumor size as a predictor of metastatic recurrence. These findings represent a novel use of clinically relevant models to assess the impact of surgery on metastatic potential and may guide optimal timing of treatments in neoadjuvant (presurgical) and adjuvant (postsurgical) settings to maximize patient benefit.

Migration and orientation of endothelial cells on micropatterned polymers: A simple model based on classical mechanics [11]
Team participants: Thierry Colin, Clair Poignard, Olivier Saut
Other participants: Julie Joie, Marie-Christine Durrieu (IMB - Institut de Mathématiques de Bordeaux), Yifeng Lei (French Institute of Health and Medical Research, Paris)

Understanding the endothelial cell migration on micropatterned polymers, as well as the cell orientation is a critical issue in tissue engineering, since it is the preliminary step towards cell polarization and that possibly leads to the blood vessel formation. In this work, we derive a simple agent-based model to describe the migration and the orientation of endothelial cells seeded on bioactive micropatterned polymers. The aim of the modeling is to provide a simple model that corroborates quantitatively the experiments, without considering the complex phenomena inherent to cell migration. Our model is obtained thanks to a classical mechanics approach based on experimental observations. Even though its simplicity, it provides numerical results that are quantitatively in accordance with the experimental data, and thus our approach can be seen as a preliminary way towards a simple modeling of cell migration.

7.3.2. Tumor-host crosstalk

Host age is a systemic regulator of gene expression impacting cancer progression [3]
Team participant: Sebastien Benzekry
Other participants: Afshin Beheshti, Lili Ma, Philip Hahnfeldt, Lynn Hlatky (CCSB - Center of Cancer and Systems Biology), J. Tyson Mcdonald (University of Houston), Michael Peluso (Cancer Risk Factor Branch, Molecular Biology Laboratory)

Aging is the major determinant of cancer incidence, which, in turn, is likely dictated in large part by processes that influence the progression of early subclinical (occult) cancers. However, there is little understanding of how aging informs changes in aggregate host signaling that favor cancer progression. In this study, we provide direct evidence that aging can serve as an organizing axis to define cancer progression-modulating processes. As a model system to explore this concept, we employed adolescent (68 days), young adult (143 days), middle-aged (551 days), and old (736 days) C57BL/6 mice as syngeneic hosts for engraftment of Lewis lung cancer to identify signaling and functional processes varying with host age. Older hosts exhibited dysregulated angiogenesis, metabolism, and apoptosis, all of which are associated with cancer progression. TGFβ1, a central player in these systemic processes, was downregulated consistently in older hosts. Our findings directly supported the conclusion of a strong host age dependence in determining the host tumor control dynamic. Furthermore, our results offer initial mechanism-based insights into how aging modulates tumor progression in ways that may be actionable for therapy or prevention.

Capturing the Driving Role of Tumor-Host Crosstalk in a Dynamical Model of Tumor Growth [4]
Team participant: Sebastien Benzekry
Other participants: Afshin Beheshti, Philip Hahnfeldt, Lynn Hlatky (CCSB - Center of Cancer and Systems Biology)

In 1999, Hahnfeldt et al. proposed a mathematical model for tumor growth as dictated by reciprocal communications between tumor and its associated vasculature, introducing the idea that a tumor is supported by a dynamic, rather than a static, carrying capacity. In this original work, the carrying capacity was equated with the variable tumor vascular support resulting from the net effect of tumor-derived angiogenesis stimulators and inhibitors. This dynamic carrying capacity model was further abstracted and developed in our recent publication to depict the more general situation where there is an interaction between the tumor and its supportive host tissue; in that case, as a function of host aging. This allowed us to predict a range of host changes that may be occurring with age that impact tumor dynamics. More generally, the basic formalism described here can be (and has been), extended to the therapeutic context using additional optimization criteria. The model depends on three parameters: one for the tumor cell proliferation kinetics, one for the stimulation of the stromal support, and one for its inhibition, as well as two initial conditions. We describe here the numerical method to estimate these parameters from longitudinal tumor volume measurements.

7.3.3. Metronomic oncology

Metronomic Reloaded: Theoretical Models Bringing Chemotherapy into the Era of Precision Medicine [5]
Team participant: Sebastien Benzekry
Other participants: Eddy Pasquier, Dominique Barbolosi, Joseph Ciccolini, Nicolas André (CRO2 - Centre de recherches en oncologie biologique et oncopharmacologie), Bruno Lacarelle (Clinical Pharmacokinetics), Fabrice Barlési (Service d'Oncologie Multidisciplinaire et d'Innovations Thérapeutiques)

Oncology has benefited from an increasingly growing number of groundbreaking innovations over the last decade. Targeted therapies, biotherapies, and the most recent immunotherapies all contribute to increase the number of therapeutic options for cancer patients. Consequently, substantial improvements in clinical outcomes for some disease with dismal prognosis such as lung carcinoma or melanoma have been achieved. Of note, the latest innovations in targeted therapies or biotherapies do not preclude the use of standard cytotoxic agents, mostly used in combination. Importantly, and despite the rise of bioguided (a.k.a. precision) medicine, the administration of chemotherapeutic agents still relies on the maximum tolerated drug (MTD) paradigm, a concept inherited from theories conceptualized nearly half a century ago. Alternative dosing schedules such as metronomic regimens, based upon the repeated and regular administration of low doses of chemotherapeutic drugs, have emerged as possible strategies to improve response rates while reducing toxicities. The recent changes in paradigm in the way we theorize cancer biology and evolution, metastatic spreading and tumor ecology, alongside the recent advances in the field of immunotherapy, have considerably strengthened the interest for metronomic approaches. This work aims at reviewing the recent evolutions in the field of theoretical biology of cancer and computational oncology, with a focus on the consequences these changes have on the way we administer chemotherapy. In particular, a step towards developing adaptive dosing should help to further optimize the efficacy of metronomic therapy. There is a rising trend to establish personalized medicine in oncology. Developing extensive bio-guided strategies for decision-making in the choice of drugs to be administered is now a common practice at the bedside. Similarly, developing extensive model-guided strategies for decision-making in refining dosing and scheduling should be undertaken to achieve precision medicine in oncology.

7.3.4. Protein-protein interaction networks

Design principles for cancer therapy guided by changes in complexity of protein-protein interaction networks [7]
Team participant: Sebastien Benzekry
Other participants: Jack A Tuszynski (Alberta University), Edward Rietman, Giannoula Lakka Klement (Newman-Lakka Institute)

The ever-increasing expanse of online bioinformatics data is enabling new ways to, not only explore the visualization of these data, but also to apply novel mathematical methods to extract meaningful information for clinically relevant analysis of pathways and treatment decisions. One of the methods used for computing topological characteristics of a space at different spatial resolutions is persistent homology. This concept can also be applied to network theory, and more specifically to protein-protein interaction networks, where the number of rings in an individual cancer network represents a measure of complexity. Results: We observed a linear correlation of R = -0.55 between persistent homology and 5-year survival of patients with a variety of cancers. This relationship was used to predict the proteins within a protein-protein interaction network with the most impact on cancer progression. By re-computing the persistent homology after computationally removing an individual node (protein) from the protein-protein interaction network, we were able to evaluate whether such an inhibition would lead to improvement in patient survival. The power of this approach lied in its ability to identify the effects of inhibition of multiple proteins and in the ability to expose whether the effect of a single inhibition may be amplified by inhibition of other proteins. More importantly, we illustrate specific examples of persistent homology calculations, which correctly predict the survival benefit observed effects in clinical trials using inhibitors of the identified molecular target. Conclusions: We propose that computational approaches such as persistent homology may be used in the future for selection of molecular therapies in clinic.

The technique uses a mathematical algorithm to evaluate the node (protein) whose inhibition has the highest potential to reduce network complexity. The greater the drop in persistent homology, the greater reduction in network complexity, and thus a larger potential for survival benefit. We hope that the use of advanced mathematics in medicine will provide timely information about the best drug combination for patients, and avoid the expense associated with an unsuccessful clinical trial, where drug(s) did not show a survival benefit.

7.4. Other new results

Superconvergent Cartesian Methods for Poisson type Equations in 2D–domains [21]
Team participants: Olivier Gallinato, Clair Poignard

In this work, we present three superconvergent Finite Difference methods on Cartesian grids for Poisson type equations with Dirichlet, Neumann or Robin conditions. Our methods are based on finite differences and high-order discretizations of the Laplace operator, to reach the superconvergence properties, in the sense that the first-order (and possibly the second-order) derivatives of the numerical solution are computed at the same order as the solution itself. We exhibit the numerical conditions that have to be fulfilled by the schemes to get such superconvergences and extensively illustrate our purpose by numerical simulations. We conclude by applying our method to a free boundary problem for cell protrusion formation recently proposed by the authors and colleagues. Note that quasistatic Stefan-like problem can be accurately solved by our methods.

Adaptive radiotherapy in routine: The radiation oncologist's point of view [14]
Team participant: Olivier Saut
Other participants: Bénédicte Henriques de Figueiredo, Adeline Petit, Paul Sargos, Guy Kantor, Claudia Pouypoudat, Christina Zacharatou, Mikael Antoine (Institut Bergonié, radiology department)

Adaptive radiotherapy is defined as all processes leading to the modification of a treatment plan on the basis of patient-specific variations observed during the course of a treatment. This concept is currently of particular relevance due to the development of onboard volumetric imaging systems, which allow for daily viewing of variations in both tumour and organs at risk in terms of position, shape or volume. However, its application in routine clinical practice is limited due to the demanding nature of the processes involved (re-delineation and replanning) and increased dependence on available human resources. Even if "online" strategies, based on deformable image registration (DIR) algorithms, could lead to a reduction in both work and calculation time, for the moment their use is limited to the research field due to uncertainties surrounding the validity of results gathered. Other strategies without DIR can be used as "offline" or "hybrid offline-online" strategies that seem to offer a compromise between time consumption and therapeutic gain for the patient.

MYCENAE Project-Team

6. New Results

6.1. Numerical and theoretical studies of slow-fast systems with complex oscillations

6.1.1. Canard-Mediated (De)Synchronization in Coupled Phantom Bursters

Participants: Elif Köksal Ersöz, Mathieu Desroches, Maciej Krupa, Frédérique Clément.

In [32], we study canard-mediated transitions in mutually coupled phantom bursters. We extend a multiple-timescale model which provides a sequence of dynamic events, i.e. transition from a frequency modulated relaxation cycle to a quasi-steady state and resumption of the relaxation regime through small amplitude oscillations. Folded singularities and associated canard solutions have a particular impact on the dynamics of the original system, which consists of two feedforward coupled FitzHugh-Nagumo oscillators, where the slow subsystem (regulator) controls the periodic behavior of the fast subsystem (secretor). We first investigate the variability in the dynamics depending on the canard mechanism that occurs near the folded singularities of the 4D secretor- regulator configuration. Then, we introduce a second secretor and focus on the slow-fast transitions in the presence of a linear coupling between the secretors. In particular, we explore the impact of the relationship between the canard structures and the coupling on patterns of synchronization and desynchronization of the collective dynamics of the resulting 6D system. We identify two different sources of desynchronization induced by canards, near a folded-saddle singularity and a folded-node singularity, respectively.

Part of these results have also been presented as posters at the *SIAM Conference on Applications of Dynamical Systems* (Snowbird, May 17-21, 2015) and 1st *International Conference on Mathematical Neuroscience* (Antibes Juan les Pins, June 8-10-2015).

6.1.2. Mixed-Mode Oscillations in a piecewise linear system with multiple time scale coupling

Participants: Soledad Fernández García, Maciej Krupa, Frédérique Clément.

We analyze a four dimensional slow-fast piecewise linear system with three time scales presenting Mixed-Mode Oscillations. The system possesses an attractive limit cycle along which oscillations of three different amplitudes and frequencies can appear, namely, small oscillations, pulses (medium amplitude) and one surge (largest amplitude). In addition to proving the existence and attractiveness of the limit cycle, we focus our attention on the canard phenomena underlying the changes in the number of small oscillations and pulses. We analyze locally the existence of secondary canards leading to the addition or subtraction of one small oscillation and describe how this change is globally compensated for or not with the addition or subtraction of one pulse.

6.1.3. Noise-induced canard and mixed-mode oscillations in large stochastic networks with multiple timescales

Participants: Jonathan Touboul, Maciej Krupa, Mathieu Desroches.

We investigate in [28] the dynamics of large stochastic networks with different timescales and nonlinear mean-field interactions. After deriving the limit equations for a general class of network models, we apply our results to the celebrated Wilson-Cowan system with two populations with or without slow adaptation, paradigmatic example of nonlinear mean-field network. This system has the property that the dynamics of the mean of the solution exactly satisfies an ODE. This reduction allows to show that in the mean-field limit and in multiple populations with multiple timescales, noise induces canard explosions and Mixed-Mode Oscillations on the mean of the solution. This sheds new light on the qualitative effects of noise and sensitivity to precise noise values in large stochastic networks. We further investigate finite-sized networks and show that systematic differences with the mean-field limits arise in bistable regimes (where random switches between different attractors occur) or in mixed-mode oscillations, were the finite-size effects induce early jumps due to the sensitivity of the attractor.

6.1.4. Canard explosion in delayed equations with multiple timescales, applications to the delayed Fitzhugh-Nagumo system

Participants: Maciej Krupa, Jonathan Touboul.

In two contributions, we investigated theoretically the presence of canard explosions of delayed differential equations, and have applied these results to the FitzHugh-Nagumo neuronal model.

- In [21] we analyze canard explosions in delayed differential equations with a one-dimensional slow manifold. This study is applied to explore the dynamics of the van der Pol slow-fast system with delayed self-coupling. In the absence of delays, this system provides a canonical example of a canard explosion. We show that as the delay is increased a family of 'classical' canard explosions ends as a Bogdanov-Takens bifurcation occurs at the folds points of the S-shaped critical manifold.
- Motivated by the dynamics of neuronal responses, we analyze in [21] the dynamics of the Fitzhugh-Nagumo slow-fast system with delayed self-coupling. Beyond the regime of small delays, delays significantly enrich the dynamics, leading to mixed-mode oscillations, bursting and chaos. These behaviors emerge from a delay-induced subcritical Bogdanov-Takens instability arising at the fold points of the S-shaped critical manifold. Underlying the transition from canard-induced to delay-induced dynamics is an abrupt switch in the nature of the Hopf bifurcation.

6.1.5. Canard-induced loss of stability across a homoclinic bifurcation

Participants: Mathieu Desroches, Jean-Pierre Françoise, Lucile Megret.

In [16], we investigate the possibility of bifurcations which display a dramatic change in the phase portrait in a very small (on the order of 10^{-7} in the example presented here) change of a parameter. We provide evidence of existence of such a very rapid loss of stability on a specific example of a singular perturbation setting. This example is strongly inspired of the explosion of canard cycles first discovered and studied by E. Benoît, J.-L. Callot, F. Diener and M. Diener. After some presentation of the integrable case to be perturbed, we present the numerical evidences for this rapid loss of stability using numerical continuation. We discuss then the possibility to estimate accurately the value of the parameter for which this bifurcation occurs.

6.1.6. Analysis of Interspike-Intervals for the General Class of Integrate-and-Fire Models with Periodic Drive

Participant: Justyna Signerska-Rynkowska.

In [27], we study one-dimensional integrate-and-fire models of the general type $\dot{x} = F(t, x)$ and analyze properties of the firing map which iterations recover consecutive spike timings. We impose very week constraints for the regularity of the function $F(t, x)$ e.g. often it suffices to assume that F is continuous. If additionally F is periodic in t, using mathematical study of the displacement sequence of an orientation preserving circle homeomorphism, we provide a detailed description of the regularity properties of the sequence of interspike-intervals and behaviour of the interspike-interval distribution.

6.1.7. A geometric mechanism for mixed-mode bursting oscillations in a hybrid neuron model

Participants: Justyna Signerska-Rynkowska, Jonathan Touboul, Alexandre Vidal.

In [35], we exhibit and investigate a new type of mechanism for generating complex oscillations featuring an alternation of small oscillations with spikes (MMOs) or bursts (MMBOs) in a class of hybrid dynamical systems modeling neuronal activity. These dynamical systems, called nonlinear adaptive integrate-and-fire neurons, combine nonlinear dynamics modeling input integration in a nerve cell with discrete resets modeling the emission of an action potential and the subsequent return to reversal potential. We show that presence of complex oscillations in these models relies on a fundamentally hybrid structure of the flow: invariant manifolds of the continuous dynamics govern small oscillations, while discrete resets govern the emission of spikes or bursts. The decomposition into these two mechanisms leads us to propose a purely geometrical interpretation of these complex trajectories, and this relative simplicity allows to finely characterize the MMO patterns through the study of iterates of the adaptation map associated with the hybrid system. This map is however

singular: it is discontinuous and has unbounded left- and right-derivatives. We apply and develop rotation theory of circle maps for this class of adaptation maps to precisely characterize the trajectories with respect to the parameters of the system. In contrast to more classical frameworks in which MM(B)Os were evidenced, the present geometric mechanism neither requires no more than two dimensions, does not necessitate to have separation of timescales nor complex return mechanisms.
Part of these results have also been presented as posters at the *SIAM Conference on Applications of Dynamical Systems* (Snowbird, May 17-21, 2015) and 1st *International Conference on Mathematical Neuroscience* (Antibes Juan les Pins, June 8-10-2015).

6.2. Non conservative transport equations for cell population dynamics

6.2.1. Cell-kinetics based calibration of a multiscale model: application to cell population dynamics in ovarian follicles

Participants: Benjamin Aymard [ICL], Frédérique Clément, Danielle Monniaux [INRA], Marie Postel.

In [30], we present a strategy for tuning the parameters of a multiscale model of structured cell populations in which physiological mechanisms are embedded into the cell scale. This strategy allows one to cope with the technical difficulties raised by such models, that arise from their anchorage in cell biology concepts: localized mitosis, progression within and out of the cell cycle driven by time- and possibly unknown-dependent, and nonsmooth velocity coefficients. We compute different mesoscopic and macroscopic quantities from the microscopic unknowns (cell densities) and relate them to experimental cell kinetic indexes. We study the expression of reaching times corresponding to characteristic cellular transitions in a particle-like reduction of the original model. We make use of this framework to obtain an appropriate initial guess for the parameters and then perform a sequence of optimization steps subject to quantitative specifications. We finally illustrate realistic simulations of the cell populations in cohorts of interacting ovarian follicles.

6.2.2. Dimensional reduction of a multiscale cell population model

Participants: Frédérique Clément, Frédéric Coquel [CMAP], Marie Postel, Kim Long Tran.

We have designed a dimensional reduction of a multiscale structured cell population model, consisting of a system of 2D transport equations, into a system of twice as many 1D transport equations. The reduced model is obtained by computing the moments of the 2D model with respect to one space variable. The 1D solution is defined from the solution of the 2D model starting from an initial condition that is a Dirac mass in the direction removed by reduction. Long time properties of the 1D model solution are obtained in connection with properties of the support of the 2D solution for general case initial conditions. Finite volume numerical approximations of the 1D reduced model can be used to compute the moments of the 2D solution with satisfying accuracy. The numerical robustness is studied in the scalar case and a full scale vector case is presented.

6.3. Macroscopic limits of stochastic neural networks and neural fields

6.3.1. Pinwheel-Dipole configuration in cat visual cortex

Participants: Jérôme Ribot [CIRB], Alberto Romagnoni [CIRB], Chantal Milleret [CIRB], Daniel Bennequin [CIRB], Jonathan Touboul.

One fascinating aspect of the brain is its ability to process information in a fast and reliable manner. The functional architecture is thought to play a central role in this task, by encoding efficiently complex stimuli and facilitating higher level processing. In the early visual cortex of higher mammals, information is processed within functional maps whose layout is thought to underlie visual perception. The possible principles underlying the topology of the different maps, as well as the role of a specific functional architecture on information processing, is however poorly understood.

- In [25], we show that spatial frequency representation in cat areas 17 and 18 exhibits singularities around which the map organizes like an electric dipole potential. These singularities are precisely co-located with singularities of the orientation map: the pinwheel centers. We first show, using high resolution optical imaging, that a large majority (around 80%) of pinwheel centers exhibit in their neighborhood semi-global extrema in the spatial frequency map. These extrema created a sharp gradient that was confirmed with electrophysiological recordings. Based on an analogy with electromagnetism, a mathematical model of a dipolar structure is proposed, that was accurately fitted to optical imaging data for two third of pinwheel centers with semi-global extrema.
- Mathematically, this pinwheel-dipole architecture is fascinating. We demonstrated mathematically in [26] that two natural principles, local exhaustivity of representation and parsimony, would indeed constrain the orientation and spatial frequency maps to display co-located singularities around which the orientation is organized as a pinwheel and spatial frequency as a dipole. Moreover, using a computational model, we showed that this architecture allows a trade-off in the local perception of orientation and spatial frequency, but this would occur for sharper selectivity than the tuning width reported in the literature. We therefore re-examined physiological data and show that indeed the spatial frequency selectivity substantially sharpens near maps singularities, bringing to the prediction that the system tends to optimize balanced detection between different attributes.

These results shed new light on the principles at play in the emergence of functional architecture of cortical maps, as well as their potential role in processing information.

6.3.2. Absorption properties of stochastic equations with Hölder diffusion coefficients

Participants: Jonathan Touboul, Gilles Wainrib [ENS].

In [29], we address the absorption properties of a class of stochastic differential equations around singular points where both the drift and diffusion functions vanish. According to the Hölder coefficient alpha of the diffusion function around the singular point, we identify different regimes. Stability of the absorbing state, large deviations for the absorption time, existence of stationary or quasi-stationary distributions are discussed. In particular, we show that quasi-stationary distributions only exist for alpha < 3/4, and for alpha in the interval (3/4, 1), no quasi-stationary distribution is found and numerical simulations tend to show that the process conditioned on not being absorbed initiates an almost sure exponential convergence towards the absorbing state (as is demonstrated to be true for alpha = 1). Applications of these results to stochastic bifurcations are discussed.

6.3.3. On a kinetic FitzHugh-Nagumo model of neuronal network

Participants: Stéphane Mischler [CEREMADE], Cristóbal Quiñinao [CIRB], Jonathan Touboul.

We investigate in [33] the existence and uniqueness of solutions of a McKean-Vlasov evolution PDE representing the macroscopic behavior of interacting Fitzhugh-Nagumo neurons. This equation is hypoelliptic, nonlocal and has unbounded coefficients. We proved existence of a solution to the evolution equation and non trivial stationary solutions. Moreover, we demonstrated uniqueness of the stationary solution in the weakly nonlinear regime. Eventually, using a semigroup factorisation method, we showed exponential nonlinear stability in the small connectivity regime.

6.4. Modeling of neurogenesis and brain development

6.4.1. Lhx2 regulates the timing of β-catenin-dependent cortical neurogenesis

Participants: Lea-Chia-Ling Hsu [Taipei], Sean Nama [Taipei], Yi Cui, Ching-Pu Chang [Taipei], Chia-Fang Wang [Taipei], Hung-Chih Kuo [Taipei], Jonathan Touboul, Shen-Ju Chou [Taipei].

The timing of cortical neurogenesis has a major effect on the size and organization of the mature cortex. The deletion of the LIM-homeodomain transcription factor Lhx2 in cortical progenitors by Nestin-cre leads to a dramatically smaller cortex. In [19] we report that Lhx2 regulates the cortex size by maintaining the cortical progenitor proliferation and delaying the initiation of neurogenesis. The loss of Lhx2 in cortical progenitors results in precocious radial glia differentiation and a temporal shift of cortical neurogenesis. We further investigated the underlying mechanisms at play and demonstrated that in the absence of Lhx2, the Wnt/β-catenin pathway failed to maintain progenitor proliferation. We developed and applied a mathematical model that reveals how precocious neurogenesis affected cortical surface and thickness. Thus, we concluded that Lhx2 is required for β-catenin function in maintaining cortical progenitor proliferation and controls the timing of cortical neurogenesis.

6.4.2. Competition and boundary formation in heterogeneous media: Application to neuronal differentiation

Participants: Cristóbal Quiñinao [CIRB], Benoît Perthame [LJLL], Jonathan Touboul.

We analyze in [22] an inhomogeneous system of coupled reaction-diffusion equations representing the dynamics of gene expression during differentiation of nerve cells. The outcome of this developmental phase is the formation of distinct functional areas separated by sharp and smooth boundaries. It proceeds through the competition between the expression of two genes whose expression is driven by monotonic gradients of chemicals, and the products of gene expression undergo local diffusion and drive gene expression in neighboring cells. The problem therefore falls in a more general setting of species in competition within a non-homogeneous medium. We show that in the limit of arbitrarily small diffusion, there exists a unique monotonic stationary solution, which splits the neural tissue into two winner-take-all parts at a precise boundary point: on both sides of the boundary, different neuronal types are present. In order to further characterize the location of this boundary, we use a blow-up of the system and define a traveling wave problem parametrized by the position within the monotonic gradient: the precise boundary location is given by the unique point in space at which the speed of the wave vanishes.

6.4.3. Local homeoprotein diffusion can stabilize boundaries generated by graded positional cues

Participants: Cristóbal Quiñinao [CIRB], Alain Prochiantz [CIRB], Jonathan Touboul.

Boundary formation in the developing neuroepithelium decides on the position and size of compartments in the adult nervous system. In [23], we started from the French Flag model proposed by Lewis Wolpert, in which boundaries are formed through the combination of morphogen diffusion and of thresholds in cell responses. In contemporary terms, a response is characterized by the expression of cell-autonomous transcription factors, very often of the homeoprotein family. Theoretical studies suggest that this sole mechanism results in the formation of boundaries of imprecise shapes and positions. Alan Turing, on the other hand, proposed a model whereby two morphogens that exhibit self-activation and reciprocal inhibition, and are uniformly distributed and diffuse at different rates lead to the formation of territories of unpredictable shapes and positions but with sharp boundaries (the 'leopard spots'). Here, we have combined the two models and compared the stability of boundaries when the hypothesis of local homeoprotein intercellular diffusion is, or is not, introduced in the equations. We find that the addition of homeoprotein local diffusion leads to a dramatic stabilization of the positioning of the boundary, even when other parameters are significantly modified. This novel Turing/Wolpert combined model has thus important theoretical consequences for our understanding of the role of the intercellular diffusion of homeoproteins in the developmental robustness of and the changes that take place in the course of evolution.

6.4.4. Designing a mathematical model of the dynamics of progenitor cell populations in the mouse cerebral cortex

Participants: Marie Postel, Alice Karam [UPMC], Mérina Latbi [UPMC], Guillaume Pezeron [UPMC], Kim Long Tran, Frédérique Clément, Sylvie Schneider-Maunoury [UPMC].

The mammalian cortex is a laminar structure in the dorsal telencephalon, composed of distinct cell types with different spatial and temporal origins. Cortical projection neurons display different patterns of layering and connectivity that depend on their birth date. We have designed a multi-scale mathematical model of structured cell populations, taking into account three main cell types: apical progenitors (APs), intermediate progenitors (IPs) and neurons (N). APs self-renew and produce IPs that divide to give Ns. The main originality of this spatio-temporal model is to explicitly represent the different phases of the cell cycle, G1, S, G2 and M. Biological data from the experiments and from the literature provide values for parameters of the model (e.g. duration of each cell cycle phase and division rates for each cell type). The outputs of the model are interpretable in terms of cell kinetics (e.g. mitotic index, labelling index, cell numbers). They are adjusted to experimental observations by numerical simulation.

NUMED Project-Team (section vide)

REO Project-Team

7. New Results

7.1. Mathematical and numerical analysis of fluid-structure interaction problems

Participants: Matteo Aletti, Faisal Amlani, Benoit Fabrèges, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Mikel Landajuela Larma, Damiano Lombardi, Marina Vidrascu.

In [55] we present a numerical study in which several partitioned solution procedures for incompressible fluid-structure interaction are compared and validated against the results of an experimental FSI benchmark. The numerical methods discussed cover the three main families of coupling schemes: strongly coupled, semi-implicit and loosely coupled. Very good agreement is observed between the numerical and experimental results. The comparisons confirm that strong coupling can be efficiently avoided, via semi-implicit and loosely coupled schemes, without compromising stability and accuracy.

In [14] we introduce a Nitsche-XFEM method for fluid-structure interaction problems involving a thin-walled elastic structure (Lagrangian formalism) immersed in an incompressible viscous fluid (Eulerian formalism). The fluid domain is discretized with an unstructured mesh not fitted to the solid mid-surface mesh. Weak and strong discontinuities across the interface are allowed for the velocity and pressure, respectively. The fluid-solid coupling is enforced consistently using a variant of Nitsche's method with cut-elements. Robustness with respect to arbitrary interface intersections is guaranteed through suitable stabilization. Several coupling schemes with different degrees of fluid-solid time splitting (implicit, semi-implicit and explicit) are investigated. A series of numerical tests in 2D, involving static and moving interfaces, illustrates the performance of the different methods proposed.

In [15] we investigated the autoregulation in the retinal haemodynamics by means of three-dimensional simulations. The autoregulation is a key phenomenon from a physiological standpoint, consisting in the ability of the vasculature to control the flow in different pressure conditions. A simplified fluid-structure interaction method was devised in order to render the vessels wall contraction in a large network, with an affordable computational cost. Several test cases were performed on a patient-specific arteriolar network, whose geometry was reconstructed by using fundus camera images. The tests were in agreement with experimental trends and confirm the ability of the approach to reproduce the phenomena involved.

In [33] we study an unsteady nonlinear fluid-structure interaction problem which is a simplified model to describe blood flow through viscoelastic arteries. We consider a Newtonian incompressible two-dimensional flow described by the Navier-Stokes equations set in an unknown domain depending on the displacement of a structure, which itself satisfies a linear viscoelastic beam equation. The fluid and the structure are fully coupled via interface conditions prescribing the continuity of the velocities at the fluid-structure interface and the action-reaction principle. We prove that strong solutions to this problem are global-in-time. We obtain in particular that contact between the viscoleastic wall and the bottom of the fluid cavity does not occur in finite time. To our knowledge, this is the first occurrence of a no-contact result, but also of existence of strong solutions globally in time, in the frame of interactions between a viscous fluid and a deformable structure.

In [27] and [45] we study the effect of wall bending resistance on the motion of an initially spherical capsule freely suspended in shear flow or in a a planar hyperbolic flow. We consider a capsule with a given thickness made of a three–dimensional homogeneous elastic material. A numerical method is used to model the coupling of a boundary integral method for the fluids with a shell finite element method for the capsule envelope. For a given wall material, the capsule deformability strongly decreases when the wall bending resistance increases. In addition, if one expresses the same results as a function of the two–dimensional mechanical properties of the mid–surface, which is how the capsule wall is modeled in the thin–shell model, the capsule deformed shape is identical to the one predicted for a capsule devoid of bending resistance. The bending rigidity is found to have

a negligible influence on the overall deformation of an initially spherical capsule, which therefore depends only on the elastic stretching of the mid–surface. Still, the bending resistance of the wall must be accounted for to model the buckling phenomenon, which is observed locally at low flow strength and persist at steady state. We show that the wrinkle wavelength only depends on the bending number, which compares the relative importance of bending and shearing phenomena, and provide the correlation law. Such results can then be used to infer values of the bending modulus and wall thickness from experiments on spherical capsules in simple shear flow.

In [57] we consider the motion of an elastic structure represented by the nonlinear Saint-Venant Kirchhoff model immersed in a compressible fluid modeled by the compressible Navier-Stokes equations. Existence and uniqueness of a regular solution defined locally in time is proved.

7.2. Numerical methods for biological flows

Participants: Chloé Audebert, Benoit Fabrèges, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Sanjay Pant, Marc Thiriet, Irène Vignon-Clementel.

In [37] we present a closed-loop global lumped parameter model for pre stage-II single-ventricle physiology. This model, which is built on a fibre mechanics based description of the heart chambers, benefits from a novel method to describe regurgitant valves. As many as 33 model parameters are estimated from uncertain clinical measurements in two patients—with and without atrioventricular valve regurgitation—through the method of data assimilation. Results are validated qualitatively through measurements and clinical estimates that were not included in the parameter estimation procedure. The methods are shown to successfully capture patient-specific clinical observations such as double peaked nature of valvular flows and abnormalities in electrocardiogram readings.

In [39] we propose a methodology for full propagation of uncertainty from clinical data to model results that enables estimation of the confidence associated with model predictions. We illustrate this problem in a pre stage-II single-ventricle physiology, for which coherence of simulations and clinical data indicated that the flow split to the right lung was highly uncertain. We want to assess here how such uncertainty translates into surgical planning of removing the stenosis or not. Taking into account the effect of the rest of the circulation is also studied in the uncertainty propagation.

In [21] 3D blood flow simulations are carried out for the design of a stented valve reducer in enlarged ventricular outflow tracts. Different device designs are built and compared with the initial device-free state, or with the reducer alone. Results suggest that pressure loss is higher for the reducer alone than for the full device, and that the latter successfully restores hemodynamics to a healthy state. Pressure forces on the reducer and on the valve have the same magnitudes. Migration would occur towards the right ventricle rather than the pulmonary arteries.

In [44] we aim at developing a mathematical model in order to reproduce hemodynamics changes due to liver ablation surgeries. First, a 0D closed-loop model is developed, to simulate hepatectomy and compute post-operative average values. Due to the closed loop, the surgery impact both on and from the whole circulation can be captured, including bleeding and infusion. Then, a one-dimensional artery model is implemented to improve the closed-loop model and simulate better the changes in arterial waveforms due to surgery.

In [54] we investigate the spatial and time discretization of the transient Oseen equations. Finite elements with symmetric stabilization in space are combined with several time-stepping schemes (monolithic and fractional-step). Quasi-optimal (in space) and optimal (in time) error estimates are established for smooth solutions in all flow regimes. We first analyze monolithic time discretizations using the Backward Differentation Formulas of order 1 and 2 (BDF1 and BDF2). We derive a new estimate on the time-average of the pressure error featuring the same robustness with respect to the Reynolds number as the velocity estimate. Then, we analyze fractional-step pressure-projection methods using BDF1. The stabilization of velocities and pressures can be treated either implicitly or explicitly. Numerical results illustrate the main theoretical findings.

In [26] we study the effects of inserted needle on the subcutaneous interstitial flow. A goal is to describe the physical stress affecting cells during acupuncture treatment. The model consists of the convective Brinkman equations to describe the flow through a fibrous medium. Numerical studies in FreeFem++ are performed to illustrate the acute physical stress developed by the implantation of a needle that triggers the physiological reactions of acupuncture. We emphasize the importance of numerical experiments for advancing in modeling in acupuncture. In [40] we show that the acupoint must contain a highly concentrated population of mastocytes (e.g., very-high–amplitude, small-width Gaussian distribution) to get an initial proper response. Permanent signaling is provided by chemotaxis and continuous recruitment of mastocytes. Therefore, the density and distribution of mastocytes are crucial factors for efficient acupuncture as well as availability of circulating and neighboring pools of mastocytes.

In [61] we carry out a three-dimensional blood flow simulation through a complete macrovascular circuit, the cerebral venous network, rather than using reduced order simulation and partial vascular network. The bio-mechanical modeling step is carefully performed and leads to the description of the flow governed by the Navier-Stokes equations for an incompressible viscous fluid. We then numerically solve the equations with a free finite element software in five meshes of a realistic geometry obtained from medical images to prove the feasibility of the pipeline. Some particularities of the venous network, as asymmetry for example, are discussed.

7.3. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi.

In [58] we investigate the monodomain equation which describes the evolution of the cardiac electrical potential and which corresponds to a coupled system involving a reaction-diffusion equation and an ordinary differential equation. Lipschitz stability inequalities are shown for the identification of some parameters of the model from measurements on the cardiac potential and the ionic variable.

In [32] we studied the application of a Reduced-Order Modeling method (Approximated Lax Pairs) to the solution of the partial differential equations describing the polarisation of tissues. Due to the complexity of the scenarios involved and the presence of propagating waves, the performances of the standard methods proposed in the literature to provide a low computational cost solution are not always satisfactory. The ALP method consists of the construction of an adaptive time dependent basis that diagonalises, at each time, a Schrödinger-type operator. Its application to several 2D and 3D test-cases on the equations arising in electrophysiology was investigated, showing that the performances of the method in terms of speed-up and accuracy are promising.

In [62] we considered the simulation of full cycles of the electrical activity of the heart and the corresponding body surface potential. The model is based on a realistic torso and heart anatomy, including ventricles and atria. One of the specificities of our approach is to model the atria as a surface, which is the kind of data typically provided by medical imaging for thin volumes. The bidomain equations are considered in their usual formulation in the ventricles, and in a surface formulation on the atria. Two ionic models are used: the Courtemanche-Ramirez-Nattel model on the atria, and the " Minimal model for human Ventricular action potentials " (MV) by Bueno-Orovio, Cherry and Fenton in the ventricles. The heart is weakly coupled to the torso by a Robin boundary condition based on a resistor-capacitor transmission condition. Various ECGs are simulated in healthy and pathological conditions (left and right bundle branch blocks, Bachmann's bundle block, Wolff-Parkinson-White syndrome). To assess the numerical ECGs, we use several qualitative and quantitative criteria found in the medical literature. Our simulator can also be used to generate the signals measured by a vest of electrodes. This capability is illustrated at the end of the article.

In [24] we address the inverse problem of electrocardiography from a new perspective, by combining electrical and mechanical measurements. Our strategy relies on the definition of a model of the electromechanical contraction which is registered on ECG data but also on measured mechanical displacements of the heart tissue typically extracted from medical images. In this respect, we establish in this work the convergence of a sequential estimator which combines for such coupled problems various state of the art sequential data

assimilation methods in a unified consistent and efficient framework. Indeed we aggregate a Luenberger observer for the mechanical state and a Reduced Order Unscented Kalman Filter applied on the parameters to be identified and a POD projection of the electrical state. Then using synthetic data we show the benefits of our approach for the estimation of the electrical state of the ventricles along the heart beat compared with more classical strategies which only consider an electrophysiological model with ECG measurements. Our numerical results actually show that the mechanical measurements improve the identifiability of the electrical problem allowing to reconstruct the electrical state of the coupled system more precisely. Therefore, this work is intended to be a first proof of concept, with theoretical justifications and numerical investigations, of the advantage of using available multi-modal observations for the estimation and identification of an electromechanical model of the heart.

7.4. Lung and respiration modeling

Participants: Laurent Boudin, Muriel Boulakia, Céline Grandmont, Jessica Oakes, Nicolas Pozin, Irène Vignon-Clementel.

In silico models of flow and transport in the lung are increasingly being used to predict regional deposition in healthy and diseased lungs. However, very few models have been validated with in vivo human or animal experimental data. In [36], we create a physiologically-based simulation of airflow and particle transport in healthy and emphysematous rat lungs. Excellent agreement between the numerical predictions and experimental data is found for the healthy lungs. However, the numerical predictions are unable to predict the experimental findings of enhanced deposition in the normal regions of the emphysematous lungs and thus more sophisticated models of transport in the deep regions of the lung are needed. This is what is being explored in [42], where interactions of flow and transport between 3D upper-parts and 1D downstream respiratory trees are captured for inspiration and expiration for the first time.

While several groups have investigated detailed flow and particle transport in the acinar regions of the healthy lung, little is currently known about diseased acini. In [35] we perform numerical simulations of flow and transport in healthy and emphysematous acini. As the alveolar septa is deteriorated in emphysema there is less surface area available for particles to deposit on. Therefore, fewer particles deposit in the diseased models. In addition, we find that particle deposition is more heterogeneously distributed in emphysema, a phenomenon that was also found in the in vivo animal experiments.

7.5. Methods for the interaction data - simulation

Participants: Jean-Frédéric Gerbeau, Damiano Lombardi, Sanjay Pant, Irène Vignon-Clementel.

In [38] we proposed an information theoretical framework to study the practical identifiability of dynamical systems. The fundamental question arising in parameter estimation problems is whether, given a set of observations of the system, it is possible to retrieve the parameters values. The method proposed exploits a database of direct numerical simulations and study the parameters-to-observables map by means of differential entropies. Contrary to other approaches proposed in the literature it is not restricted to ordinary differential equations and take the experimental noise into account. Several test cases were performed on a large spectrum of bio-physical systems, providing promising results.

In [60] we studied a differential entropy estimator based on $kp-$neighbours, aiming at applying a Bayesian framework and some information-theoretic ideas to inverse problems. The goal of this work is to estimate the Shannon differential entropy in high dimensional settings, in possible presence of functional or nearly functional dependences. A modification of the Kozachenko-Leonenko estimator is proposed, consisting of introducing a local gaussian approximation to the probability measure. Test-cases were performed to assess the properties of the method and to compare its performances with other methods proposed in the literature.

The articles [37], presented in the section about biological flows, and [24], presented in the section about electrophysiology, also present methods concerning the interaction data - simulation.

7.6. Miscellaneous

Participants: Laurent Boudin, Irène Vignon-Clementel.

In [34] we develop a quantitative single cell-based model for multi-cellular tumor spheroids of a specific lung cancer cell line, growing under various nutrient conditions: we confront the simulations performed with this model with data on the growth kinetics and spatial labeling patterns for cell proliferation, extracellular matrix, cell distribution and cell death. We stepwise arrive at a model that mimics the spheroid growth under two conditions, and can predict two other ones. The number of mechanisms the model contains is necessary and sufficient to explain the data.

In [19] we consider a kinetic model describing some mechanisms of opinion formation in the framework of referendums, where the individuals, who can interact between themselves and modify their opinion by means of spontaneous self-thinking, are moreover under the influence of mass media. We study, at the numerical level, both the transient and the asymptotic regimes. In particular, we point out that a plurality of media, with different orientations, is a key ingredient to allow pluralism and prevent consensus. The forecasts of the model are compared to some surveys related to the Scottish independence referendum of 2014.

In [56] we review various results on the compactness of the linearized Boltzmann collision operator and of its generalization to mixtures of non-reactive monatomic gases.