RESEARCH CENTER

FIELD
**Computational Sciences for Biology, Medicine and the Environment**

# Activity Report 2012

# Section New Results

Edition: 2013-04-24

COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

COMPUTATIONAL MEDICINE AND NEUROSCIENCES

OBSERVATION AND MODELING FOR ENVIRONMENTAL SCIENCES

OBSERVATION, MODELING, AND CONTROL FOR LIFE SCIENCES

# ABS Project-Team

# 5. New Results

## 5.1. Modeling Interfaces and Contacts

Docking, scoring, interfaces, protein complexes, scoring functions, Voronoi diagrams, arrangements of balls.

### *5.1.1. Modeling Macro-molecular Complexes : a Journey Across Scales*

**Participants:** Frédéric Cazals, Tom Dreyfus.

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

While proteins and nucleic acids are the fundamental components of an organism, Biology itself is based on the interactions they make with each other. Analyzing macromolecular interactions typically requires handling systems involving from two to hundreds of polypeptide chains. After a brief overview of the modeling challenges faced in computational structural biology, the text [16] reviews concepts and tools aiming at improving our understanding of the link between the static structures of macromolecular complexes and their biophysical/biological properties. We discuss geometrical approaches suited to atomic-resolution complexes and to large protein assemblies; for each, we also present examples of their successful application in quantifying and interpreting biological data. This methodology includes state-of-the-art geometric analyses of surface area, volume, curvature, and topological properties (isolated components, cavities, voids, cycles) related to Voronoi constructions in the context of structure analysis. On the applied side, we present novel insights into real biological problems gained thanks to these modeling tools.

### *5.1.2. CSA: Comprehensive Comparison of Pairwise Protein Structure Alignments*

**Participant:** Noël Malod-Dognin.

*In collaboration with I. Wohlers (CWI / VU University Amsterdam, Netherlands), R. Andonov (Irisa / Rennes University, France), G.W. Klau (CWI / VU University Amsterdam, Netherlands).*

Protein structural alignment is a key method for answering many biological questions involving the transfer of information from well-studied proteins to less well-known proteins. Since structures are more conserved during evolution than sequences, structural alignment allows for the most precise mapping of equivalent residues. Many structure-based scoring schemes have been proposed and there is no consensus on which scoring is the best. Comparative studies also show that alignments produced by different methods can differ considerably. Based on the alignment engine derived from A_purva, we designed CSA (Comparative Structural Alignment), the first web server for computation, evaluation and comprehensive comparison of pairwise protein structure alignments at single residue level [15]. It offers the exact computation of alignments using the scoring schemes of DALI, Contact Map Overlap (CMO), MATRAS and PAUL. In CSA, computed or uploaded alignments can be explored in terms of many inter-residue distances, RMSD, and sequence-based scores. Intuitive visualizations also help in grasping the agreements and differences between alignments. The user can thus make educated decisions about the structural similarity of two proteins and, if necessary, post-process alignments by hand. CSA is available at http://csa.project.cwi.nl.

Upon publication [15], CSA was selected by *Nucleic Acids Research* as featured article of July 2012 (top 5% of papers in terms of originality, significance and scientific excellence.

## 5.2. Modeling Macro-molecular Assemblies

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

### 5.2.1. Stoichiometry Determination for Mass-spectrometry Data: the Interval Case

**Participants:** Deepesh Agarwal, Frédéric Cazals, Noël Malod-Dognin.

In structural proteomics, given the individual masses of a set of protein types and the exact mass of a protein complex, the *exact stoichiometry determination problem (SD)*, also known as the money-change problem, consists of enumerating all the stoichiometries of these types which allow to recover the target mass. If the target mass suffers from experimental uncertainties, the *interval SD problem* consists of finding all the stoichiometry vectors compatible with a target mass within an interval.

We make contributions in two directions [18]. From a theoretical standpoint, we present a constant-memory space algorithm (`DIOPHANTINE`) and an output sensitive dynamic programming based algorithm (`DP++`), both inherently addressing the interval SD problem. From an applied perspective, we raise three points. First, we show that `DIOPHANTINE` and `DP++` yield an improvement from 3 to 4 orders of magnitude over state-of-the-art exact SD algorithms, for typical protein complexes facing uncertainties on the target mass in the range 0.1-1%. Second, we show that `DIOPHANTINE` behaves like an output-sensitive algorithm—especially when the interval width increases, albeit such a property cannot be expected in general. Third, from a biological perspective, using a panel of biological complexes (eukaryotic translation factor, yeast exosome, 19S proteasome sub-unit, nuclear pore complex), we stress the importance of enumeration, even at a null noise level.

The programs accompanying this paper are available from http://team.inria.fr/abs/addict/.

## 5.3. Algorithmic Foundations

Voronoi diagrams, $\alpha$-shapes,

The work undertaken in this vein in 2012 will be finalized in 2013.

## 5.4. Immunology

Immune response, infection, antibodies, complementarity determining region (CDR)

### 5.4.1. Teleost Fish Mount Complex Clonal IgM and IgT Responses in Spleen Upon Systemic Viral Infection

**Participant:** Frédéric Cazals.

*In collaboration with*

- *R. Castro, L. Journeau, A. Benmansour and P. Boudinot (INRA Jouy-en-Josas, France)*
- *H.P. Pham and A. Six (Univ. of Paris VI, France)*
- *O. Bouchez (INRA Castanet Tolosan, France)*
- *V. Giudicelli and M-P. Lefranc (IMGT / CNRS, Montpellier, France)*
- *E. Quillet (INRA Jouy-en-Josas, France)*
- *S. Fillatreau (Leibniz Institute, Berlin, Germany)*
- *O. Sunyer (Univ. of Pennsylvania, USA)*

Upon infection, B-lymphocytes expressing antibodies specific for the intruding pathogen develop clonal responses triggered by pathogen recognition via the B-cell receptor. The constant region of antibodies produced by such developing clones dictates their functional properties. In teleost fish, the clonal structure of B-cell responses and the respective contribution of the three isotypes IgM, IgD, and IdT remains unknown. The expression of IgM and IgT are mutually exclusive, leading to the existence of two B-cell subsets expressing either both IgM and IgD or only IgT. In [12], we undertook a comprehensive analysis of the variable heavy chain (VH) domain repertoires of the IgM, IgD and IgT in spleen of homozygous isogenic rainbow trout (Onchorhynus mykiss), before and after challenge with a rhabdovirus, the Viral Hemorrhagic Septicemia Virus (VHSV), using CDR3-length spectratyping and pyrosequencing of immunoglobulin (Ig) transcripts. In healthy fish, we observed distinct repertoires for IgM, IgD and IgT respectively, with a few amplified $\mu$

and $\tau$ junctions, suggesting the presence of IgM and IgT secreting cells in the spleen. In infected animals, we detected complex and highly diverse IgM responses involving all VH subgroups, and dominated by a few large public and private clones. A lower number of robust clonal responses involving only a few VH were detected for the mucosal IgT, indicating that both IgM$^+$ and IgT$^+$ spleen B cells responded to systemic infection but at different degrees. In contrast, the IgD response to the infection was faint. Although the IgD and IgT present different structural features and evolutionary origin compared to mammalian IgD and IgA respectively, their implication in the B-cell response evokes these mouse and human counterparts. Thus, it appears that the general properties of antibody responses were already in place in common ancestors of fish and mammals, and were globally conserved during evolution with possible functional convergences.

## AMIB Project-Team

# 5. New Results

## 5.1. RNA structures

### 5.1.1. RNA structure alignment

It is widely accepted that, for a large number of RNA families, the structure is more conserved than the sequence. Therefore, any reasonable notion of homology should consider the similarity in the secondary structure, i.e. how well the base-pairing positions in two structures can be put in correspondence, or aligned. In collaboration with a significant part of the French bioinformatics community, an assessment of the quality of existing algorithms for the problem was proposed [6]. Furthermore, a review of the state-of-the-art in RNA comparison algorithms is to be published [11], and a chapter in a forthcoming book on RNA computational biology was written in collaboration with Robert Giegerich (University Bielefeld) during his stay.

Most existing alignment tools rely on the assumption that the RNA structure is free of pseudoknots, i.e. free of crossing interactions. This condition naturally arises from the intractability of the unconstrained version of the problem. In a joint work, A. Denise, Ph. Rinaudo and Y. Ponty worked around this issue by proposing a parameterized complexity algorithmic solution for the unconstrained version of the problem. One of the key feature of this algorithm is that, although exponential in the worst-case scenario, it naturally adapts its complexity to the level of intricacy of the aligned structures, and remains polynomial for large classes of pseudoknots. Preliminary results of this work were presented at the WABI'12 conference [35].

### 5.1.2. Energy-weighted RNA algorithmics

We complemented previous studies led within AMIB on RNA structures with restricted classes of pseudoknots by showing, in a collaboration with Rolf Backofen (Freiburg University), that the computational hardness of RNA folding with general pseudoknots is extremely robust to the choice of a precise energy model. It was shown that the problem is completely unapproximable when expressive – yet realistic – energy models are taken into consideration. These results were presented at CPM'12 [37] (Helsinki, Finland).

Moreover, using an interpolation technique introduced at the RECOMB'11 conference, we were able to improve both the sequential and parallel complexities of the `RNAbor` algorithm developed within P. Clote's lab. The resulting algorithm and its application to the detection of conformational switches in sequence lengths that were previously unreachable by the algorithm, are described in a manuscript accepted in *Plos One*.

### 5.1.3. RNA knowledge-based potentials and 3D studies

The building of an RNA potential proved much harder and interesting than we initially expected. A non-redundant dataset had first to be extracted from the literature as the available dataset were not suitable for our study even the very recent ones. From the collected distance data, the building of a knowledge-based potential was usually done using histograms; and the histogram interval size and data fitting was an issue. In our 2012 study, we showed that the best solution to build potentials with no interval issue is by using Dirichlet Process Mixture Models (DPMs) [24]. We also benefited of the group experience in modeling the dynamics of RNA and normal-mode experiments to obtain two good decoy sets which complemented the well-known Farna study. We also showed that in many case our high-resolution predictions were better than the Farna/Rosetta standard.

### *5.1.4. RNA 3D structure prediction*

In collaboration with PRISM at Versailles and Westhof's group at Strasbourg, we addressed the problem of ab initio prediction of RNA three-dimensional structure. We developed an algorithm for automatically predicting the topological family of any RNA three-way junction, an thus its coarse-grained local geometry, given only the information from the secondary structure: the sequence and the Watson–Crick pairings. Additionally, we showed that the results are noticeably improved if homology information is used [14]. The resulting software, `Cartaj`, is available online and downloadable at http://cartaj.lri.fr. Then we investigated a new approach for the global prediction of the coarse-grain 3D structure of RNA molecules. We model a molecule as being made of helices and junctions. Using our results above, we are able to classify junctions into topological families that determine their preferred 3D shapes. All the parts of the molecule are then allowed to establish long-distance contacts that induce a threedimensional folding of the molecule. An algorithm relying on game-theory was proposed to discover such long-distance contacts that allow the molecule to reach a Nash equilibrium. As reported by our experiments, this approach allows one to predict the global shape of large molecules of several hundreds of nucleotides that are out of reach of the state-of-the-art methods [15].

A graph-theoretic approach has been successfully used for classification and structure prediction of transmembrane beta-barrel proteins[23], [25].

## 5.2. Proteins structures and interactions

### *5.2.1. Protein-protein interaction*

Adrien Guilhot, PhD candidate in our project worked on a modified scoring function for the Rosetta software suite. After an extensive conformation generation for the two recently published benchmarks, we now have a model for protein-RNA semi-flexible docking which is currently being tested.

The prediction of the network of protein-protein interactions (PPI) of an organism is crucial for the understanding of biological processes and for the development of new drugs. Machine learning methods have been successfully applied to the prediction of PPI in yeast by the integration of multiple direct and indirect biological data sources. However, experimental data are not available for most organisms. We propose in [9] an ensemble machine learning approach for the prediction of PPI that depends solely on features independent from experimental data. New estimators of the coevolution between proteins have been developed and combined them in an ensemble learning procedure.

This method has been applied to a dataset of known co-complexed proteins in *Escherichia coli* and compared it to previously published methods. Our method allows prediction of PPI with an unprecedented precision of $95.5\%$ for the first 200 sorted pairs of proteins compared to $28.5\%$ on the same dataset with the previous best method.

A close inspection of the best predicted pairs allowed us to detect new or recently discovered interactions between chemotactic components, the flagellar apparatus and RNA polymerase complexes in *E. coli*.

## 5.3. Combinatorics and Annotation

### *5.3.1. Word counting and random generation*

A long-term research on word enumeration has been realized by the team, in order to calculate a statistical significance for a pattern occurrence according to a given background model. As a part of E. Furletova's thesis, defended in February 2012, co-advised by M. Roytberg (IMPB, Puschino, Russia) and M. Régnier, an extension to Hidden Markov Models, `SufPref`, has been proposed. It relies on a new concept of overlap graphs that efficiently overcomes the main difficulty - overlapping occurrences - in probabilities computation. An implementation is available at http://server2.lpm.org.ru/bio/online/sf/. This algorithm provides a significant space improvement over a previous algorithm, `AhoPro` developed with our former associate team MIGEC. Word statistics were used to identify mRNA targets for miRNAs involved in carcinogenesis [13].

Large deviation results have been derived in [41] that take advantage of general combinatorial properties of words. First, an approximation is derived for the double strands counting problem that refers to a counting of a given pattern in a set of sequences that arise from both strands of the genome. Here dependencies between a sequence and its complement plays a fundamental role. Second, sets of small sequences, with non-identical distributions, are addressed. Possible applications are the search of cis-acting elements in regulatory sequences that may be known, for example from ChIP-chip or ChipSeq experiments, as being under a similar regulatory control.

In [21], we developed a new algorithm for generating uniformly at random words of any regular language $L$. When using floating point arithmetics, its bit-complexity is $O(q \log n)$ in space and $O(qn \log n)$ in time, where $n$ stands for the length of the word, and $q$ stands for the number of states of a finite deterministic automaton of $L$. We implemented the algorithm and compared its behavior to the state-of-the-art algorithms, on a set of large automata from the VLTS benchmark suite. Both theoretical and experimental results show that our algorithm offers an excellent compromise in terms of space and time requirements, compared to the known best alternatives. In particular, it is the only method that can generate long paths in large automata. Moreover, in [10], in collaboration with the Fortesse group at LRI, we presented several randomised algorithms for generating paths in large models according to a given coverage criterion. This work opens new perspectives for future studies of statistical testing and model checking, mainly to fight the combinatorial explosion problem.

### *5.3.2. Analysis and design of weighted combinatorial models*

Weighted context-free grammars are natural – yet powerful – random models for biological sequence and structures. We furthered our developments on these objects, and applied them to the study of the Boltzmann ensemble of low-energy in RNA.

In collaboration with P. Clote (Boston College), we used such analytic combinatorics to establish that the average geometric distance between the terminal ends of an RNA sequence, once folded, is asymptotically constant [8].

Furthermore, in collaboration with C. Banderier, O. Bodini and H. Tafat (LIPN), we constructively showed that any predefined distribution of pattern could be attained by a (possibly ambiguous) regular expressions. We also designed a dynamic-programming algorithm to automatically build such models, adopting a segmentation approach based on a parsimony principle. This work was presented at the ANALCO'12 conference [30].

Finally, we continued with D. Gardy and J. Du Boisberranger (PRISM, Université de Versailles-St Quentin) a joint study of collisions in weighted random generation. Indeed, while performing a random generation within large collections of weighted objects, the probability of any sample can be exactly and efficiently computed. Therefore, any redundancy in the sampled set is uninformative (contrasting with situations where the probability is also estimated by the sampling procedure). Following previous results presented at GASCOM'10 (Montreal), we presented at the AOFA'12 (Montreal, Canada) conference [33], a new close formula for the waiting-time of the coupon collector problem, i.e. the average number of words that one must draw to obtain the full collection. The framework defined here has direct applications in the context of RNA : approaches based on sampling are preferred to deterministic optimizations, and algorithmic efficiency of the methods can be critically affected by the redundancy of sampled sets. .

### *5.3.3. Scientific Workflows*

Several Scientific workflow systems have been designed to support users in the tasks of designing, managing, monitoring, and executing in-silico experiments. Such systems are now equipped of provenance modules able to collect data produced and consumed during workflow runs to enhance reproducibility. In this context, we have worked in two directions. First, we have worked on the problem of reuse between scientific workflows. In particular, we have identified the presence of common or similar (sub-)workflows and workflow elements, and have deeply studied, for the first time in the literature, the problem of cross-author reuse [38].

Second, we have worked on studying the structure of scientific workflows. More precisely, we have focused on the series-parallel graph structures. Designing sub-workflows, querying or monitoring workflows leads to perform graph sub-isomorphism. This problem is NP-complete when general DAGs are considered but can be solved in polynomial time when graphs restricted to SP graphs are considered. We have designed and implemented the SPFlow algorithm that rewrites any workflow into an SP workflow while ensuring that the provenance of the rewritten workflow is the same as the original [32], [39].

We are currently working on identifying the reasons why some scientific workflows have a non SP structure. Our long-term goal is to design a *distilling procedure* for scientific workflows offering users the ability of naturally designing workflows having a structure close to SP structures. This work is done in close collaboration with the University of Manchester [31].

## 5.4. Systems Biology

### *5.4.1. Reasoning on knowledge to build signaling networks:*

We have introduced a logic-based method to infer molecular networks and show how it allows inferring signalling networks from the design of a knowledge base. Provenance of inferred data has been carefully collected, allowing quality evaluation. Our method (i) takes into account various kinds of biological experiments and their origin; (ii) mimics the scientist's reasoning within a first-order logic setting; (iii) specifies precisely the kind of interaction between the molecules; (iv) provides the user with the provenance of each interaction; (v) automatically builds and draws the inferred network [29].

### *5.4.2. Metabolic pathways*

The topological analyse of metabolic networks is a first step to understand their behaviours and is described in term of fluxes analyses. We work on the elaboration of a stoichiometric model of *Bacillus subtilis* where its fluxes analyse predicted transcriptional regulation to be more important for the dynamics induced by glucose than by malate [7].

In metabolic pathway analyses, the metabolic networks are described in term of biochemical reactions and metabolites. The integration of structural data is required for a comprehensive understanding of the metabolic networks. We represent the metabolic networks with the functional connectivity between the protein functional domains to make more relevant analyses. We used Bio$\Psi$, a formal multi-level description based on elementary actions, to assign functions on structural domains and the elementary flux modes theory to check if the already known pathways remain presents and to identify new ones.

A new version of the software has `Mpas` (Metabolic Pathway Analyser Software) been developed during a Master2'internship by Gh. Fievet. Meanwhile we have also introduced in the landscape of the cell its membranes and the numerous pumps that facilitate ions transfers, hence taking into account the pH of the cytoplasm, a parameter that fits the cell mytosis cycle and which proves to separate the cancerous/normal status of cells [22]. We now aim at study larger and more elaborate metabolic systems, including the Krebs cycle and the mitochondria influence, thus enhancing the scalability of our method [17].

### *5.4.3. Bacterial phenotypic adaptation*

We attempt to re-interpret a major event, the initiation of chromosome replication in *Escherichia coli*, in the light of scales of equilibria. This entails thinking in terms of hyperstructures as responsible for intensity sensing and quantity sensing and how this sensing might help explain the role of the DnaA protein in initiation of replication. We outline experiments and an automaton approach to the cell cycle that should test and refine the scales concept [19].

Another possible direction to study the mechanisms used by cells to integrate and respond to their environment is to search for a link between two large hyperstructures: the cytoskeleton and the general metabolic activity of the cell. There is extensive evidence for the interaction of metabolic enzymes with the eukaryotic cytoskeleton. We state the hypothesis that the cytoskeleton senses and integrates the general metabolic activity of the cell. The physical and chemical effects arising from metabolic sensing by the cytoskeleton would have major consequences on cell shape, dynamics and cell cycle progression. The hypothesis provides a framework that helps the significance of the enzyme-decorated cytoskeleton be determined [18].

In order to test these hypotheses, we have added many features to the HSIM simulation software. The main addition being a way to get both the power of expression of the "entity-centred" paradigm and the computational efficiency of global methods, such that Gillespie-like stochastic simulation algorithm (SSA). To achieve this, we have implemented two new algorithms. The first one concerns the possibility to take into account the interactions between two classes of molecules: the one we want to follow the spatial location over time (entities) and the one for which only the evolution of the number of copies over time is relevant.

The second algorithm is an enhancement of the tau-leap variant of the exact Gillespie SSA; This allows to take into account the interactions between globally treated molecules. The HSIM-SSA algorithm performs an adaptive processing of the number of reactions which may have been triggered during the time step. At each time step, the fast reactions are averaged while the slow reactions are fully stochastically treated. This allows HSIM-SSA to be more than 10 times faster than the other tau-leap SSA implementations [28].

### 5.4.4. Use of bacteria for biotechnology

Another center of interest has been to find a way to use bacteria as a mean to help us to engineer new biomolecules with specific characteristics. It is sometimes speculated that the equivalent of the polymerase chain reaction might be developed for identification of peptides, proteins or other molecules. Natural amplification systems do exist as in the case of certain autoinducer systems in bacteria. We have been outlined a possible, generic method, *the mimic chain reaction*, for obtaining peptides with 3-D structures that mimic the 3-D structure of their targets. These targets would include a variety of molecules, including proteins. There are therefore two categories of applications: the ability via amplification firstly to detect a known protein or other target at an extremely low concentration and secondly to obtain a set of peptides that mimic the structure of an unknown target and that can be used to obtain a *photofit* [20].

# BAMBOO Project-Team

# 6. New Results

## 6.1. Partial Enumeration of Traces

Traditional algorithms to solve the problem of sorting by signed reversals output just one optimal solution while the space of all optimal solutions can be huge. A so-called trace represents a group of solutions which share the same set of reversals that must be applied to sort the original permutation following a partial ordering. By using traces, we therefore can represent the set of optimal solutions in a more compact way. Algorithms for enumerating the complete set of traces of solutions were developed. However, due to their exponential complexity, their practical use is limited to small permutations. A partial enumeration of traces is a sampling of the complete set of traces and can be an alternative for the study of distinct evolutionary scenarios of big permutations. Ideally, the sampling should be done uniformly from the space of all optimal solutions. This is however conjectured to be #P-complete.

We proposed and evaluated three algorithms for producing a sampling of the complete set of traces that instead can be shown in practice to preserve some of the characteristics of the space of all solutions [7]. We analysed the distribution of the enumerated traces with respect to their height and average reversal length.

## 6.2. De-novo calling alternative splicing events from RNA-seq data

We addressed the problem of identifying and quantifying polymorphisms in RNA-seq data when no reference genome is available, without assembling the full transcripts. Based on the fundamental idea that each polymorphism corresponds to a recognisable pattern in a De Bruijn graph constructed from the RNA-seq reads, we proposed a general model for all polymorphisms in such graphs. We then introduced an exact algorithm, called KISSPLICE, to extract alternative splicing events. The first version of KISSPLICE appeared in 2011, but several important improvements were implemented in 2012 [24]. The first improvement was the memory consumption, the new version is much more memory efficient and can handle datasets of approximately $10^8$ reads. The second was in the running time, the enumeration step can now be done in parallel, which results in a significant speedup in the overall running time. Finally, an improved event quantification step was added to the method.

Application-wise, we showed that KISSPLICE enables to identify more correct events than general purpose transcriptome assemblers. Additionally, on a 71 M reads dataset from human brain and liver tissues, KISSPLICE identified 3497 alternative splicing events, out of which 56% are not present in the annotations, which confirms recent estimates showing that the complexity of alternative splicing has been largely underestimated so far.

## 6.3. Efficient bubble and/or cycle enumeration in directed/undirected graphs

Polymorphisms in DNA- or RNA-seq data lead to recognisable patterns in a de Bruijn graph representation of the reads obtained by sequencing. Such patterns have been called mouths, or bubbles in the literature. They correspond to two vertex-disjoint directed paths between a source $s$ and a target $t$. Due to the high number of such bubbles that may be present in real data, their enumeration is a major issue concerning the efficiency of dedicated algorithms. We developed the first linear delay algorithm to enumerate all bubbles with a given source [31].

By combining the insights from the most efficient but not optimal solution presented by Johnson [SIAM J. Computing, 1975] for simple cycle enumeration in undirected graphs and an amortisation technique previously established by our collaborators Roberto Grossi and Rui Ferreira [ESA, 2011] from the University of Pisa, Italy, we obtained the first optimal solution to list all the simple cycles in an undirected graph $G$ (paper accepted at SODA 2013, to appear). Moreover, we also obtained the first optimal solution to list all the simple paths from $s$ to $t$ in an undirected graph $G$. This work benefited also from discussions and work from Pierluigi Crescenzi and Marie-France Sagot. The method is not naturally extendable to directed graphs, and the challenge is now to obtain optimal solutions in this case also.

## 6.4. Simulating RNA-seq experiments

RNAseq experiments now enable to characterise the RNA complement of a cell. However, the series of steps (fragmentation, reverse transcription, sequencing) that separate the initial RNA molecules from the short DNA reads obtained in fine are not well understood although it is widely accepted that they contribute to generating noise in the signal. We introduced the FLUXSIMULATOR [14], a computer program able to reproduce the biases seen in RNAseq data. This pipeline should prove useful both to produce realistic data on which to test programs which aim at reconstructing RNA from short reads, and suggest ways of improving the experimental steps so that they produce less noise.

## 6.5. Chimeric Transcripts may be Translated

There is now increasing evidence for the existence of so-called Chimeric Transcripts. In contrast to regular transcripts, which are composed of exons located close to each other on the genome, these chimeric transcripts can be composed of exons which are located megabases away, or even on different chromosomes. We showed that these chimeras are lowly expressed, are tissue specific, and that some of them may be translated, yielding proteins with altered function or localisation [13].

## 6.6. Transcriptomics of symbiosis in the Asobara tabida-Wolbachia association

*Wolbachia* has evolved a very peculiar phenotype in the host *Asobara tabida* where it is obligatory for oogenesis. Transcriptomics approaches were developed first using Sanger sequencing of mRNA [19]. It has now been complemented by RNAseq analyses on two lines, which exhibit different ovarian phenotypes in absence of *Wolbachia*. We have currently analysed these data both to isolate genes that are differentially expressed, but also that exhibit polymorphism between the lines. Interesting candidates were detected that are under further investigation and that are involved in the regulation of early oogenesis, apoptosis, autophagy and oxidative stress. This part is in direct connection with the algorithms developed by BAMBOO for the analysis of NGS data without a reference genome (KISSPLICE).

## 6.7. MicroRNA predictor

We developed a microRNA predictor using structural and target information. The method shows 97% sensitivity and 90% specificity for the *Acyrthosiphon pisum* genome. Comparing to the results of the previous method we developed in 2010 (available in the software CRAVELA) we obtained a better performance (sensitivity 90% and specificity 88%). However, as we are working on a genome wide scale, it is important to obtain even better specificity (obviously, maintaining a reasonable sensitivity). This work is currently in development.

On the other hand, the computational search for novel miRNA precursors often involves also some sort of structural analysis with the aim of identifying which type of structures are recognised and processed by the cellular miRNA-maturation machinery. A natural way to tackle this problem is to perform clustering over the candidate structures along with known miRNA precursor structures. Mixed clusters allows then the identification of candidates that are similar to known precursors. Given the large number of pre-miRNA candidates that can be identified in single-genome approaches, even after applying several filters for precursor robustness and stability, a conventional structural clustering approach is unfeasible. We proposed a method to represent candidate structures in a feature space which summarises key sequence/structure characteristics of each candidate [21]. We showed that proximity in this feature space is related to sequence/structure similarity, and we selected candidates which have a high similarity to known precursors. Additional filtering steps were then applied to further reduce the number of candidates to those with greater transcriptional potential.

## 6.8. Genomics of symbiosis

Insect symbioses are model systems for studying the effect of symbionts and the evolution of bacterial genomes. Members from the LBBE described the symbiotic complement of different biotypes of the insect *Bemisia tabaci* in Western Africa. We further obtained the complete genome of different symbionts that co-exist in *Bemisia tabaci*, among which the the primary symbiont *Portiera* [25], *Hamiltonella*, *Rickettsia* and *Wolbachia*. Analyses are underway, that concern the possible complementation between *Hamiltonella* and *Portiera* and the comparative analyses of different *Hamiltonella* genomes.

## 6.9. Representation and curation of metabolic data: UniPathway, Rhea and MNX

These activities are carried out in collaboration with the SwissProt group at the Swiss Institute for BioInformatics (SIB). UNIPATHWAY (http://www.unipathway.org) is a manually curated database of metabolic pathways. It provides the official controlled vocabulary for pathway annotation within UNIPROTKB records since 2009. A complete description of the UNIPATHWAY database and of its relationship with UNIPROTKB has been published in *Nucleic Acids Research* (Jan. 2012 Database Issue) [22]. RHEA (http://www.ebi.ac.uk/rhea) is developed jointly with the European Institute for Bioinformatics (EBI) and the SIB. It provides a comprehensive resource of expert-curated biochemical reactions, for use in a large spectrum of applications, including metabolic network reconstruction and pathway inference. The complete description of the RHEA database appeared in the Jan. 2012 *NAR* Database issue [5]. The MNX project is developed in the context of the METANETX project (http://www.metanetx.org). It attempts to automate the reconciliation of discrepancies between metabolite or reaction information from distinct resources (BIGG, BRENDA, CHEBI/RHEA, KEGG, METACYC, UNIPATHWAY, THE SEED, REACTOME), thereby alleviating a major bottleneck in the construction of genome-scale metabolic network models. The MNXREF namespace is available at http://www.metanetx.org/mnxdoc/mnxref.html and the method to compute the MNXREF namespace is described in [8].

## 6.10. Annotation of the proteins of Angomonas deanei and Strigomonas culicis

*Angomonas deanei* and *Strigomonas culicis* are trypanosomatids that harbour only one beta-proteobacterial endosymbiont and this mutualistic association is an interesting model to study eukaryotic cell evolution. The genomes of these organisms were sequenced by our collaborators at LNCC / MCT (Brazil) and we participated in the functional annotation of these genomes as concerns their metabolism which enabled to reveal new aspects of the *Trypanosomatidae* family. This work has been submitted for publication. It was done with Ana Tereza Vasconcelos in a collaboration with Maria Cristina Machado Motta (UFRJ - Brazil).

## 6.11. Finding candidate genes for orphan enzymes

Of all biochemically characterized metabolic reactions formalized by the IUBMB, over one out of four have yet to be associated with a nucleic or protein sequence, *i.e.* are sequence-orphan enzymatic activities. Few bioinformatics annotation tools are able to propose candidate genes for such activities by exploiting context-dependent rather than sequence-dependent data, and none are readily accessible and propose result integration across multiple genomes. We introduced CANOE (Candidate genes for Orphan Enzymes), a four-step bioinformatics strategy that proposes ranked candidate genes for sequence-orphan enzymatic activities (or orphan enzymes for short) [26]. Our strategy found over 60,000 genomic metabolons in more than 1,000 prokaryote organisms from the MICROSCOPE platform developed by the group of Claudine Médigue from the Génoscope with whom this work was done, generating candidate genes for many metabolic reactions, of which more than 70 distinct orphan reactions. A computational validation of the approach was discussed and we presented a case study on the anaerobic allantoin degradation pathway in *Escherichia coli* K-12.

## 6.12. Metabolic cooperation of symbionts and their host trypanosomatids

Trypanosomatids that harbour a symbiotic bacterium (SHTs) are known to have less nutritional requirements when compared to their counterparts without symbionts (RTs). Nutritional and biochemical data indicated that the symbionts largely contributed to the routes for amino acid and vitamin biosynthesis. We analysed the genomic data of 5 SHTs and their respective symbionts and 2 RTs and we found most of the genes related to those pathways in the symbionts. This work will soon be submitted for publication. It is being done with Ana Tereza Vasconcelos in a collaboration with Erney P. Camargo, Marta M.G. Teixeira (USP - Brazil), João M.P. Alves, Gregory A. Buck (VCU - USA), and Maria Cristina Machado Motta (UFRJ - Brazil).

## 6.13. Structural and dynamical analysis of biological networks

We published a review on the structural and dynamical analysis of biological networks with as main focus explaining the cares that should be taken when this kind of analysis is performed [18]. Correctly distinguishing between potential metabolic networks and their realisations is necessary in choosing the right methods to be used and in the interpretation of their outcomes. In our review, we covered several different techniques, both static and dynamic, for the analysis of metabolic networks such as centrality techniques, flux-balance analysis and kinetic modelling of full-scale networks.

## 6.14. Network distance analysis

We addressed the diameter computation problem in the case of undirected unweighted graphs, where the diameter $D$ is defined as the maximum distance among all the pairs of nodes and the distance $d(u, v)$ between two nodes $u$ and $v$ is defined as the number of edges contained in the shortest path from $u$ to $v$. In the context of real-world networks, the textbook method based on performing a breadth-first search (in short, BFS) from *every* node of the graph, requires a prohibitive cost of $O(nm)$ time, where $n$ is the number of nodes and $m$ is the number of edges of the graph. Our main contribution consists of showing that BFS can indeed be an extremely *powerful* tool to compute the *exact* value of the diameter, whenever it is used in a more clever way. In particular, we have developed the *iterative* Fringe Upper Bound (in short, *i*FUB) algorithm to calculate the exact value of the diameter. This work has been accepted for publication in *Theoretical Computer Science* (to appear).

We then successively generalised the idea of the *i*FUB algorithm, by presenting the directed *i*FUB (in short, D*i*FUB) algorithm, in order to calculate the diameter of the strongly connected components of directed graphs [33]. As far as we know, D*i*FUB is the first algorithm which is able to compute exactly the diameter of the strongly connected components of huge real-world directed graphs. The D*i*FUB algorithm can also return a pair of nodes whose distance is exactly equal to the diameter, and a natural adaptation of it works also for weighted graphs.

## 6.15. Information spreading in dynamic graphs

We showed how a technique used to analyse the flooding completion time in the case of a special class of random evolving graph model, that is, the *edge-Markovian model*, can be used in order to prove that the flooding completion time of a random evolving graph $(G_t)_{t \geq 0}$ is bounded by $kD + 2C$, where intuitively (1) $k$ is the smallest time necessary for the rising of a giant component, (2) $D$ is the diameter of the giant component, and (3) $C$ is the time required for the nodes outside the giant component to eventually get an edge connecting them to the giant component [30]. Then, based on this result, we developed a general methodology for analysing flooding in sequences of random graphs and we applied this general methodology to the case of power-law evolving graphs (that is, sequences of mutually independent random graphs such that the number $y$ of nodes of degree $x$ distributes like $1/x^{\beta}$ for some $\beta > 0$), and to the case of an arbitrary given degree distribution.

## 6.16. Metabolic network comparison

Previous works on minimal gene sets, when analysing host-dependent bacteria, found small common sets of metabolic genes. When such analyses are restricted to bacteria with similar lifestyles, larger portions of metabolism are expected to be shared and their composition is worth investigating. Comparing the small molecule metabolism of 58 bacteria carefully selected and representing a range of lifestyles, we found not a single enzymatic reaction common to all of them. While obligate intracellular symbionts have no core of reactions within their group, extracellular and cell-associated symbionts do have a small core enriched in biosynthetic processes composed of disconnected fragments. As more genomes are added, we expect, based on our simulations, that the core of cell-associated and extracellular bacteria continues to diminish, converging to approximately 60 reactions. These results were in preparation in 2011 and are now published [17]. The work was done with Ana Tereza Vasconcelos and in a collaboration with Ludovic Cottret (INSA Toulouse).

## 6.17. Core and periphery of metabolic networks

The core metabolism can be defined as the reactions present in every organism, however it is not robust considering that adding or removing one organism in the study will modify the resulting set. An alternative way is to include in the core the reaction that is present in a large enough proportion of species. For that, we proposed a method where the threshold to decide what is large enough is not set by the user (thus relying on a subjective choice), but rather automatically selected by the method, relying on the information contained in the data. Two approaches are being proposed, one is EM (Expectation Maximization) which relies only on the information of presence / absence of a reaction in a species while the second (NEM - Neighboring Expectation Maximization) relies on a neighbouring relation between reactions. The latter tends to classify in a same group (core or periphery) a reaction for which a majority of neighbours belong to a same group. The work is being done with Ana Tereza Vasconcelos in a collaboration with Catherine Matias, Christophe Ambroise, Yolande Diaz (Genopole, CNRS).

## 6.18. Metabolic stories

Enumerating stories, *i.e.*, enumerating maximal directed acyclic graphs with sets of sources and targets contained in a given subset of the nodes, is an algorithmic approach we proposed for interpreting metabolomics experiments. The modelling, algorithms and complexity results were recently accepted for publication [2]. The complexity of the enumeration problem remains unknown. There are also further modelling issues that could be dealt with in a near future. Both considerations were also detailed in a talk given in August at St. Petersburg, in the First RECOMB Satellite Conference on Open Problems in Algorithmic Biology.

We then applied our enumerating method on real data. We analysed data on the detoxification process of yeast cells exposed to cadmium. Our method allowed to recover known pathways involved in the process but also to propose alternative scenarios. The method was also investigated in order to automatically propose metabolic pathways through an experiment in which we try to recover known metabolic pathways using only minimal information (*e.g.*, their entries and endpoints). A paper is in prepartion and should soon be submitted for publication. This work is being done in collaboration with Fabien Jourdan and Ludovic Cottret from the INRA at Toulouse, and with Christophe Junot from the CEA in Paris.

## 6.19. Minimal precursor sets

We proposed two new, more efficient algorithms for the enumeration of minimal precursor sets: PITUFINA and PAPA PITUFO [3]. The model of minimal precursor sets we had previously published was the first to formally take into account cycles, which are a common event in metabolic networks. The new methods avoid the memory issues of our previous approach by traversing directly the metabolic network structure instead of building a secondary tree representation. PAPA PITUFO additionally saves pre-computed solutions by a local modification of the network.

## 6.20. Minimum ratio cover of matrix columns by extreme rays of its induced cone

Given a matrix $S$ and a subset of columns $R$, we studied the problem of finding a cover of $R$ with extreme rays of the cone $\mathcal{F} = \{v \in \mathbb{R}^n \mid Sv = \mathbf{0}, v \geq \mathbf{0}\}$, where an extreme ray $v$ covers a column $k$ if $v_k > 0$ [34]. In order to measure how proportional a cover is, we introduced two different minimisation problems, namely the MINIMUM GLOBAL RATIO COVER (MGRC) and the MINIMUM LOCAL RATIO COVER (MLRC) problems. In both cases, we applied the notion of the *ratio* of a vector $v$, which is given by $\frac{\max_i v_i}{\min_{j \mid v_j > 0} v_j}$. These problems were originally motivated by a biological question on metabolic networks. This notion of ratio is also of interest in the field of *exact linear programming*, where current algorithms for scaling a matrix have a complexity that depends on the ratio of its elements. We showed that these two problems are NP-hard, even in the case in which $|R| = 1$. We introduced a mixed integer programming formulation for the MGRC problem, which is solvable in polynomial time if all columns should be covered, and introduce a branch-and-cut algorithm for the MLRC problem. Finally, we presented computational experiments on data obtained from real metabolic networks.

## 6.21. Optimal flux spaces of genome-scale stoichiometric models

The metabolism of organisms can be studied with comprehensive stoichiometric models of their metabolic networks. Flux balance analysis (FBA) calculates optimal metabolic performance of stoichiometric models. However, detailed biological interpretation of FBA is limited because, in general, a huge number of flux patterns give rise to the same optimal performance. The complete description of the resulting optimal solution spaces was thus far a computationally intractable problem. We introduced COPE-FBA: Comprehensive Polyhedra Enumeration Flux Balance Analysis, a computational method that solves this problem [15]. COPE-FBA indicates that the thousands to millions of optimal flux patterns result from a combinatorial explosion of flux patterns in just a few metabolic sub-networks. The entire optimal solution space can now be compactly described in terms of the topology of these sub-networks. COPE-FBA simplifies the biological interpretation of stoichiometric models of metabolism, and provides a profound understanding of metabolic flexibility in optimal states.

## 6.22. Lateral gene transfer as a support for the tree of life

We published with Sophie Abby the last results of her PhD work that apply an explicit phylogenetic model of horizontal gene transfer to bacterial and archaeal phyla [1]. We showed that lateral gene transfer allows to discriminate between phylogenetic hypotheses, and that in a typical bacterial gene family, 96-98% of tree branches result from vertical descent and 2-4% from lateral gene transfer.

## 6.23. Comparative approximability of hybridization number and directed feedback vertex set

We showed that the problem of computing the hybridization number of two rooted binary phylogenetic trees on the same set of taxa $X$ has a constant factor polynomial-time approximation if and only if the problem of computing a minimum-size feedback vertex set in a directed graph (DFVS) has a constant factor polynomial-time approximation. The latter problem, which asks for a minimum number of vertices to be removed from a directed graph to transform it into a directed acyclic graph, is one of the problems in Karp's seminal 1972 list of 21 NP-complete problems. However, despite considerable attention from the combinatorial optimisation community, it remains to this day unknown whether a constant factor polynomial-time approximation exists for DFVS. Our result thus placed the (in)approximability of hybridization number in a much broader complexity context, and as a consequence we obtained that hybridization number inherits inapproximability results from the problem Vertex Cover [16]. On the positive side, we used results from the DFVS literature to give an $O(logrloglogr)$ approximation for the hybridization number, where $r$ is the value of an optimal solution to the hybridization number problem. This work is submitted for publication.

## 6.24. Influence of symbionts on antagonistic interactions

Symbionts are often key players in antagonistic interactions between their hosts and other organisms. In host-parasitoid interactions, both players can be infected by different symbionts. We investigated how a virus and *Wolbachia*, respectively infecting a parasitoid and a drosophila, can shape the host-parasitoid interaction. While only a limited effect *Wolbachia* has been detected, the virus protects the parasitoid from the immune response of *Drosophila* [20]. Protection conferred by symbionts to their insect hosts is a promising avenue for antivectorial programs, but requires a thorough analysis of the evolutionary consequences of protection. We reviewed the literature on this topic [28].

## 6.25. Mod/Resc Parsimony Inference

We addressed a computational biology problem that aims at understanding a mechanism that could potentially be used to genetically manipulate natural insect populations infected by inherited, intra-cellular parasitic bacteria. In this problem, that we denoted by Mod/Resc Parsimony Inference, we are given a boolean matrix and the goal is to find two other boolean matrices with a minimum number of columns such that an appropriately defined operation on these matrices gives back the input. We showed that this is formally equivalent to the Biclique Edge Cover for Bipartite Graphs problem and derive some complexity results for our problem using this equivalence. We provided a new, fixed parameter tractability approach for solving both problems that slightly improves upon a previously published algorithm for the Biclique Edge Cover for Bipartite Graphs. Finally, we presented experimental results applying some of our techniques to a real-life dataset. This is the augmented journal version [23] of the conference paper that appeared in 2011.

## 6.26. On the genetic architecture of cytoplasmic incompatibility

Numerous insects carry intracellular bacteria manipulating their reproduction and thus facilitating their own spread. Cytoplasmic incompatibility (CI) is a common form of such manipulation, where a (currently uncharacterized) bacterial modification of male sperm 35 induces the early death of embryos unless the fertilized eggs carry the same bacteria, inherited from the mother. The death of uninfected embryos provides an indirect selective advantage to infected ones, thus enabling the spread of the bacteria. We used and expanded recently developed algorithms (the first being the one described in the previous item) to infer the genetic architecture underlying the complex incompatibility data from the mosquito *Culex pipiens*. We showed that CI requires more genetic determinants 40 than previously believed, and that quantitative variation in gene products potentially contributes to the observed CI patterns. In line with population genetic theory of CI, our analysis suggests that toxin factors (those inducing embryo death) are present in fewer copies in the bacterial genomes than antitoxin factors (those ensuring that infected embryos survive). In combination with comparative genomics, our approach will provide helpful guidance to 45 identify the genetic basis of CI, and more generally of other toxin / anti-toxin systems that can be conceptualised under the same framework. This work is currently submitted for publication. It was done in collaboration with Sylvain Charlat from the LBBE.

## 6.27. Viral population structure and dynamics

The work which started a few years ago with the Pasteur Institute in Cambodia (Dr. P. Buchy) and the CIRAD at Montpellier (Dr. R. Frutos) on viral population structure and dynamics has been continued in 2012, focusing on the H5N1 and Dengue viruses. The exploratory statistical approach based on MCOA (see the Bamboo annual report for 2011) was used to identify a novel H5N1 endemic sub-clade specific to Cambodia [27] and the work performed last year on Dengue serotype 1 has been extended in 2012 to serotypes 2 and 3 [11] thus providing a more precise view of the virus population dynamics over the last 12 years and demonstrating "synchronized" replacements most probably linked to climatic disasters like flood or drought.

## 6.28. Charge group partitioning in biomolecular simulation

Molecular simulation techniques are increasingly being used to study biomolecular systems at an atomic level. Such simulations rely on empirical force fields to represent the intermolecular interactions. There are many different force fields availableleach based on a different set of assumptions and thus requiring different parametrization procedures. Recently, efforts have been made to fully automate the assignment of force-field parameters, including atomic partial charges, for novel molecules. In this work, we focused on a problem arising in the automated parameterisation of molecules for use in combination with the gromos family of force fields: namely, the assignment of atoms to charge groups such that for every charge group the sum of the partial charges is ideally equal to its formal charge. In addition, charge groups are required to have size at most $k$. We showed NP-hardness and gave an exact algorithm capable of solving practical problem instances to provable optimality in a fraction of a second [32].

# BEAGLE Team

# 6. New Results

## 6.1. Model of genome reduction

To test whether the effect of the rearrangement rate on genome size holds independently of the artificial chemistry of the aevol (individual-based) model, we have written a simpler, mathematical model of genome size evolution including both genes and intergenic sequences, evolving through small insertions and deletions, large deletions and duplications and through selection based on gene number. The approach was presented this summer as a poster at the SMBE conference (Society for Molecular Biology and Evolution). We have shown analytically that without selective pressure, genomes spontaneously shrink and that large genomes are particularly unstable. When selection is included that favors the highest gene number, simulations show that genome sizes do not grow indefinitely as large genomes cannot be sustained. There is a trade-off between fitness and structural stability. A manuscript is being written and will be submitted in January.

## 6.2. The Paradoxical Effects of Allelic Recombination on Fitness

D.P. Parsons, C. Knibbe, G. Beslon. [42]

We introduced in the aevol model the possibility of DNA exchange by allelic recombination, in order to study the influence of recombination on the evolution of both fitness and genomic architecture. Surprisingly, despite the theoretical benefits it could confer, there seems to be very little (if any) differences in the fitness of the evolved organisms between the different groups of simulations.

## 6.3. Genome histories reconstructions

E. Tannier (Beagle), with B. Boussau, G. Szollosi, V. Daubin, L. Duret, M. Gouy, S. Abby (LBBE, Lyon), N. Lartillot (Univ Montreal), C. Chauve (SFU Vancouver)

Lateral gene transfer has been discovered in the 1940's and since has been seen by phylogeneticists as a noise one had to remove before analyses in molecular evolution. Thi noise was recently considered so important that it would blur the historical signal and leave no hope for reconstructing a phylogeny. In a series of papers [16], [31], [32], we model the lateral gene transfer and prove that it can be used as a signal to

- reinforce the support for the phylogeny of vertical descent [16]
- order in time some bacterial diversification events, and thus provide a unique source for dating the history of life (more than 3/4 of it is prokaryotic and the fossil record is not abundant) [31]
- have a trace of extinct species which did not leave any descendants, if they gave some genes to more successful lineages, which opens the way to include them in molecular phylogenies [32]

We devised methods to trace whole genome evolution, with multi-scale mutations: from nucleotide substitutions to large-scale rearrangements. We provided a mammalian phylogeny accounting fot the evolution of several thousand genes [17], and a method to sample among evolutionary scenarios [27].

Eventually we built a model of evolution of relations between pairs of genes, enable us to reconstruct ancestral genome structure or ancestral systems of interactions [18]. In the case of genome structure we also published a method to linearize a set of ancestral relations [26].

## 6.4. A Theory of Rate Coding Control by Intrinsic Plasticity Effects

H. Berry (Beagle), J. Naudé and B. Delord (ISIR, CNRS UMR 7222, Univ P&M Curie, Paris) and J.T. Paz (Stanford Univ Medical Center, CA, USA).

Over the past decades, experimental and theoretical studies of the cellular basis of learning and memory have mainly focused on synaptic plasticity, the experience-dependent modification of synapses. However, behavioral learning has also been correlated with experience-dependent changes of non-synaptic voltage-dependent ion channels. This intrinsic plasticity changes the neuron's propensity to fire action potentials in response to synaptic inputs. Thus a fundamental problem is to relate changes of the neuron input-output function with voltage-gated conductance modifications. Using a sensitivity analysis in biophysically realistic models, we depicted a generic dichotomy between two classes of voltage-dependent ion channels [28]. These two classes modify the threshold and the slope of the neuron input-output relation, allowing neurons to regulate the range of inputs they respond to and the gain of that response, respectively. We further provide analytical descriptions that enlighten the dynamical mechanisms underlying these effects and propose a concise and realistic framework for assessing the computational impact of intrinsic plasticity in neuron network models. Our results account for a large repertoire of empirical observations and may enlighten functional changes that characterize development, aging and several neural diseases, which also involve changes in voltage-dependent ion channels.

## 6.5. The influence of topology on calcium wave propagation in 3D astrocyte networks

H. Berry, Jules Lallouette (Beagle)

Glial cells are non-neuronal cells that constitute the majority of cells in the human brain and significantly modulate information processing via permanent cross-talk with the neurons. Astrocytes are also themselves inter-connected as networks and communicate via chemical wave propagation. How astrocyte wave propagation depends on the local properties of the astrocyte networks is however unknown. We have investigated the influence of the characteristics of the network topology on wave propagation [38]. Using a model of realistic astrocyte networks ($>1000$ cells embedded in a 3D space), we show that the major classes of propagations reported experimentally can be emulated by a mere variation of the topology. Our study indicates that calcium wave propagation is favored when astrocyte connections are limited by the distance between the cells, which means that propagation is better when the mean-shortest path of the network is larger. This unusual property sheds new light on consistent reports that astrocytes in vivo tend to restrict their connections to their nearest neighbors.

## 6.6. Dynamics of protein aggregation in Escherichia coli

H. Berry, Anne-Sophie Coquel (Beagle) and Ariel Lindner (INSERM U1001, Cochin Medical School, Paris).

Protein aggregation plays a key role in cell decline and leads to several human disease linked to ageing like Alzheimer or Parkinson disease and prion disease. In Escherichia coli bacteria, accumulation of damaged proteins and their asymmetric segregation allowed to show ageing signs. This work [14] is focused on the in vivo spatial dynamics of protein aggregates in E. coli. Protein aggregates can be classified as inclusion bodies and they are amorphous or amyloid with a high order level due to $\beta$ sheets. Combining a double theoretical and experimental approach, based on modeling and time-lapse and microfluidic microscopy, we studied the mechanism governing the motion of protein aggregates and the long-term vertical transmission of prionoid aggregates for about 10 generations. Our results show clearly that Brownian diffusion governs the motion of protein aggregates and the diffusion coefficient depends on the molecule size. The amyloid proteinopathy study shows the existence of lineages propagating two kind of aggregates : globular or comet-like. Lineages maintaining globular aggregates present an increase of the aggregate size until inhibition of the growth rate while comet-like aggregates are mildly detrimental to growth. We observed also at low frequency in some lineages the presence of both aggregates and a switch between them. Globular foci give born to comet-like aggregates.

## 6.7. Model of membrane domains emergence

HA Soula, A Coulon, G Beslon (Beagle)

In the classical view, cell membrane proteins undergo isotropic random motion, that is a 2D Brownian diffusion that should result in an homogeneous distribution of concentration. It is, however, far from the reality: Membrane proteins can assemble into so-called microdomains (sometimes called lipid rafts) which also display a specific lipid composition. The amount of this so-called overconcentration at equilibrium is simply related to the ratio of diffusion coefficients between zones of high and low diffusion. Expanding the model to include particle interaction, we show that inhomogeneous diffusion can impact particles clusterization as well. The clusters of particles were more numerous and appear for a lower value of interaction strength in the zones of low diffusion compared to zones of high diffusion. Provided we assume stable viscosity heterogeneity in the membrane, our model proposes a simple mechanism to explain particle concentration heterogeneity and hence domains.

## 6.8. Deleterious effect of receptor clustering on canonical signaling pathways

BR Caré, HA Soula (Beagle)

Classical framework for analyzing system biology pathways assumed that the cells are a well mixed and stirred medium. This hypothesis can dramatically fail in the case of membrane based stage of signaling. Due to microdomains membrane receptors are colocalized. Using individual based-model we show that this clustering seriously impairs the overall ligands binding as well as several pathways downstream. We contend that this unexpected effect is a very simple tool available for a cell to adjust its response.

## 6.9. Novel mathematical model of Adipose tissue cells size distribution

HA Soula (Beagle) C. Soulage, A Géloën (CARMEN)

We present a novel model to explain bimodality of size distribution of adipocytes: adipose tissue cells. These cells are dedicated to storing energy excess in form of fat and therefore can experience wide variations of sizes. Ubiquitous to all the species, we tested so far the size distributions are bimodal with no characteristic size. Using data from experiments, we provide a simple surface based model of circulating fats that cells can exchange. We show that in the physiological range for the parameters of the model, we obtain bimodal distribution. We also provide prediction of the size evolution during severe caloric restriction that we were able to verify experimentally as well.

## BONSAI Project-Team

# 6. New Results

## 6.1. High-throughtput sequence processing

- Within the PhD of T. T. Tran, we proposed a new indexing structure adapted to GPUs. We studied an indexing scheme with perfect hashing functions, and developed a prototype written in openCL for a read mapper. This read mapper has a sensitivity comparable to state-of-the-art read mappers, and provides substantial time gains in some cases.
- Within our collaboration with the Lille hospital on the follow-up of leukemia residual disease, we proposed a new heuristic to study immunological VDJ recombinations and follow their evolution along the time. The method is under testing on several datasets obtained from the Ion Torrent sequencer at IRCL (Institut de Recherche sur le Cancer de Lille).
- Within the PhD of E. Kopylova, we designed an new algorithm to filter out ribosomal RNA sequences from RNA raw data produced in metatranscriptomic sequencing. The method combines text indexing techniques, with the Burst trie, and Universal Levenshtein automaton to allow for seraching with errors. An article has been published the journal *Bioinformatics* [4].

## 6.2. Noncoding RNAs

- We designed a new algorithm to produce all locally optimal secondary structures of an RNA sequence. Locally optimal secondary structures are thermodynamically stable RNA structures that are maximal for inclusion: they cannot be extended without producing a conflict between base pairs in the secondary structure, or increasing the free energy. This was published in *Journal for Computational Biology* [7].
- We took part to a collaborative work on benchmarking for RNA structure comparison. This work has been published in *Advances in bioinformatics* [2].

## 6.3. Genomic rearrangements

- Within the context of the PhD of A. Thomas:
  - We designed an algorithm for finding the minimal number of block interchanges required to transform a duplicated linear genome into a tandem duplicated linear genome. We provide a formula for the distance as well as a polynomial time algorithm for the sorting problem. This work was published in the conference *Bioinformatics* [13].
  - We explored a new problem concerning tandem halving, that is reconstructing a non-duplicated ancestor to a partially duplicated genome in a model where duplicated content is caused by several tandem duplications. We provide a distance in $O(n)$ time and a scenario in $O(n^2)$ time. We considered several problems related to multiple tandem reconstruction and proved that the simpliest of reconstructing 2 tandems is NP-hard. This work was published in the conference WABI 2012 [14].
- In the context of ancestral genome reconstruction, we designed an algorithm for the identification of Minimal Conflicting Sets (MCS) rows in a biological binary matrix. We provided a $O(n^2m^2 + nm^7)$ time algorithm, largely improving the up-to-date best algorithm in $O(m^6n^5(m+n)^2log(m+n))$ time. This work was published in the conference CPM 2012 [11].

- In the context of the comparison of sets of alternative gene transcripts, we designed a general framework to compare sets of transcripts that are transcribed from orthologous loci of several species. The model is based on the construction of a common reference sequence, and on annotations that allow the reconstruction of ancestral sequences, the identification of conserved events, and the inference of gains and losses of donor/acceptors sites, exons, introns and transcripts. This work was published in the conference ISBRA 2012 [12].

## 6.4. Nonribosomal peptides

- With the arrival of Ammar Hasan, a postdoc researcher, we started a new project on the prediction of nonribosomal peptides activity. We defined a novel peptide fingerprint based on monomer composition. This fingerprints is used for peptide similarity searching and for activity prediction. This work was published in *Journal of Computer-Aided Molecular Design* [1].
- We participated in the writing of a review dedicated to kurstakin, a nonribosomal lipopeptide synthetized by several Bacillus genus and published in *Applied microbiology and biotechnology* [3].
- The collaboration with members of EPI Orpailleur (CRI Nancy) succeeded in designing a protocol to discover new nonribosomal peptide synthetases in bacterial genomes and then annotate them in order to predict the peptide they produce. It was published in JOBIM 2012 [16].

# DYLISS Team

# 6. New Results

## 6.1. Data integration

**Participants:** Jacques Nicolas [**contact**], Geoffroy Andrieux, Andres Aravena, Pierre Blavy, Jérémie Bourdon, Guillaume Collet, Damien Eveillard, Michel Le Borgne, Sylvain Prigent, Anne Siegel, Sven Thiele, Valentin Wucher.

- **Identification of key regulators by the integration of flux and regulatory information** [*P. Blavy, A. Siegel*] We introduced a new method to combine reaction-based "flux" information (consumption and prediction of molecules) and regulatory "causal" information (effect of the variation of a molecule on the variation of another molecule) in order to find potential key regulators of a set of molecules. It has been validated by recovering among the causal graph derived from the Transpath database the main regulators of 190 groups of genes which are known to share a transcription factor according to the TRED database. [22][Online publication]
- **Reconstruction of transcriptional networks** [*A. Aravena, A. Siegel*] Transcriptional regulatory network models can be reconstructed ab initio from DNA sequence data by locating the binding sites, defined by position specific score matrices, and identifying transcription factors by homology with known ones in other organisms. In general the resulting network contains spurious elements. We use differential expression experimental data, in the form of Mutual Information, as ASP logical constraints to be satisfied by any valid regulatory network subgraph. These rules are used to determine the minimal sets of motif and transcription factors which constitute a genetic regulatory network compatible with experimental data [20][Online publication].
- **Studying diversity in marine environment** [*D. Eveillard*] We proposed a statistical-based data analysis of environmental microarrays. It shows that similar physical parameters drive bacterial and archae communities that share common ammonia oxidizing capacities [12][Online publication]
- **Brown algae metabolic network reconstruction** [*S. Prigent, S. Thiele, A. Siegel*] In order to better understand the functioning of cellular metabolism in the model brown alga *E. siliculosus*, metabolic networks are under construction based on genomic information. Two approaches are conducted in parallel to complete the network, a stochastic one that proceeds by sampling the solution space and a combinatorial one that tries to minimize the number of added reactions [23].

## 6.2. Asymptotic dynamics

**Participants:** Anne Siegel [**contact**], Oumarou Abdou-Arbi, Geoffroy Andrieux, Pierre Blavy, Jérémie Bourdon, Damien Eveillard, Michel Le Borgne, Vincent Picard, Sven Thiele, Santiago Videla.

- **Probabilistic sources for sequences and systems biology** [*J. Bourdon*] The habilitation thesis surveys how methods based on average-case analysis of algorithms can be used to model the quantitative response of a biological system from a biomolecular to a physiological scale [28].
- **Learning the early-response of protein signaling networks.** [*S. Videla, S. Thiele, A. Siegel*] We demonstrated the usefulness of the Answer Set Programming approach (ASP) to learn Boolean models from high-throughput phospho-proteomics data. Exact constraint solving showed a quantum leap over heuristic (state-of-the-art) methods in terms of efficiency and scalability, and guarantees global optimality of solutions as well as provides a complete set of solutions [19][Online publication]
- **Numerical model of signaling pathways** [*G. Andrieux, M. Le Borgne*] We have proposed an integrative numerical (ODE) model for the dynamic regulation of $TGF\beta$ Signaling by $TIF1\gamma$. The model successfully unifies the seemingly opposite roles of $TIF1\gamma$, and reveals how changing $TIF1\gamma/Smad4$ ratios affect the cellular response to stimulation by $TGF\beta$, accounting for a highly graded determination of cell fate. [10].

- **Identification of regulatory networks in ecology** [*D. Eveillard*] A clustering data-based approach emphasizes regulatory networks at the bacterial population scale. It allowed the identification of antagonistic interactions between heterotrophic bacteria as a potential regulator of community structure of hypersaline microbial mats. [15][Online publication]

## 6.3. Sequence annotation

**Participants:** François Coste [**contact**], Catherine Belleannée, Gaëlle Garet, Clovis Galiez, Laurent Miclet, Jacques Nicolas.

- **Expressive pattern matching** [*C. Belleannée, J. Nicolas*] We have presented for the first time Logol, a new application designed to achieve pattern matching in possibly large sequences with realistic biological motifs. Logol consists in both a language for describing patterns, and the associated parser for effectively scanning sequences (RNA, DNA or protein) with such motifs. The language, based on an high level grammatical formalism, allows to express flexible patterns (with misparings - improper alignment of DNA strands - and indels) composed of both sequential and structural elements (such as repeats or pseudoknots)[21][Online publication]. Logol has been applied to the detection of -1 frameshifts, a structure including pseudoknots, on a reference benchmark (Recode2) [26][Online publication].
- **Analysis of sequence repeats** [*J. Nicolas*] We have participated to a book that introduces up-to-date methods for the identification and study of transposable elements in genomes. J. Nicolas contributed with a chapter that provides an overview of the formal underpinnings of the search for these highly repeated elements in genomic sequences and describes a selection of practical tools for their analysis. It concludes with the interest of syntactic analysis in this domain [24][Online publication].
- **Grammatical models for local patterns** [*G. Garet, J. Nicolas, F. Coste*] We studied the annotation of new proteins with respect to banks of already annotated protein sequences. For this task, we are developping grammatical inference methods. We introduced new classes of substitutable languages and new generalization criterion based on local substitutability concept and illustrated the great potential of the approach on a benchmark considering a real non trivial protein family. [16][Online publication]
- **Local maximality** [*L. Miclet*] Starting from locally maximal subwords and locally minimal superwords common to a finite set of words, we have defined the corresponding sets of alignments. We gave a partial order relation between such sets of alignments, as well as two operations between them and showed it has a lattice structure that can be used for inducing a generalization of the set of words [18][17].
- **Searching for Smallest Grammars on Large Sequences and Application to DNA** [*F. Coste*] We are motivated by the inference of the structure of genomic sequences, that we address as an instance of the smallest grammar problem. Previously, we reduce it to two independent optimization problems: choosing which words will be constituents of the final grammar and finding a minimal parsing with these constituents. This year we made these ideas applicable on large sequences. First, we improved the complexity of existing algorithms by using the concept of maximal repeats for constituents. Then, we improved the size of the grammars by cautiously adding a minimal parsing optimization step. Together, these approaches enabled us to propose new practical algorithms that return smaller grammars (up to 10%) in approximately the same amount of time than their competitors on a classical set of genomic sequences and on whole genomes. [14] [Online publication].
- **CyanoLyase: a database of phycobilin lyase sequences, motifs and functions** [*F. Coste*] In collaboration with our partners of the ANR project Pelican, we have set up CyanoLyase (http://cyanolyase.genouest.org/), a manually curated sequence and signature database of phycobilin lyases and related proteins. Protomata-Learner has been used to establish the signature of the 32 known subfamilies that are used to rapidly retrieve and annotate lyases from any new genome [13] [Online publication]

# GENSCALE Team

# 6. New Results

## 6.1. Next Generation Sequencing

**Participants:** Alexan Andrieux, Rayan Chikhi, Liviu Ciortuz, Dominique Lavenier, Fabrice Legeai, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo, Erwann Scaon, Raluca Uricaru.

- **Ultra-low memory data structure for de novo genome assembly** : We propose a new encoding of the de Bruijn graph, which occupies an order of magnitude less space than current representations. The encoding is based on a Bloom filter, with an additional structure to remove critical false positives. [24]
- **Transcriptomic variant detection** : We developped a new method, called kissplice, that calls splicing variant events from sets of RNA-seq NGS reads. It constructs the de-Bruijn graph from the reads and then detects in this graph all patterns corresponding to alternative splicing events. [21]
- **Targeted assembly of NGS data**: The method is based on an iterative targeted assembler which processes large datasets of reads on commodity hardware. Basically, it checks for the presence of given regions of interest in the reads and reconstructs their neighborhood, either as a plain sequence (consensus) or as a graph (full sequence structure). [20]
- **Mapping reads on a graph:** We developped a strategy for directly mapping sequences on bi-directed de-Bruijn graphs. Based on a seed-and-extend algorithm it can be applied on large datasets.[31]
- **Pea aphid genomics and evolution.** Using some of the softwares developped by Genscale, genomic variants and expression data of the pea aphid were analysed, revealing candidate regions involved in the adaptation to host plant, and genes involved in the reproduction mode, either with differential expression patterns or particular patterns of evolutionary rates in other aphid species. [11], [12], [19]

## 6.2. Protein structures

**Participants:** Rumen Andonov, Guillaume Chapuis, Dominique Lavenier, Mathilde Le Boudic-Jamin, Antonio Mucherino.

- **Comparison of pairwise protein structure alignments.** The method provides either optimal, top-scoring alignments or heuristic alignments with quality guarantee for some inter-residue distance-based measures. Alignments are compared using a number of quality measures and intuitive visualizations. The methodology brings new insight into the structural relationship of the protein pairs and is a valuable tool for studying structural similarities. [23]
- **Alignment graph.** This object is the main input to find similarities between biomolecules (ARN, proteins). This kind of graph has to model physical and/or chemical properties of the biomolecules and need to take into account constraints dictated by the type of applications (3D comparison, docking, etc.). Our research aims to provide a strategy to automate the building of alignment graphs. A prototype software, called MAGE, is currently under test to validate our approach.
- **Mathematical model and exact algorithm for optimally aligning protein structures.** The algorithm proposes for the first time, to evaluate the popular DALI heuristic in sound mathematical terms. The results indicate that DALI usually computes optimal or close to optimal alignments. However, we detect a subset of small proteins for which DALI fails to generate any significant alignment, although such alignments do exist [22].

- **Modeling the protein flexibility by distance geometry.** We suggest a strategy for modeling protein flexibility that is based on the discretization of the space of possible molecular conformations for a protein. The same discretization process was previously employed for discretizing Molecular Distance Geometry Problems (MDGPs) [30].
- **NMR problems.** We introduce formally the Discretizable Molecular Distance Geometry Problem (DMDGP) for solving the 3D structure of a protein based on Nuclear Magnetic Resonance data together with an algorithm, which we named the "Branch & Prune" (BP), for the solution of DMDGPs [16].We also provide surveys on these recent works about DMDGPs [15], [27].
- **Improvements and variants of the DMDGP**. We exploit symmetries in DMGP trees. We consider similar or related problems (re-ordering of the vertices, relaxing vertices consecutivity assumtion, including side chains and finding low energy homopolymer conformations). Parallelism has also been investigated. [17], [14], [18], [28], [26], [29]

## 6.3. High Throughput Sequence Analysis

**Participants:** Rayan Chikhi, Erwan Drezen, Dominique Lavenier, Claire Lemaitre, Nicolas Maillet, Pierre Peterlongo.

- **Comparing metagenomes.** This research aims to define new ways of comparing billions of sequences generated by NGS sequencers. Standard techniques don't scale with such volume of data, both in terms of memory fingerprint and execution time. We have successfully tested a new method based on probabilistic data structures (Bloom filter) allowing large sets of sequences to be indexed in a short time on standard computers. [25]
- **Bank-to-bank comparison.** In cooperation with the Korilog company we improve the PLAST technology developed for bank-to-bank sequence similarity search. Structuration of the index has been revisited to reduce the memory fingerprint and the execution time. The Korilog company has successfully integrated this improvments sofware component in its own software and has just began its promotion with promising responses from several potential clients.[Korilog promotion]

## 6.4. HPC and Parallelism

**Participants:** Rumen Andonov, Guillaume Chapuis, Charles Deltel, Dominique Lavenier, Fabrice Legeai, François Moreews.

- **High performance pipelines for annotation**. We participed to effort of URGI (INRA Versailles) to set up TriAnnot, a modular architecture allowing for the annotation of genomes. The TriAnnot pipeline is parallelized on a 712 CPU computing cluster that can run a 1-Gb sequence annotation in less than 5 days. [13]
- **Bioinformatics Workflows.** SLICEE is an environment to capture and parallelize time-consuming bioinformatics applications on grid or cloud platforms. In 2012, a web interface has been designed to interactively draw and run workflows from standard browsers ([workflow portal]). Several workflows used in the BioWIC ANR project have been successfully tested on this platform (http://biowic.inria.fr/)
- **Parallelization of a pseudo-clique solver.** Following such solvers as DAST and A_purva, we develop a pseudo-clique solver for alignment graphs. Looking for pseudo-cliques allows us to relax some of the constraints that are inherent to clique finding and thus maintain polynomial run times. We focus on defining a parallel algorithm and developping an implementation that benefits from multiple levels of parallelism: fine grain parallelism (bit-level parallelism, SSE instructions) and coarse grain parallelism (mutli-core parallelism). Intended applications range from protein local similarity search to protein surface similarity search or even docking.

## IBIS Project-Team

# 5. New Results

## 5.1. Analysis of gene regulatory networks by means of piecewise-linear (PL) models

GENETIC NETWORK ANALYZER (GNA) is a tool for the qualitative modeling and simulation of the dynamics of gene regulatory networks by means of PL models, as described in Section 4.1 . GNA has been integrated with the other bioinformatics tools distributed by Genostar (http://www.genostar.com/). Version 8.4 of GNA was released by IBIS and Genostar this year. This version is an update of version 8.0, deposited at the Agence pour la Protection des Programmes (APP). Some bugs have been corrected in the new version and the program has been adapted to the latest versions of Java and the software platform of Genostar. A book chapter describing the current version of GNA has been published in a volume on the modeling of bacterial molecular networks [13]. The chapter is a tutorial illustrating the practical use of recent functionalities of GNA like the network editor and the formal verification module by means of an example network in *E. coli*.

The predictions obtained with the help of GNA are purely qualitative, describing the dynamics of the network by means of a state transition graph. While a qualitative analysis is appropriate for certain problems, the absence of precise quantitative predictions may not be desirable in others, such as the analysis of a limit cycle or the design of a controller for a synthetic network. The quantitative study of PL models of gene regulatory networks is hindered by the fact that the step functions describing the logic of regulatory interactions lead to discontinuities in the right-hand side of the PL models (Section 3.2 ). This has motivated extensions of the PL models based on differential inclusions and Filippov solutions. As of now, no numerical simulation tool for the simulation of these Filippov extensions is available. In collaboration with the BIPOP project-team, we have shown how tools developed for the simulation of nonsmooth mechanical, electrical and control systems can be adapted for this purpose. A paper describing these results is being prepared for submission.

## 5.2. Experimental mapping of gene regulatory networks in bacteria

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 3.3 ). 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 4.2 ), to allow biologists to make the most of the information contained in reporter gene expression data. Several improvements of the platform for measuring gene expression are the subject of ongoing work, including a novel method for efficiently cloning reporter gene constructions on the chromosome of *E. coli*.

These tools have been used in a series of studies directed at the experimental mapping of gene regulatory networks in *E. coli*. One example, carried out in the framework of the PhD thesis of Guillaume Baptist, is the development of a new screening methodology for identifying all genes that control the expression of a target gene through genetic or metabolic interactions. The screen combines mutant libraries with luciferase reporter constructs. Instead of a static picture of gene expression, this method allows dynamical monitoring in different environmental conditions. Mutants with interesting phenotypes can thus be selected based on multiple criteria, and the expression dynamics of the target gene can be extensively characterized. The method has been applied to the identification of the direct and indirect regulators of the gene *acs* in *Escherichia coli*. We confirmed known genetic regulators of the gene and identified new regulatory influences, many of which involve metabolic intermediates or metabolic sensing. An analysis of mutants involved in glycolysis and glucose transport demonstrates that the classical model of catabolite repression in *E. coli* needs to be amended. A paper describing the above work is currently under revision.

Other examples of on-going work are the analysis of the network involved in motility and sessility and the modulation of the RpoS regulon in *E. coli* by Stephan Lacour, the analysis of the regulation of cAMP levels in the bacterial cell by Claire Villiers, and the analysis of various aspects of the regulation of carbon metabolism by Valentin Zulkower and Stéphane Pinhal.

## 5.3. Analysis of metabolic coupling in gene regulatory networks

The regulation of gene expression is tightly interwoven with metabolism and signal transduction. A realistic view of genetic regulatory networks should therefore not only include direct interactions resulting from transcription regulation, but also indirect regulatory interactions mediated by metabolic effectors and signaling molecules. We coined the term metabolic coupling to denote these indirect interactions mediated by metabolism. Ignoring metabolic coupling during the analysis of the network dynamics may lead crucial feedback loops to be missed.

In previous work, published in *PLoS Computational Biology* in 2010, we showed how indirect interactions arising from metabolic coupling can be derived from a model of the underlying biochemical reaction network. We applied this approach to the carbon assimilation network in *Escherichia coli* investigating how the structural properties of the network are modified by the inclusion of metabolic interactions. Our results showed that the derived gene regulatory network is densely connected, contrary to what is usually assumed. Moreover, we found that the signs of the indirect interactions are largely fixed by the direction of metabolic fluxes, independently of specific parameter values and rate laws, and that a change in flux direction may invert the sign of indirect interactions. This leads to a feedback structure that is at the same time robust to changes in the kinetic properties of enzymes and that has the flexibility to accommodate radical changes in the environment.

It remains an open question, however, to which extent the indirect interactions induced by metabolic coupling affect the dynamics of the system. This is a key issue for understanding the relative contributions of the regulation of gene expression and metabolism during the adaptation of the cell to changes in its environment. In collaboration with Valentina Baldazzi, formerly post-doctoral fellow in IBIS and now research scientist at INRA (Avignon), we have carried out a dynamic analysis by developing a qualitative PL model of the gene regulatory network, including both the direct and indirect interactions.

In order to obtain a clearer view of the dynamic role of metabolic coupling in the adaptation of gene expression, we developed several qualitative models corresponding to a network topology including all, some, or none of the indirect interactions. The dynamical properties of the models were analyzed and compared with available experimental data using the computer tool GNA (Section 4.1 ). In particular, we compared the steady-state concentrations of enzymes and transcription regulators during growth on glucose and acetate, as well as the dynamic response of gene expression to the exhaustion of glucose and the subsequent assimilation of acetate. We find significant differences between the dynamics of the system in the absence and presence of metabolic coupling. This confirms that indirect interactions are essential for correctly reproducing the observed adaptation of gene expression to a change in carbon source. Our work thus underlines the importance of metabolic coupling in gene regulatory networks, and shows that such indirect interactions cannot be neglected when studying the adaptation of an organism to changes in its environment. A paper describing these results has been published in the *Journal of Theoretical Biology* [5]. Another publication, reviewing the applicability of these and other ideas for multi-scale modeling in plants, has appeared in *Trends in Plant Science* [4].

## 5.4. Parameter estimation for kinetic models of carbon metabolism in bacteria

Kinetic models capture the dynamics of the large and complex networks of biochemical reactions that endow bacteria with the capacity to adapt their functioning to changes in the environment. In comparison with the qualitative PL models described in Sections 5.1 and 5.3 , these more general classes of ODE models are intended to provide a quantitative description of the network dynamics, both on the genetic and metabolic level. New experimental techniques have led to the accumulation of large amounts of data, such as time-course measurements of metabolite, mRNA and protein concentrations and measurements of metabolic fluxes under different growth conditions. However, the estimation of parameter values in the kinetic models from these data remains particularly challenging in biology, mostly because of incomplete knowledge of the molecular mechanisms, noisy, indirect, heterogeneous, and partial observations, and the large size of the systems, with dynamics on different time-scales. We have addressed parameter estimation in the context of the analysis of the interactions between metabolism and gene expression in carbon metabolism in *E. coli*.

In collaboration with Matteo Brilli and Daniel Kahn (INRA and Université Claude Bernard in Lyon), we previously developed an approximate model of central metabolism of *E. coli*, as described in an article published in *Bioinformatics* in 2011. The model was based on the use of so-called linlog functions to approximately describe the rates of enzymatic reactions. More precisely, linlog models define reactions rates as proportional to both the enzyme concentrations and a linear combination of the logarithms of metabolite concentrations. The estimation of parameters in the linlog model from metabolomics, transcriptome, proteomics data sets required the development of a new approach, adapted to the occurrence of numerous missing values in the data sets. When applied to the above-mentioned linlog model, exploiting a high-throughput dataset published in the literature, we were able to obtain reasonable estimates of the 100 parameters.

The results of the above application also revealed the fundamental role played by the identifiability of the model parameters, an issue often overlooked in systems biology. This prompted us for a thorough investigation of the concepts of structural identifiability (in presence of perfect, idealized data), practical identifiability (in presence of noisy and limited amounts of data), and the relations among the two. In addition, we looked into the implications of this analysis for the reduction of nonidentifable to identifiable models. While having a solid mathematical basis, the study was tailored to the actual experimental practice, and resulted in a practical model reduction method that improves upon our previous approach in case of large measurement noise. This study, and the results from its application to both *in-silico* case studies and state-of-the-art datasets, were reported in a paper that has been accepted for publication in the *Journal of Mathematical Biology* [6] (see also [11] for a short version with preliminary results).

A second line of work is based on the use of classical kinetic models that are, in comparison with the above-mentioned linlog models, much reduced in scope (the focus is on the metabolic and genetic regulation of the glycolysis pathway) and granularity (individual reactions are lumped together). The models, developed by Delphine Ropers, have been calibrated using experimental data from the experimental part of the IBIS group for the gene expression measurements and the group of Jean-Charles Portais at INSA in Toulouse for the measurements of metabolism. The model with the estimated parameter values is currently being tested and used to understand some key mechanisms in the adaptation of *E. coli* to the exhaustion of glucose. The PhD thesis of Manon Morin, which started at the end of this year in the framework of a collaboration supported by a Contrat Jeune Scientifique INRA-Inria, will further develop these research directions.

## 5.5. Structural identification of gene regulatory networks

In general, structural identification of genetic regulatory networks involves fitting appropriate network structures and parameters to the data. While modern measurement techniques such as reporter gene systems provide data of ever-increasing quality, the problem remains challenging because exploring all possible network structures in the search of the best fitting model is prohibitive.

In order to address the structural identification problem, Eugenio Cinquemani developed in collaboration with the Automatic Control Lab at ETH Zürich (Switzerland) and the Computer Engineering & Systems Science Department of the University of Pavia (Italy), an ODE modelling framework based on so-called unate-like functions, and a method that exploits monotonicity properties of these functions to effectively prune models that are incompatible with the data from the family of all unate-like modelling alternatives. This model invalidation step is based on simple preprocessing of time-course protein concentration and synthesis rate profiles, assumed available, and allows one to reduce the search of the best fitting model to a small subset of viable model structures.

The method, first published in *Bioinformatics* in 2010 and demonstrated on real data from the synthetic network IRMA, allows one to integrate *a-priori* knowledge on the expected network dynamics in a natural way. Leveraging on this, in the context of the same international collaboration, the method has been further developed in particular by considering relevant subclasses of the family of unate-like models that also enjoy certain quasi-convexity properties. For this restricted class, combined use of monotonicity and quasi-convexity properties allows one to ameliorate the model invalidation step, *i.e.* retain even fewer viable model structures based on affordable data preprocessing. These developments have been presented and demonstrated *in silico* in a paper published in the 2012 special issue on System Identification for Biological Systems of the *International Journal of Robust and Nonlinear Control* [9] .

We are currently applying the above methods to actual, known or partially unknown, networks. In the framework of the PhD thesis of Diana Stefan, the network inference method has been applied to gene expression data from the network regulating motility of *E. coli*. First encouraging results have suggested further experimental and computational investigations that are currently in progress.

## 5.6. Stochastic modeling and identification of gene regulatory networks in bacteria

At the single-cell level, the processes that govern gene expression are often 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 is tremendous, ranging from a better comprehension of the biochemical regulatory mechanisms underlying life, to the development of new strategies for the control of cell populations and even of single cells. General modeling paradigms, such as the Chemical Master Equation, exist for the description of stochastic dynamics at the single-cell level. However, due to the complexity of the interactions, current studies have often preferred to focus on specific cases of interest by *ad-hoc* modeling and analysis. In addition, theoretical and practical challenges inherent in the inference of stochastic models from biological experimental data have limited the development of general identification approaches.

In view of the potential and the relevance of the subject, one research line of IBIS is dedicated to the probabilistic modeling of the dynamics of gene regulatory networks at the level of individual cells. Our activity is centered around two main challenges. On the one hand, we address the problem of developing methods for fitting unknown network parameters of stochastic models to experimental data. As a reference case study we consider the network regulating the inset of the arabinose uptake process in *E. coli* upon depletion of glucose in the growth medium. For this system, Eugenio Cinquemani and Michel Page are developing and implementing methods for the inference of unknown parameters from fluorescence microscopy data. On the other hand, we investigate several alternative modelling approaches in an attempt to determine their relevance to different systems and application scenarios. This activity is being developed in collaboration with Gregory Batt (CONTRAINTES, Inria Paris-Rocquencourt), Giancarlo Ferrari-Trecate (University of Pavia, Italy), and Alfonso Carta (COMORE, Inria Sophia-Antipolis - Méditerranée). First results connected to control applications on real and simulated data have been submitted for presentation at the European Control Conference to be held in 2013. Finally, further ongoing work concerns the study of noise propagation in gene regulatory networks, in collaboration with Irina Mihalcescu (Université Joseph Fourier), and the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments, in collaboration with Marianna Rapsomaniki and Zoi Lygerou (University of Patras, Greece) and John Lygeros (ETH Zürich, Switzerland).

## 5.7. Control of regulatory networks in bacteria

While systems biology is primarily concerned with natural systems shaped by evolution, synthetic biology opens up a new generation of fundamental research by trying to redesign natural systems or create novel systems from scratch. Mathematical modeling and analysis are essential components of synthetic biology, as they help understanding the consequences of (changes in) the network of interactions on the dynamical behavior of the system. More specifically, a model can be a powerful tool for the control and regulation of the system towards a desired goal.

Within the projects ColAge and GeMCo (Section 7.2 ), we attempt to control one of the fundamental physiological properties of bacterial cells, their growth rate. In particular, in order to control the growth rate, we propose to focus on the gene expression machinery of *E. coli*, whose activity is controlled by a complex regulatory network with many components and intertwined feedback loops. Delphine Ropers is developing models of the gene expression machinery and Jérome Izard, in the context of his PhD thesis, is rewiring part of the network to enable control of the network dynamics. The results on these projects are currently being prepared for publication.

## 5.8. Shared control of gene expression by global physiological effects and specific regulators

Gene expression is controlled by the joint effect of (i) the global physiological state of the cell, in particular the activity of the gene expression machinery, and (ii) DNA-binding transcription factors and other specific regulators. While many studies have focused on networks of transcription factors, the analysis of the relative contributions of both transcription factors and global effects of the physiological state has received relatively little attention thus far.

In the framework of the PhD thesis of Sara Berthoumieux, we have developed a model-based approach to distinguish between these two effects using time-resolved measurements of promoter activities. We have demonstrated the strength of the approach by analyzing a circuit involved in the regulation of carbon metabolism in *E. coli*, consisting of two pleiotropic regulators of the cell (Crp and Fis), the gene *acs* encoding the enzyme acetyl-CoA synthetase (Acs), and the signaling metabolite cyclic AMP (cAMP) which activates Crp. *acs* is strongly expressed in the absence of glucose and is thus an excellent indicator of the transcriptional response of carbon metabolism to a growth-phase transition.

Our results show that the transcriptional response of the network is controlled by the physiological state of the cell and the signalling metabolite cAMP. The (surprising) absence of a strong regulatory effect of transcription factors suggests that they are not the main coordinators of gene expression changes during growth transitions, but rather that they complement the effect of global physiological control mechanisms. This change of perspective has important consequences for the interpretation of transcriptome data and the design of biological networks in biotechnology and synthetic biology. An article presenting the above results has been accepted for *Molecular Systems Biology* [7].

# MAGNOME Project-Team

# 6. New Results

## 6.1. Yeast comparative genomics

**Participants:** Pascal Durrens [correspondant], Tiphaine Martin, David James Sherman.

By using MAGNOME's MAGUS system and YAGA software, we have successfully realized a full annotation and analysis of seven new genomes, provided to the Génolevures Consortium by the CEA–Génoscope (Évry)[15]. Two distant genomes from the *Debaryomycetaceae* and *mitosporic Saccharomycetales* clades of the *Saccharomycetales* were annotated using previously published Génolevures genomes [6], [10], [11] as references (in prep.). A further group of five species, comprised of pathogenic and nonpathogenic species, was analyzed with the goal of identifying virulence determinants [37]. By choosing species that are highly related but which differ in the particular traits that are targeted, in this case pathogenicity, we are able to focus of the few hundred genes related to the trait (in rev.). The approximately 40,000 new genes from these studies were classified into existing Génolevures families as well as branch-specific families.

In collaboration with partners in the ISVV, Bordeaux, we have assembled and analyzed 12 wine starter yeasts, with the goal of understanding genetic determinants of performance (in prep.).

## 6.2. Assembly, annotation and comparison of bacterial Omics data

**Participants:** Elisabeth Bon [correspondant], Laetitia Bourgeade, Pascal Durrens, Aurélie Goulielmakis, Tiphaine Martin, David James Sherman.

*Oenococcus oeni* is part of the natural microflora of wine and related environments, and is the main agent of the malolactic fermentation (MLF), a step of wine making that generally follows alcoholic fermentation (AF) and contributes to wine deacidification, improvement of sensorial properties and microbial stability. The start, duration and achievement of MLF are unpredictable since they depend both on the wine characteristics and on the properties of the *O. oeni* strains. In collaboration with Patrick Lucas's lab of the ISVV Bordeaux that is currently proceeding with genome sequencing, explorative and, and comparative genomics, Elisabeth Bon coordinates our efforts into the OENIKITA project (since 2009), a scale switching challenge including highthrouput exploratory and comparative genomics for oenological bacterial starters, and the development of an online web-collaborative multigenomic comparative platform based on the the Génolevures database architecture and MAGUS / YAGA systems.

**OENI-Genomics**: In comparative genomics, we investigated gene repertoire and genomic organization conservation through intra- and inter-species genomic comparisons, which clearly show that the *O. oeni* genome is highly plastic and fast-evolving. Results reveal that the optimal adaptation to wine of a strain mostly depends on the presence of key adaptive loops and polymorphic genes. They also point up the role of horizontal gene transfer and mobile genetic elements in O. oeni genome plasticity, and give the first clues of the genetic origin of its oenological aptitudes[3], [14], [29], [33], [35], [36]. As a result of the scaling out challenge, we participated to the assembly and annotation of 19 fully sequenced *O. oeni* genome variants.

**KITA-Genomics (E. Bon, D. Sherman)**: This project that is focused on the sequencing, assembly, exploration and comparison of the *O. kitaharae* genome, has benefited to an international collaboration involving Dr V. Makeev. MAGNOME was involved into the pilot assembly, exploration and comparison of the *O. kitaharae* genome.

**Transcriptomic axis (E. Bon, A. Goulielmakis, P. Durrens)**: Under the supervision of E. Bon, Aurélie Goulielmakis has completed for the ANR DIVOENI a detailed manual annotation of a new reference strain of *O. oeni* and performed comparative transcriptome analysis to identify genes differentially expressed under different culture conditions. We explored and compared how the expression system is solicited when *O. oeni* strains adapted to grow in some niches are placed under stress-exposure conditions. The monitoring of gene expression status between strains, through the definition of a global expression pattern proper to each gene, partially lift the veil on how *O. oeni* genome adapts function to its environment. The weight of genetic background and ecological niche pressure on gene expression flexibility was evaluated, and the *O. oeni* pan-transcriptome architecture characterized. The first guidelines revealed a supra-spatial organization of stress response into activated and repressed larger macro-domains defining functional landmarks and intra-chromosomal territories. Decryption of stress-sensitive gene repertoires promises to be an efficient tool in the conquest of *O. oeni* “domestication” through the identification of molecular markers responsible for different physiological capabilities, and the selection of the best adapted strains [21], [43].

**Gene plasticity modelisation (E. Bon, L. Bourgeade)**: A novel axis of research recently emerged under the initiative of E. Bon (pseudOE project) around the detection, characterization and conservation of pseudogenes populations in *Oenococcus* bacteria. Such topic presents a double interest: phylogenetic at first because it should allow to better estimate the degree of genic/genomic plasticity of these bacteria, and algorithmic then because the pseudogenes are a source of confusion for the automatic prediction of genes. Through a transversal collaboration and a cooperative supervision with the Algorithms for Analysis of Biological Structures Group (P. Ferraro, J. Allali) at LaBRI, Laetitia Bourgeade (PhD, Univ. Bordeaux1) was recruited to develop dedicated methods to improve pseudogenes automatic detection, and therefore gene predictions, and to reconstruct fossil and modern genes evolutionary history [20], [23].

## 6.3. Big Data in comparative genomics

**Participants:** David James Sherman [correspondant], Pascal Durrens, Natalia Golenetskaya, Florian Lajus, Tiphaine Martin.

Data growth in comparative genomics presents a significant scaling challenge that requires novel informatic methods. Increase in sequence data is already a challenge, but in addition, the *relations* between the biological objects increase supralinearly (geometrically in the worst case) for every linear increase in sequence data.

MAGNOME’s Tsvetok system proposes a highly-scalable distributed approach for data and computation in comparative genomics, targeting projects of the “comparative genomics of related species” type, where a set of genomes is sequenced and analyzed as part of the same process. Tsvetok combines a novel NoSQL storage schema with domain-specific MapReduce algorithms, to efficiently handle the fundamentally data-parallel analyses encountered in comparative genomics. Natalia Golenetskaya with Florian Lajus derived use cases from web site log analyses to identify standard queries, define an appropriate query-oriented storage schema, and map structured values to this schema. This was tested in MAGNOME’s dedicated computing cluster.

Natalia Golenetskaya furthermore defined new distributed algorithms for two important large-scale analyses in MAGNOME’s pipeline: systematic identification of gene fusion and fission events in eukaryote genomes (following [7]), and large-scale consensus clustering for protein families (following [9]). For fusions and fissions, she defined a new MapReduce algorithm that avoids graph-based analysis (which is notoriously slow in MapReduce), to achieve both significant speed ups and excellent scaling to much larger data sets. For protein family clustering, she defined a novel iterative sampling strategy that combines parallel clustering of submatrices of pairwise relations, to succesively approximate the result of a complete clustering, without the need to store the entire matrix of relations in memory.

## 6.4. Inferring metabolic models

**Participants:** David James Sherman [correspondant], Pascal Durrens, Razanne Issa, Anna Zhukova.

In collaboration with Prof Jean-Marc Nicaud's lab at the INRA Grignon, we developed the first functional genome-scale metabolic model of an oleaginous yeast. Most work in producing genome-scale metabolic models has focused on model organisms, in part due to the cost of obtaining well-annotated genome sequences and sufficiently complete experimental data for refining and verifying the models. However, for many fungal genomes of biotechnological interest, the combination of large-scale sequencing projects and in-depth experimental studies has made it feasible to undertake metabolic network reconstruction for a wider range of organisms.

An excellent representative of this new class of organisms is *Yarrowia lipolytica*, an oleaginous yeast studied experimentally for its role as a food contaminant and its use in bioremediation and cell factory applications. As one of the hemiascomycetous yeasts completely sequenced in the Génolevures program it enjoys a high quality manual annotation by a network of experts. It is also an ideal subject for studying the role of species-specific expansion of paralogous families, a considerable challenge for eukaryotes in genome-scale metabolic construction. To these ends, we undertook a complete reconstruction of the *Y. lipolytica* metabolic network.

Methods: A draft model was extrapolated from the *S. cerevisiae* model iIN800, using *in silico* methods including enzyme conservation predicted using Génolevures and reaction mapping maintaining compartments. This draft was curated by a group of experts in *Y. lipolytica* metabolism, and iteratively improved and validated through comparison with experimental data by flux balance analysis. Gap filling, species-specific reactions, and the addition of compartments with the corresponding transport reactions were among the improvements that most affected accuracy. These methods, initially implemented in an *ad hoc* way in the *Pathtastic* software tool, have been redefined and formalized by Razanne Issa using a novel logical framework.

Results: We produced an accurate functional model for *Y. lipolytica*, MODEL1111190000 in Biomodels.net, that has been qualitatively validated against gene knockouts. This model has been enriched by Anna Zhukova with ontology terms from ChEBI and GO.

## 6.5. Summarized visualization of metabolic models

**Participants:** David James Sherman [correspondant], Anna Zhukova.

In collaboration with Romain Bourqui and Antoine Lambert of the LaBRI, we defined new strategies for exploring whole genome metabolic models. There is an inherent tension between detail and understandability in these large networks: on the one hand, detailed description of individual reactions is needed for accurate simulation, but on the other hand, high-level views of reactions are needed for describing partways in human terms. We are defining knowledge-based simplification rules, that permit the user to factor similar reactions into one "generic" reaction in order to visualize a whole pathway or compartment, while maintaining the underlying model so that the user can later "drill down" to the specific reactions if need be. New layout rules implemented in the Tulip platform are used to draw the resulting networks in a familiar way.

In collaboration with Bruno Pinaud of the LaBRI, rule-based rewriting of metabolic models was used to define these simplifications using his PORGY software tool.

## 6.6. Hierarchical modeling with BioRica

**Participants:** David James Sherman [correspondant], Rodrigo Assar Cuevas, Nicolás Loira.

A recurring challenge for *in silico* modeling of cell behavior is that experimentally validated models are so focused in scope that it is difficult to repurpose them. Hierarchical modeling is one way of combining specific models into networks. Effective use of hierarchical models requires both formal definition of the semantics of such composition, and efficient simulation tools for exploring the large space of complex behaviors.

BioRica is a high-level hierarchical modeling framework for models combining continuous and discrete components. By providing a reliable and functional software tool backed by a rigorous semantics, we hope to advance real adoption of hierarchical modeling by the systems biology community. By providing an understandable and mathematically rigorous semantics, this will make is easier for practicing scientists to build practical and functional models of the systems they are studying, and concentrate their efforts on the system rather than on the tool.

Building on previous work that formalized strategies for integrating discrete control with continuous models, Rodrigo Assar defined a new framework for BioRica models using Kaufman's Quantized State Systems (in prep.).

## MORPHEME Team

# 5. New Results

## 5.1. Imaging

### 5.1.1. ML estimation of wavelet regularization hyperparameters in inverse problems

**Participant:** Laure Blanc-Féraud.

*This work was made in collaboration with Caroline Chaux from LATP (Marseille) and Roberto Cavicchioli and Luca Zanni from University of Modena (Italy).*

Parameter estimation, Maximum likelihood estimation, Wavelet transforms, Deconvolution, Gradi- ent methods

We are interested in regularizing hyperparameter estimation by maximum likelihood in inverse problems with wavelet regularization. One parameter per subband is estimated by gradient ascent algorithm.We have to face with two main difficulties: i) sampling the a posteriori image distribution to compute the gradient of the objective function; ii) choosing a suited step-size to ensure good convergence properties of the gradient ascent algorithm. We first show that introducing an auxiliary variable makes the sampling feasible using classical Metropolis-Hastings algorithm and Gibbs sampler. Secondly, we propose an adaptive step-size selection and a line-search strategy to improve the gradient-based method. Good performances of the proposed approach are demonstrated on both synthetic and real data.

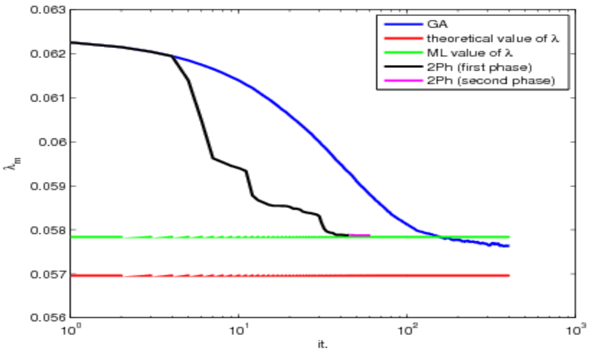


*Figure 1. $\lambda_m$ behavior over iterations of ascent algorithm for a sub band at first level of wavelet decomposition*

### 5.1.2. Joint optimization of noisy image coding and denoising

**Participants:** Mikael Carlavan, Laure Blanc-Féraud.

*this work was made in collaboration with Marc Antonini (I3S), Roberto Camarero and Christophe Latry (CNES) and Yves Bobichon (TAS).*

coding, denoising, wavelet transform, global rate-distortion optimization

This work concerns the study of optimal noisy source coding/denoising. A global optimization of the problem is usually difficult to perform as the global fidelity criterion needs to be optimized in the same time over the sets of both coding and denoising parameters. Most of the bibliography in this domain is based on the fact that, for a specific criterion, the global optimization problem can be simply separated into two independent optimization problems: The noisy image should be first optimally denoised and this denoised image should then be optimally coded. In many applications however, the layout of the acquisition imaging chain is fixed and can not be changed, that is a denoising step can not be inserted before coding. For this reason, we are concerned here with the problem of global joint optimization in the case the denoising step is performed, as usual, after coding/decoding. In this configuration, we showed on a simple case how to express the global distortion as a function of the coding and denoising parameters. We presented an algorithm to minimize this distortion to get the optimal values of these parameters. Figure 2 shows results of this joint optimization algorithm, on the classical test image *Barbara*, in comparison to the usual disjoint optimization technique, which consists in selecting the coding and the denoising parameters such that the coding and the denoising errors are independently minimized. On the range of validity of the proposed model, we see that the joint optimized distortion slightly outperforms the disjoint optimized distortion (in the presented example, the PSNR of the reconstructed image increases of $0.4dB$ at $1.85$ bits/pixels). The interesting point of the proposed method is that it allows to reach the same global error than the disjoint optimized technique but for a lower coding rate. For example, on this image, the joint optimization technique reaches at $1.42$ bits/pixel the same distortion than the one obtained at $2.04$ bits/pixels for the disjoint optimization technique. The benefit in terms of compression performances of the joint optimization appears then to be very significant.

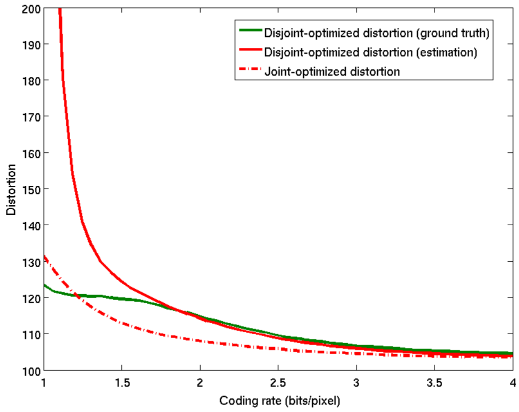


*Figure 2. Comparison of the disjoint optimized distortion (ground truth and estimation) to the joint optimized distortion Barbara.*

### 5.1.3. Blind deconvolution

**Participants:** Saima Ben Hadj, Laure Blanc-Féraud.

*This research takes place within the ANR DIAMOND. This work was made in collaboration with Gilles Aubert, Laboratoire J. Dieudonné (CNRS,UNS).*

One of our tasks within the ANR Diamond project is the blind restoration of images coming from Confocal laser scanning microscopy (CLSM). CLSM is a powerful technique for studying biological specimens in three dimensions by optical sectioning. Nevertheless, it suffers from some artifacts. First, CLSM images are affected by a depth-variant (DV) blur due to spherical aberrations induced by refractive index mismatch between the different media composing the system as well as the specimen. Second, CLSM images are corrupted with a

Poisson noise due to low illumination. Because of these intrinsic optical limitations, it is essential to remove both DV blur and noise from these images by digital processing.

In this context, we first study space-variant (SV) blur models and prove that a model where the SV point spread function (PSF) is approximated by a convex combination of a set of space-invariant (SI) PSFs is efficient and adequate to the inversion problem [30] [10]. Afterwards, we focus on the non-bind restoration problem and we fit a fast restoration method based on a domain decomposition technique [33] to our DV blur model [10], [9].

Recently, we focus on the blind case. In fact, in practice it is difficult to obtain the DV PSF in spite of the existence of theoretical PSF models [34], because these models are dependent on some unknown acquisition parameters (e.g. the refractive index (RI) of the specimen). Therefore a blind or semi-blind restoration algorithm is needed for this system. We propose two methods for this problem : In the first method, we define a criterion to be jointly minimized w.r.t to the image and the PSF set. In this method, the intensities of each SI PSF are estimated at every voxel. Although the big number of parameters to be estimated, the method allows more freedom on the shape of the PSF which could be more or less deformed according to spherical aberration level. We provide a theoretical proof of the existence of a minimizer of the considered problem [23]. Then, we perform the minimization by following an alternate minimization scheme, each elementary minimization is performed using the recently proposed scaled gradient projection (SGP) algorithm that has shown a fast convergence rate [29]. Results on simulated CLSM images and comparison with another alternate scheme based on a regularized version of the Richardson–Lucy algorithm [31] are shown in Fig. 3 . In the second blind method, we use a Gaussian approximation of each of the SI PSFs. This presents the advantage of significantly reducing the number of parameters to be estimated but constraints the PSF shape. We prove on simulated data that the first method provides more accurate restoration result than the second one.

(a) (b) (c) (d)

*Figure 3. $(Y, Z)$ sections of blind restoration results on a simulated CLSM image. (a) original image, (b) simulated observation, (c) restoration using our blind restoration method using SGP algorithm, (d) restoration using a regularized Richardson–Lucy algorithm embedded in an alternate scheme.*

### *5.1.4. Morphogenesis of living organisms*

**Participant:** Grégoire Malandain.

*This research takes place within the Inria Large-scale initiative Morphogenetics.*

*This work was made in collaboration with Christophe Godin and Léo Guignard from Virtual Plants.*

super-resolution, SPIM, morphogenesis

We extended a previous work [32] for the reconstruction of microscopic images. In particular, we extended the super-resolution image reconstruction (where several images, acquired from different viewpoints, are fused) to the lightsheet (or SPIM for Selective Plane Illumination Microscope) microscope modality. This modality offers a high acquisition speed, allowing imaging an organism frequently. As an exemple, Phallusia mammillata and Ciona intestinalis embryos can be imaged from 32 cells to around 1000 cells. The organism is captured from four different angles every 2 minutes during 2 hours (collaboration with CRBM Montpellier and EMBL Heidelberg).

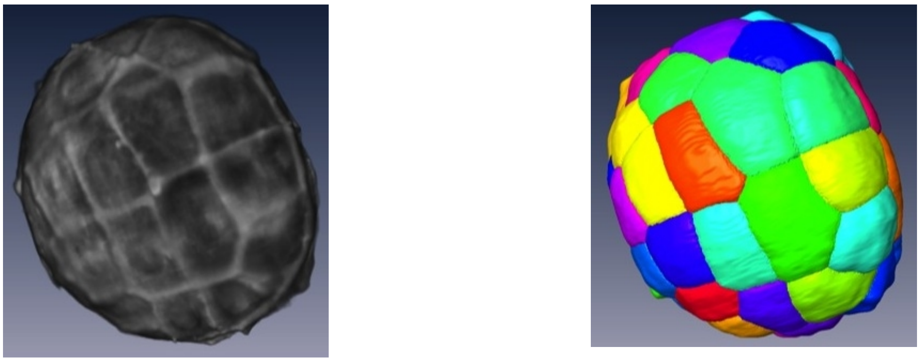


*Figure 4. Left: 3D rendering of a reconstructed image of a Phallusia mammillata embryo; right: segmentation of cells.*

## 5.2. Features Extraction

### *5.2.1. Axon extraction from fluorescent confocal microscopy images*

**Participants:** Alejandro Mottini, Xavier Descombes, Florence Besse.

It it known that the analysis of axonal topologies allows biologists to study the causes of neurological diseases such as Fragile X Syndrome and Spinal Muscular Atrophy. In order to perform the morphological analysis of axons, it is first necessary to segment them. Therefore, the automatic extraction of axons is a key problem in the field of neuron axon analysis.

For this purpose, biologists label single neurons within intact adult Drosophila fly brains and acquire 3D fluorescent confocal microscopy images of their axonal trees. These images need to be segmented.

In our work presented in [16], we propose a new approach for the automatic extraction of axons from fluorescent confocal microscopy images which combines algorithms for filament enhancement, binarization, skeletonization and gap filling in a pipeline capable of extracting the axons containing a single labeled neuron. Unlike other segmentation methods found in the literature, the proposed is fully automatic and designed to work on 3D image stacks. This allows us to analyze large image databases.

The method performance was tested on 12 real 3D images and the results quantitatively evaluated by calculating the RMSE between the tracing done by an experienced biologist and the automatic tracing obtained by our method. The good results obtained in the validation show the potential use of this technique in helping biologists for extracting axonal trees from confocal microscope images (see figures 5 and 6 ).

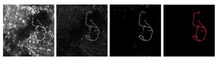


*Figure 5. Results obtained on each step of the algorithm for one image stack (maximum intensity projections). From left to right: original image, filament enhancement, binarization and final result.*

### *5.2.2. Dendrite spine detection from X-ray tomographic volumes*

**Participants:** Anny Hank, Xavier Descombes, Grégoire Malandain.

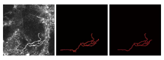


*Figure 6. Comparison between original image (left), our result (middle) and ground truth (right) for two images (maximum intensity projections).*

We have developped an automated algorithm for detecting dendritic spines from XRMT data. XRMT data allows imaging a large volume of tissue, and therefore a higher number of spines than laser scanning microscopy. We have shown that despite the lower image quality compared to microscopic data, we were able to extract dendritic spines. The main idea of the proposed approach is to define a mask for performing the spine detection without facing the false alarms problem as we introduce some information on spines localization. We therefore first extract the dendrites themselves and then compute the spine mask based on prior knowledge on their distance to dendrites. To extract dendrite we first compute the medial axis thanks to a multi-scale Hessian-based method. Then, we extract segments by a 3D Hough transform and reconstruct the dendrites using a conditional dilation. The spine mask is defined nerby the detected dendrites using anatomical parameters described in the literature. A point process defined on this mask provides the spine detection.

To exemplify the proposed approach, a subvolume ($220 \times 180 \times 100$) has been extracted from a XRMT volume that is given on figure 7 . As expected, the spines appear as small objects, whose size is close to the image resolution, along the tubular structures representing dendrites. Using the localization information to detect spine is essential to prevent false alarms due to noise or to the deviation of dendrites from a cylinder model. Figure 7 shows the detected dendrite medial axis and the obtained spine detection. The obtained results are promising and correspond to a visual inspection of the data. Forthcoming validation study will allow to better assess the quality of the detection by providing a quantitative evaluation.

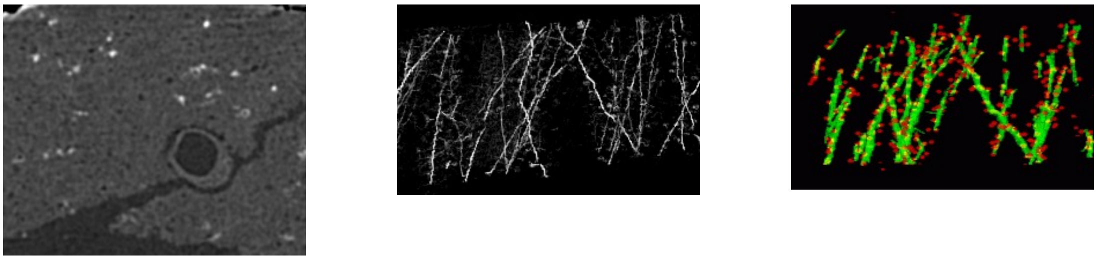


*Figure 7. XMRT slice (left), dendrites medial axis (middle) and spine detection (in red) (right)*

### 5.2.3. Cell detection

**Participant:** Xavier Descombes.

*This work was done in collaboration with Emmanuel Soubies and Pierre Weiss from ITAV (Toulouse)*

We have proposed some improvements of the Multiple Birth and Cut algorithm (MBC) in order to extract nuclei in 2D and 3D images. We have introduced a new contrast invariant energy that is robust to degradations encountered in fluorescence microscopy (e.g. local radiometry attenuations). Another contribution of this work is a fast algorithm to determine whether two ellipses (2D) or ellipsoids (3D) intersect. Finally, we propose a new heuristic that strongly improves the convergence rates. The algorithm alternates between two birth steps. The first one consists in generating objects uniformly at random and the second one consists in perturbing the current configuration locally. Performance of this modified birth step is evaluated and examples on various image types show the wide applicability of the method in the field of bio-imaging.

Figure 8 left shows the segmentation result on a Drosophila embryo obtained using SPIM imaging. This is a rather easy case, since nuclei shapes vary little. The images are impaired by various defects: blur, stripes and attenuation. Despite this relatively poor image quality, the segmentation results are almost perfect. The computing time is 5 minutes using a C++ implementation. The image size is 700 × 350. Figure 8 right presents a more difficult case, where the image is highly deteriorated. Nuclei cannot be identified in the image center. Moreover, nuclei variability is important meaning that the state space size $\chi$ is large. Some nuclei are in mitosis (see e.g. top-left). In spite of these difficulties, the MBC algorithm provides acceptable results. They would allow to make statistics on the cell location and orientation, which is a major problem in biology. The computing times for this example is 30 minutes.

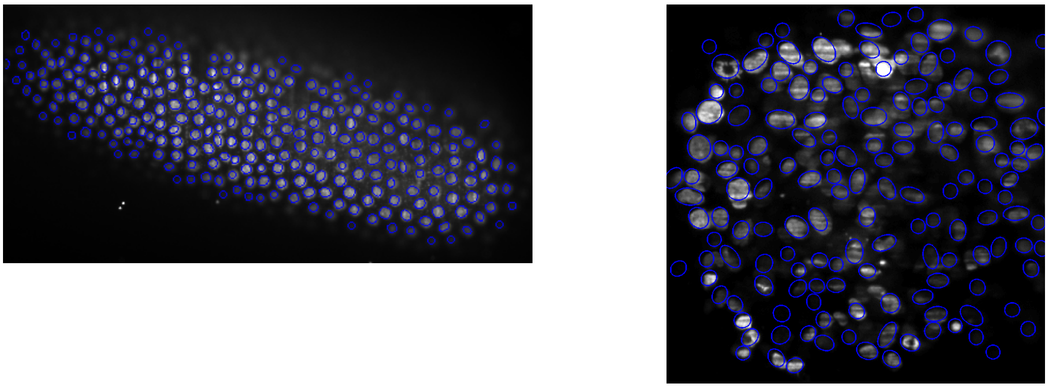


*Figure 8. 2D segmentations of a nuclei of Drosophila embryo (left) and a multicellular tumor spheroid (right).*

### 5.2.4. Spermatozoid tracking

**Participants:** Clarens Caraccio, Xavier Descombes.

In this work, we have proposed an algorithm for tracking spermatozoid in a sequence of confocal images. We first detect the spermatozoids by thresholding the result of a top hat operator. The thresold is automatically estimated using Otsu's method. We then analyse the different connected components to detect overlaps between adjacent spermatozoids. Temporal neighbors are selected based on the spatial consistency of the object sets between two consecutive time. A first result is given on figure 9 .

## 5.3. Classification

### 5.3.1. Axon morphology comparison using elastic shape analysis

**Participants:** Alejandro Mottini, Xavier Descombes, Florence Besse.

It is known that neuronal morphology impacts network connectivity, thus providing information on its functioning. Moreover, it allows the characterization of pathological states. Therefore, the analysis of the morphological differences between normal and pathological structures is of paramount importance.

We present a new method for comparing reconstructions of axonal trees (obtained, for example, by applying our segmentation method on confocal microscopy images of normal and mutant axonal trees) which takes into account both topological and geometrical information and is based on the Elastic Shape Analysis Framework. The method computes the geodesic between two axons in a space of tree like shapes, and the distance between the two is defined as the length of the geodesic. Moreover, our method is capable of showing how one axon transforms into the other by taking intermediate points in the geodesic.

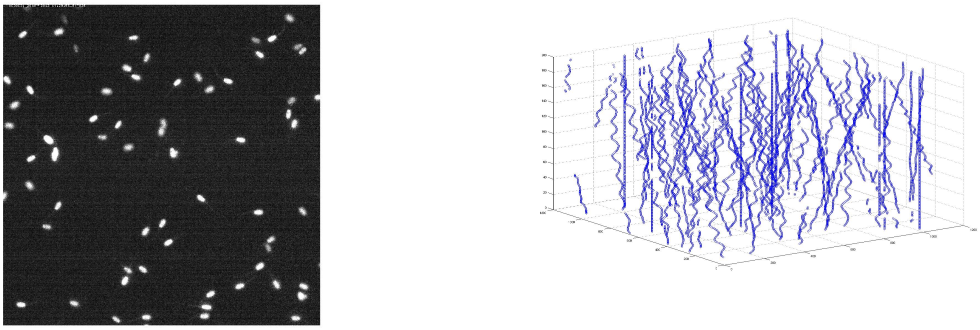


Figure 9. Original confocal image and estimated spermatozoid trajectories.

We consider two axonal trees $T_1$ and $T_2$, each consisting of an axon and several branches (and possibly sub branches). All are represented by 3D open curves in $\mathbb{R}^3$ (see Figure 10 ). We start by defining the matching function $M$ such that $M:(0,1,2,...n)\times(0,1,....,m)$, where $n$ and $m$ are the number of branches in $T_1$ and $T_2$ respectively. The matching function matches the branches of the two trees, for example, by assigning branch $i=1$ of $T_1$ to branch $j=3$ of $T_2$. We then define a branch function $C$ which indicates, for a given time $t_c$, how many branches remain after $\beta(t_c)$ (see Figure 10 ). We only take into account branches which have a match in the other axonal tree. Finally, we define the distance between two axonal trees $T_1, T_2$ as:

$$D(T_1,T_2)=\min_M d((\beta_1(t),C1(t,M)),(\beta_2(t),C2(t,M)))+\sum_{(i,j)}\alpha_{i,j}M(i,j)D(T_1(i),T_2(j)) \qquad (1)$$

where $\beta_k$ is the main curve (axon) of tree $k$, $C_k$ its branch function, $M$ the matching function, $\alpha_{i,j}$ a weight parameter and $D(T_1(i),T_2(j))$ the distance between the matched branches of the two trees. All distances between simple curves are calculated using the elastic shape analysis framework.

The method performance was tested on a group of 22 (11 normal and 11 mutant) 3D images, each containing one axonal tree manually segmented by an experienced biologist from a set of real confocal microscopy images. The mean and standard deviation of the inter and intra class distances between the neurons were calculated and results suggest that the proposed method is able to distinguish between the two populations (an average interpopulation to intrapopulation distance ratio of 1:21 and 1:28 were obtained). In addition, we computed the optimum transformations between axons. An example is shown in figure 11 . This result was obtained by taking intermediate points along the geodesic between the two trees.

Figure 10. Axonal tree diagrams (a) and their corresponding $C$ functions for a given $M$ (b).

### 5.3.2. Vascular network segmentation from X-ray tomographic volumes

**Participant:** Xavier Descombes.

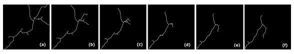


*Figure 11. Optimum transformation between two axonal trees (transformation starts in (a) and finishes in (f), maximum intensity projections).*

*This work was made in collaboration with Franck Plouraboué and Abdelakim El Boustani from IMFT, Caroline Fonta from CerCo, Géraldine LeDuc from ESRF, Raphael Serduc from INSERM and Tim Weitkamp from Synchrotron Soleil.*

Micro-tomography produces high resolution images of biological structures such as vascular networks. We have defined a new approach for segmenting vascular network into pathological and normal regions from considering their micro-vessel 3D structure only. We consider a partition of the volume obtained by a watershed algorithm based on the distance from the nearest vessel. Each territory, defined as Local Vascular Territory (*a Local Vascular Territory (LVT) is a connected region corresponding to the catchment bassin associated with a vascular element. It can be obtained through the watershed computation on the opposite distance map from the vessels and is not connected to the sample border.* ), is characterized by its volume and the local vascular density. The volume and density maps are first regularized by minimizing the total variation, within a Markov Random Field framework, using a graph cut algorithm . Then, a new approach is proposed to segment the volume from the two previous restored images using an iterative algorithm based on hypothesis testing. We consider the variables density and volume for each LVT and the populations constituted by the different classes obtained by the segmentation at a given step. Classes which are not statistically significantly different are merged using a MANOVA. This blind segmentation provides different regions which have been interprated by expert as tumor, necrosis, tumor periphery and sane tissue 12 .

### 5.3.3. Statistical analysis of skin pigmentation under treatment

**Participants:** Sylvain Prigent, Xavier Descombes.

*This work was partially funded by a contract with Galderma R&D [http://www.galderma.com/RampD.aspx]. It was made in collaboration with J. Zerubia from Ayin team.*

multispectral imaging, skin, hyperpigmentation, hypothesis tests, statistical inferences

One of the steps to evaluate the efficacy of a therapeutic solution is to test it on a clinical trial involving several populations of patients. Each population receives a studied treatment and a reference treatment for the disease. For facial hyper-pigmentation, a group of $N_e$ patients receives the treatment on one cheek and a comparator on the other. The comparator can be a reference treatment or a placebo. To this end patients are selected to have the same hyper-pigmentation severity on the two cheeks. Then multi-spectral images are taken at different time $t$ along the treatment period.

We propose a methodology to assess the efficacy a treatment by calculating three differential criteria: the darkness, the area and the homogeneity. The darkness measure the average intensity of the disease on a gray scaled image $I$ obtained by a linear combination of the spectral bands of the original multi-spectral image. A differential darkness is then obtained by measuring the deviation between the initial measurement at time $t_0$, and the measurement at time $t_k$. The differential area criterion is calculated by analyzing the histogram of $I_{diff} = I_{t_0} - I_{t_k}$ a difference gray scale image between two measurements in a time series. The differential homogeneity criterion is obtaining with a multi-scale analysis of $I_{diff}$ adapted from the Statistical Parametric Mapping (SPM) methodology. Indeed, statistical inferences allow to assign a probability of change to each region of $I_{diff}$ above a set of thresholds. These probabilities are calculated with respect to the maximum intensity and the spatial extend of each region. An integration of the obtained statistical map denoted $SM$, allows to get a homogeneity criterion.

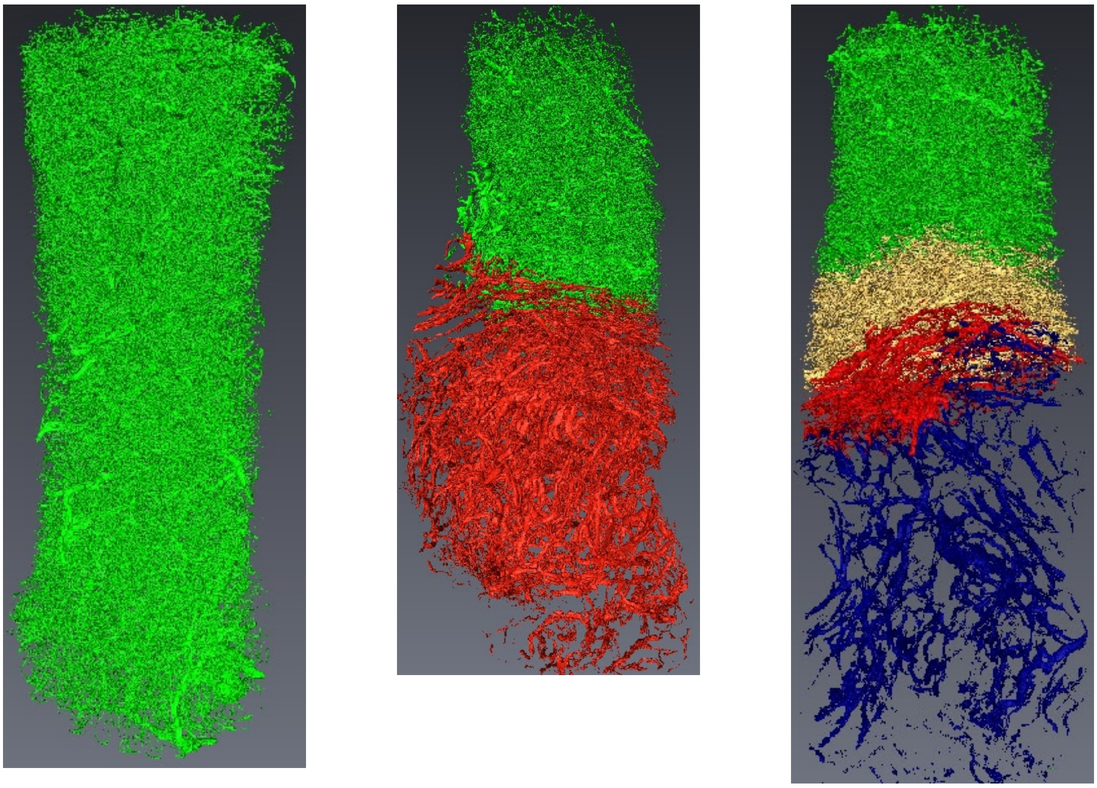
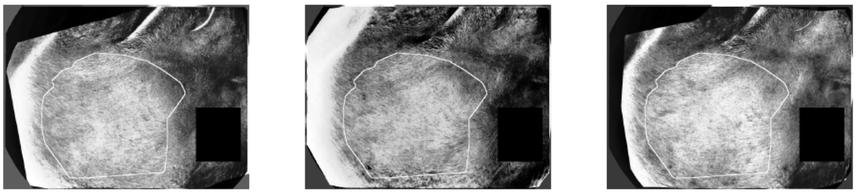
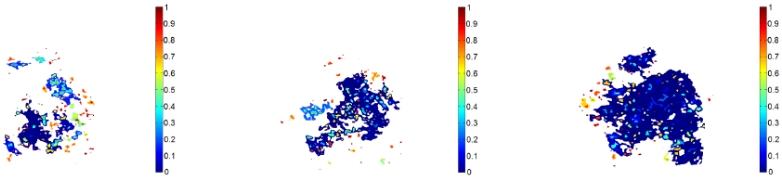


*Figure 12. Examples of segmentation: tumor (red), necrosis (blue), tumor periphery (yellow) and sane (green)*

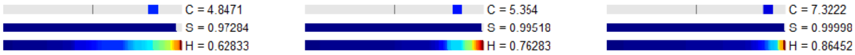
The figure 13 illustrates the differential score calculated on a patient whose pathology decreases during the clinical trial. The proposed differential score have been tested in a full clinical study and provided results that agreed with the clinical analysis. This work have been patented and published in Inria research reports [25], [26].



(a) $I^{t1}_{diff}$ equalised (b) $I^{t2}_{diff}$ equalised (c) $I^{t3}_{diff}$ equalised



(d) $SM_{t_1}$ (e) $SM_{t_2}$ (f) $SM_{t_3}$



(g) score $t_1$ (h) score $t_2$ (i) score $t_3$

*Figure 13. $I_{diff}$, $SM$ and differential score for the three measurements $t_1$, $t_2$, $t_3$ calculated for a patient whose disease decreases.*

### 5.3.4. A Recursive Approach For Multiclass Support Vector Machine: Application to automatic classification of endomicroscopic videos

**Participants:** Alexis Zubiolo, Eric Debreuve.

*This work is made in collaboration with Barbara André (Mauna Kea Technologies)*

The problem of automatic image (or video, or object) classification is to find a function that maps an image to a class or category among a number of predefined classes. An image can be viewed as a vector of high-dimension. In practice, it is preferable to deal with a synthetic signature of lower dimension. Therefore, the two classical steps of image classification are: image signature extraction and signature-based image classification. The classification rule can be learned from a set of training sample images manually classified by experts. This is known as supervised statistical learning where *statistical* refers to the use of samples and *supervised* refers to the sample classes being provided. We are interested in the learning aspect of the multiclass [1] problem when using a binary classification approach as a building block. We chose the Support Vector Machine (SVM), a well-known binary classifier.

Among the proposed extensions of binary classification methods to multiclass (three classes or more), the one-versus-one and one-versus-all approaches are the most popular ones. Let us suppose that there are $p \geq 3$ classes. The idea of the one-versus-all strategy is to oppose to any of the classes the union of the remaining $p-1$ classes. Then, $p$ SVM classifiers are determined, each one scoring, say, positively for one of the classes.

---

[1] Traditionally in classification, *multiclass* means “three classes or more” while the two-class case is referred to as binary classification.

The one-versus-one strategy opposes the classes by pair for all possible pairs. Therefore, $\frac{p(p-1)}{2}$ SVMs are determined and classification is performed by a majority vote.

As an alternative to these aforementioned strategies (as well as to other, less popular ones), we developed a recursive learning strategy. A tree of SVMs is built, achieving three goals: a fair balance in the number of samples used in each binary SVM learnings, a logarithmic complexity for classification ($\log_2(p)$ compared to the linear or quadratic complexities of one-versus-all or one-versus-one, respectively), and a coherent, incremental classification procedure (as opposed to selecting the final class based on possibly competing partial decisions). During learning, at each node of the tree, a combinatorial search is performed to determine an optimal separation of the current classification problem into two sub-problems. The proposed method was applied to automatic classification of endomicroscopic videos.

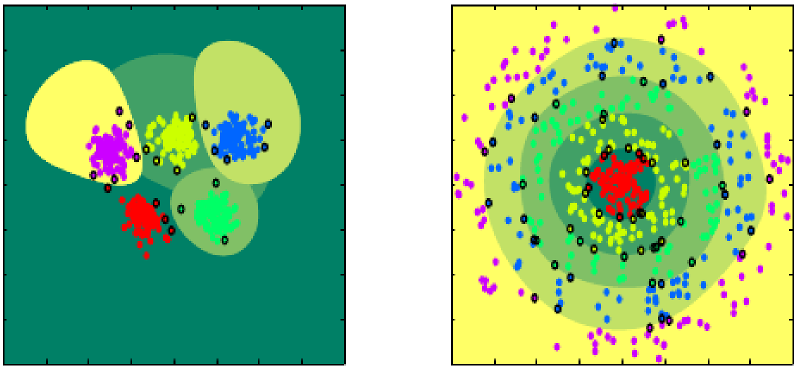


*Figure 14. Illustration of the proposed recursive approach for multiclass Support Vector Machine. Colored dots: learning feature samples; Encircled dots: computed support vectors; Colored areas: computed class regions. Left: classical example; Right: concentric example.*

## 5.4. Modeling

### *5.4.1. Tracking Growing Axons in Fluorescent Microscopy Images*

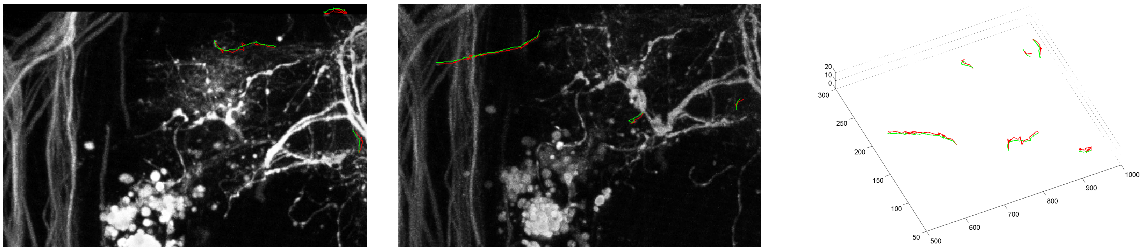
**Participants:** Huei Fang Yang, Florence Besse, Xavier Descombes.

*This work has been done in collaboration with Caroline Medioni from iBV.*

Analyzing how growing axons correctly reach their target neurons is essential for biologists to better understand the development of a nervous system. Analysis of the properties of axon growth requires detecting axonal tips and tracking their trajectories within complex and large data sets. When performed manually, the tracking task is arduous and time-consuming. To this end, we proposed a tracking method, based on the particle filtering technique, to follow the traces of axonal tips that appear as small bright spots in the 3D $+ t$ fluorescent two-photon microscopy images exhibiting low signal-to-noise ratios (SNR) and complex background. Our tracking method uses multiple dynamic models in the proposal distribution to predict the positions of the growing axons. Moreover, it incorporates object appearance, motion characteristics of the growing axons, and filament information in the computation of the observation model. The integration of these three sources results in improved accuracy of recovered trajectories. The experimental results obtained from the microscopy images, presented in Figure 15 , showed that the proposed method can successfully estimate trajectories of growing axons, demonstrating its effectiveness even under the presence of noise and complex background.

### *5.4.2. Trajectory Simulation of Growing Axons:*

**Participants:** Huei Fang Yang, Florence Besse, Xavier Descombes.



(a) (b) (c)

*Figure 15. Visual comparison between the tracking results of the proposed method and the manually created ground truth in 2D and 3D. The red trajectories are produced by the proposed method, and the green are the ground truth manually created by the expert; both are overlaid on the MIPs (a) and (b) and visualized in 3D (c). The computer generated tracks are consistent with the ground truth in general, with minor differences between the estimated positions and the ground truth positions. The differences are caused by the noise and by the effect of complex background.*

*This work has been done in collaboration with Caroline Medioni from iBV.*

It is established in biology that axons reach their target cells in the developing nervous system by the guidance of molecular gradients. To better understand how growing axons react to the molecular cues, either attractant or repellent, we simulated the trajectories of growing axons using a mathematical model that investigates the effect of molecular gradients on the axon's growth angle. Figure 16 shows the simulated trajectories of 50 growing axons. The initial position of axons is $(0, 0)$, and the red point on the right denotes the target cell that secretes the attractant cue.

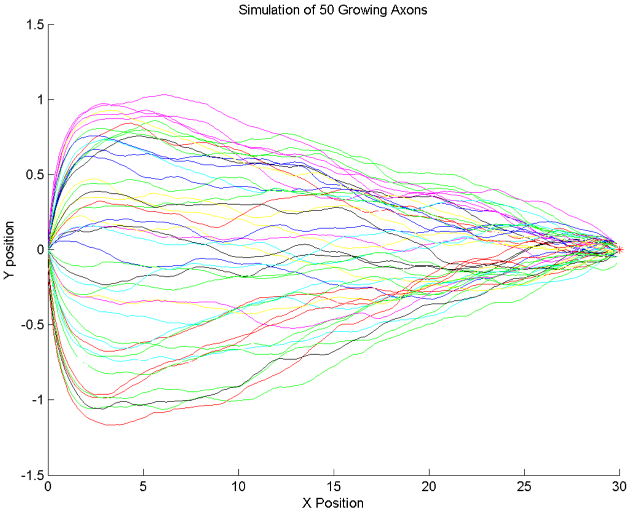


*Figure 16. Simulated trajectories of 50 growing axons. The initial position of 50 axons is set to $(0, 0)$, and the red point on the right represents the target cell that secretes the attractant cue.*

# SERPICO Team

# 6. New Results

## 6.1. Robust parametric stabilization of moving cells

**Participants:** Solène Ozeré, Patrick Bouthemy, Charles Kervrann.

**Paper under review.**

Analysing the dynamic behaviour of individual particles (e.g., proteins, vesicles) inside a cell is of primary importance in cell biology. However, the motion of these particles observed in live cell microscopy image sequences is the addition of the global movement of the cell and their own single motions. Hence, automatically stabilizing the cell (or a group of cells), i.e. compensating for its global motion or equivalently registering its successive positions over time, is previously required. We have proposed a cell stabilization method based on a realtime robust multiresolution scheme (Motion2D software [36]). It can simultaneously handle the estimation of 2D parametric global motions (e.g., affine motion model) and of temporal intensity variations. Three temporal intensity models have been investigated: constant additive model, exponential decay model (corresponding to the photobleaching effect), continuity equation. We have carried out experiments on three biological situations: development of cells, displacements of endosomes, protein recruitment by the Golgi. We have demonstrated the accuracy of our method on these challenging examples and its capacity to efficiently reveal the own motion of subcellular particles. It yields better results than the StackReg method (http://bigwww.epfl.ch/thevenaz/stackreg/), classically used in the field, in cases involving strong local dynamics (see Fig. 6 ).

**Partners:**: Perrine Paul-Gilloteaux (UMR 144 CNRS PICT IBiSA Institut Curie)

## 6.2. Motion classification for interpreting subcellular dynamics

**Participants:** Antoine Basset, Patrick Bouthemy, Charles Kervrann.

We have just started to address the classification of motions of subcellular particles in light microscopy time-lapse image sequences. For the while, we are considering the following three general classes: diffused motion, obstructed motion and directed motion. We are investigating three approaches. First, we can design likelihood ratio tests for deciding the relevant configuration on local patches. Second, we can define a short-term classification framework based on optical flow computed at time $t$. The third approach is a mid-term one exploiting pieces of trajectories (tracklets) computed by tracking a set of points.

**Partners:**: Jérôme Boulanger (UMR 144 CNRS Institut Curie)

## 6.3. Aggregation of patch-based estimations for illumination-invariant optical flow in live cell imaging

**Participants:** Denis Fortun, Charles Kervrann, Patrick Bouthemy.

**Paper under review.**

Live cell image sequences provide a large variety of challenging situations for motion estimation. We have developed a novel optical flow estimation method in the line of work of [11], based on a two-stage aggregation framework and designed to handle this diversity of issues. First, semi-local candidates are estimated with a combination of patch correspondences and illumination-invariant affine motion estimations. Then, one candidate is selected at each pixel in a graph-cut based global aggregation stage. This approach allows us to overcome usual limitations of existing methods such as loss of small structures with large displacements, dependency on illumination changes and over-smoothing of the motion discontinuities. The method jointly estimates the motion and illumination fields. We have compared our approach to state-of-the-art methods and have demonstrated its ability to outperform existing methods in challenging cases frequently arising in live cell imaging (see Fig. 7 ).

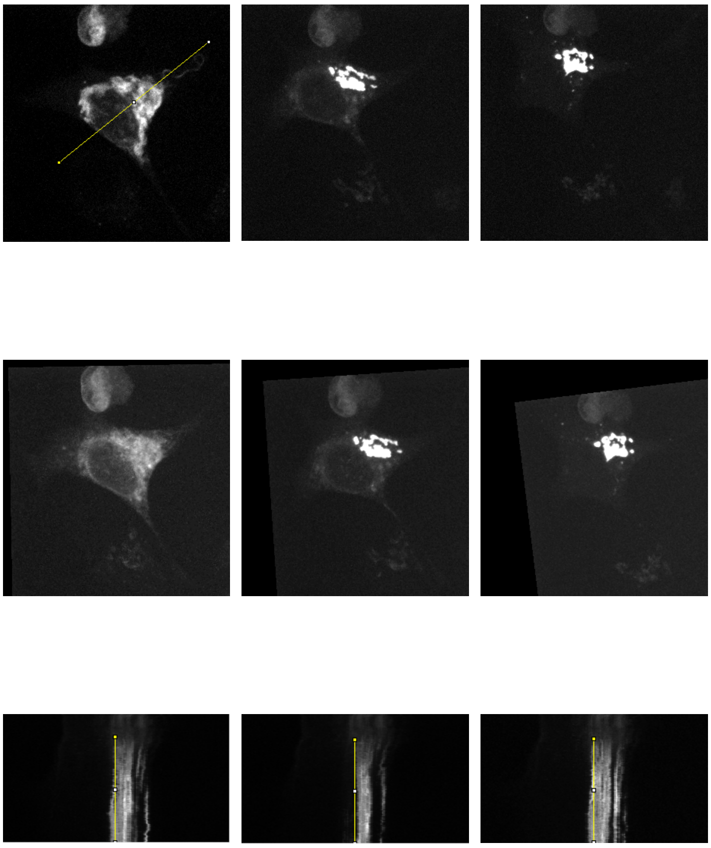
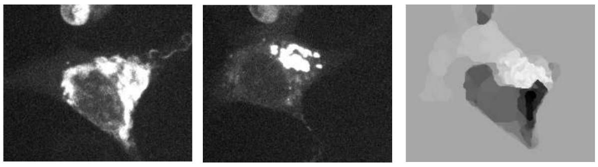
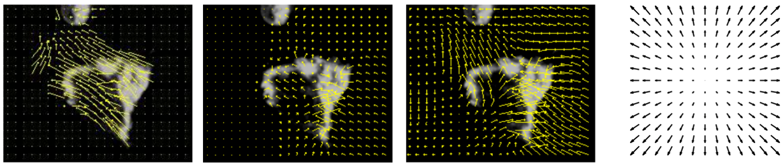


*Figure 6. Real-time imaging of the synchronized trafficking of ManII-SBP-EGFP [21]. HeLa cells were transfected to express Ii-streptavidin as a hook and ManII-SBP-EGFP as a reporter. Release of the reporter was induced by addition of biotin and monitored using a spinning disk confocal microscope (F. Pérez, UMR 144 CNRS Institut Curie): 1st row: original images. 2nd row: images out of the motion-compensated sequence at time* $t = 20, 60$ *and 140 computed with our method; 3rd row: kymographs of the backwarped sequence computed respectively with the baseline motion equation (left), with the exponential decay (middle) and with* STACKREG *software (right).*



frame 1 frame 2 illumination change map

color code map



our method method [22] method [42] vector code map

*Figure 7. Comparison of our method with the methods of [22] and [42] on a sequence of “HeLa cells” (courtesy of F. Pérez UMR 144 CNRS Institut Curie, PICT-IBiSA).*

**Partners:**: Perrine Paul-Gilloteaux (UMR 144 CNRS PICT IBiSA Institut Curie)

## 6.4. Correlation and variational approaches for motion and diffusion estimation in fluorescence imaging

**Participants:** Denis Fortun, Charles Kervrann.

**Paper under review.**

In this work, we have compared a correlation-based approach and a variational method for both motion and diffusion estimation in representative cell biology studies in fluorescence imaging. The so-called Spatio-Temporal Image Correlation Spectroscopy (STICS) is widely used in fluorescence imaging to recover physical parameters (e.g. direction of flow or Brownian motion of molecules). We have investigated recent advances in variational dense motion estimation and we have proposed to adapt the variational framework to the estimation of diffusion (i.e. Brownian motion). We have demonstrated the influence of the regularization parameter in the variational approach and its ability to capture motion of individual or clusters of moving objects. We have evaluated the advantages and limits of the two approaches for different biological studies (see Fig. 8 ).

**Partners:**: Perrine Paul-Gilloteaux, Francois Waharte and Chen Chen (UMR 144 CNRS PICT IBiSA Institut Curie)

## 6.5. Noise modeling and denoising for intensified camera in fluorescence imaging

**Participants:** Philippe Roudot, Charles Kervrann.

**Two papers under review.**

Image intensifiers are commonly used in low light level biological imaging, especially for fluorescence imaging. In this study, we have proposed a statistical framework for noise variance estimation dedicated to image sequences acquired with ICCD (Intensifier CCD). The model has been exploited for fluorescence lifetime estimation (Fluorescence lifetime imaging microscopy, FLIM) [13], [12] and image denoising. We have investigated an alternative approach to [41] and we have shown that intensifier gain variation cannot be neglected in the variance estimation as opposed to a CCD sensor gain. Additionally, we have suggested to correct the noise model spatially to cope with microscopical aberration which are common in experimental setups (see Fig. 9 ). Finally, we have proposed a novel denoising algorithm based on the NL-means filter [23] which does rely on variance stabilization. The novel patch-based filter is able to adapt to local intensity-based noise statistics (see Fig. 10 ).

**Partners:** F. Waharte and J. Boulanger (UMR 144 CNRS PICT IBiSA Institut Curie)

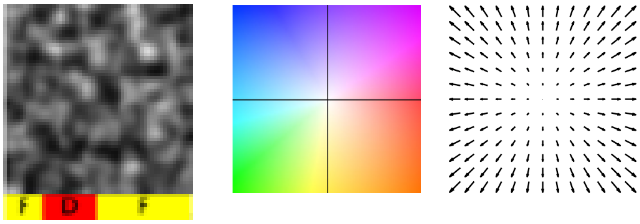
## 6.6. Microtubules modeling for variational assimilation analysis

**Participants:** Pierre Allain, Charles Kervrann.

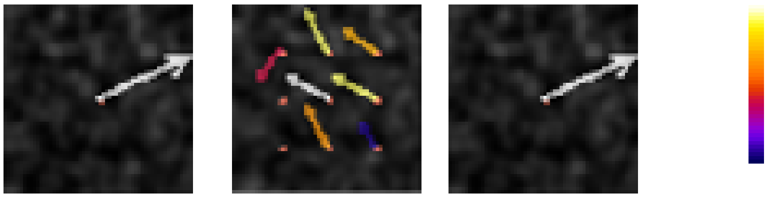
In this project, we propose a bio-physical modeling of growing microtubules at the scale of a single cell. The theoretical advantage of such a modeling is to step aside empirically-based heuristics often carrying artificial parameters which can be hard to tune and to make sense in a data analysis context. We thus propose to model microtubules as rigid and growing cylinders alike (Nedelec and Foethk 2007) [34] but including Newtonian dynamics.

Using both this modeling and fluorescence microscopy, we aim at controlling simulated microtubules to satisfy in vitro observations. We plan to use variational assimilation with adjoint method in the future to achieve such an estimation. We believe that this approach should be able to provide information both on microtubules properties and on vesicle transport dynamics.

motion scenario color code map vector code map

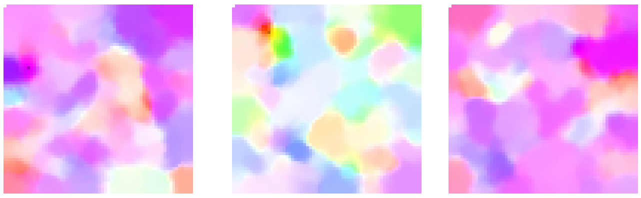


STICS-based estimation of different flux and diffusion phases



Flux 1 Diffusion Flux 2

Variational method with two different regularization parameters



Flux 1 Diffusion Flux 2

Flux 1 Diffusion Flux 2

*Figure 8. Analysis of STICS and variational methods on artificial image time series with three phases. First row: first frame of the sequence and temporal description of the 3 phases: F/Flux (i.e. directed flow), D/Diffusion (20 - 30 - 40 images) (left); coding maps of vector fields (middle and right). Second row: STICS analysis for each phase. The arrows show the direction of the displacement and the color code is used to represent orientation and magnitude of estimated velocities. Third and fouth rows: Variational estimation for image pairs of each phase with a low regularization parameter (third row) and a high regularization parameter (fourth row).*

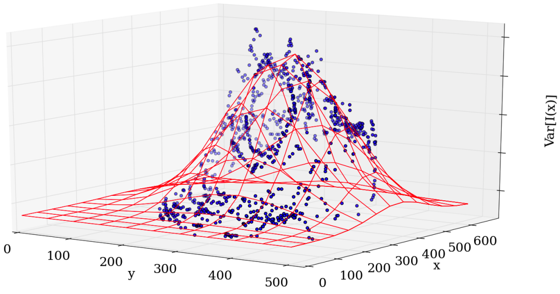


*Figure 9. Variance prediction after noise model calibration using a fluorescein FD-FLIM reference stack acquired with a wide-field (WF) microscope (parametric model in red and measurements in blue).*

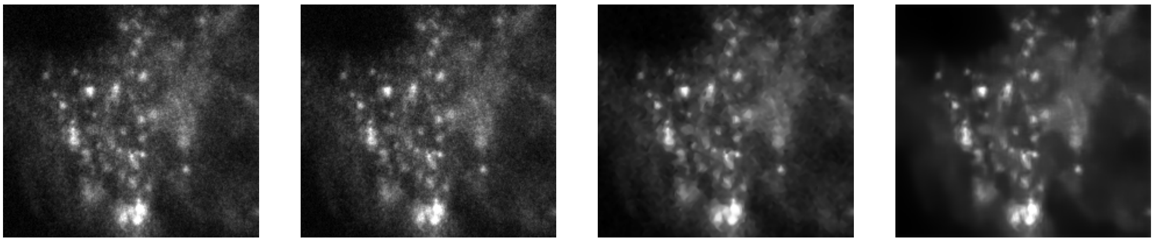


*Figure 10. Denoising performance on a live cell image acquired in FD-FLIM (fluorescent tagged caveolin protein) using a confocal setup. From left to right: original image, results with* BM3D *[[27]]* ND-SAFIR *[[4]] and our method.*

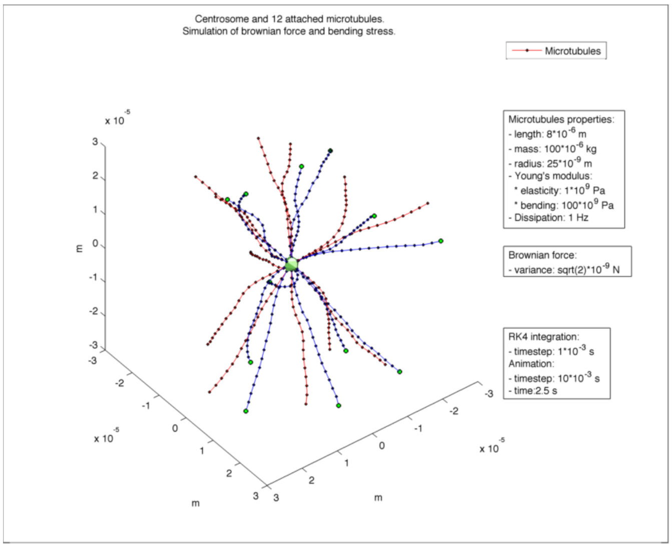


*Figure 11. Simulation of growing microtubules in 3D.*

## 6.7. Single versus dual-axis cryo-electron tomography for reconstruction of microtubules assembled in vitro

**Participant:** Charles Kervrann.

Single-axis cryo-electron tomography of vitrified specimens has become a method of choice to reconstruct in three dimensions macromolecular assemblies in their cellular context or prepared from purified components. In [9], we described a dual-axis acquisition scheme able to improve three-dimensional reconstructions of microtubules assembled in vitro. We showed that in single-axis tomograms, microtubules oriented close to the perpendicular of the tilt axis display diminished contrast, and ultimately transform into sets of parallel lines oriented in the direction of the electron beam when observed in cross-section. We analyzed projections in three-dimensional Fourier transform to demonstrate that imaging artifact is due to a decrease in the angular sampling of their equatorial components. Although the second orthogonal series of images does not fully complement the first one at the specimen level due to increased radiation damage, it still allows elongated features oriented in any directions to be correctly reconstructed, which might be essential for highly heterogeneous specimens such as cells.

**Partners:** Denis Chrétien, Audrey Guesdon and Sophie Blestel (UMR 6290 CNRS University of Rennes 1)

## 6.8. Analysis of lateral organization of ordered domains at the plasma membrane surface

**Participant:** Charles Kervrann.

**Paper under review.**

In this study, we have analysed a recently designed probe, di-4-ANEPPDHQ, that can change its fluorescent properties depending on whether it is residing in two distinct phases (ordered phase vs disordered phase) of the tobacco cell plasma membranes. We performed a spatial analysis of small (<200 nm) ordered domains observed in multispectral confocal microscopy. We focused on relevant binary images, assumed to be realizations of a MRF-Ising model, depicting the spatial organization of ordered domains. The Ising model depends on 2 parameters: the external field parameter $h$ which controls the total fraction of the "ordered" phase and the interaction parameter which controls the spatial coupling. Maximum pseudo-likelihood methods were investigated to estimate parameters able to describe the spatial properties of ordered domains at the scale of 200 nm $\times$ 200 nm. Almost all estimates of the coupling parameter were positive excluding complete spatial randomness of ordered domains and showing a tendency to spatial aggregation at small distance. We then measured the strength of spatial aggregation through the calculation of the variability fraction explained by the spatial coupling. The mean fraction is low (0.5%) suggesting positive and limited interacting forces between neighbor ordered pixels. Altogether our simulations and analyses provided a probabilistic spatial characterization of PM ordered domains, indicating that recorded images showed a two-scale organization with spatial randomness at large scales (several micrometers) associated with spatial aggregation due to short-range interactions (up to 400 nm).

**Partners:** P. Gerbeau-Pissot, F. Simon-Plas (UMR 1088 PME INRA, Dijon) and K. Kiêu (MIA Unit INRA, Jouy-en-Josas)

## 6.9. Line detection in microarray scanner images

**Participants:** Alice Bergonzoni, Charles Kervrann.

In this study, we have studied two approaches to automatically detect straight lines in images (tool-slide) for calibrating scanners designed by Innopsys company. The Hough transform has been investigated and is able to produce satisfying results provided the algorithm parameters are carefully adjusted (see Fig. 12 ). To overcome this difficulty, we have evaluated the potential of *a contrario* approach [28] which is well ground theoretically and requires no object prior and parameter adjustment. According to the Helmholtz principle which is based on the *a contrario* approach, any structure is considered in a white noise image as a deviation from randomness.

A meaningful segment is detected when the expectation of its number of occurrences in a white noise image (i.e. number of false alarms) is low. We have evaluated the potential of this method and performed experiments using the LSD algorithm [46] inspired from [28].

**Partners:**: V. Paveau (Innopsys)

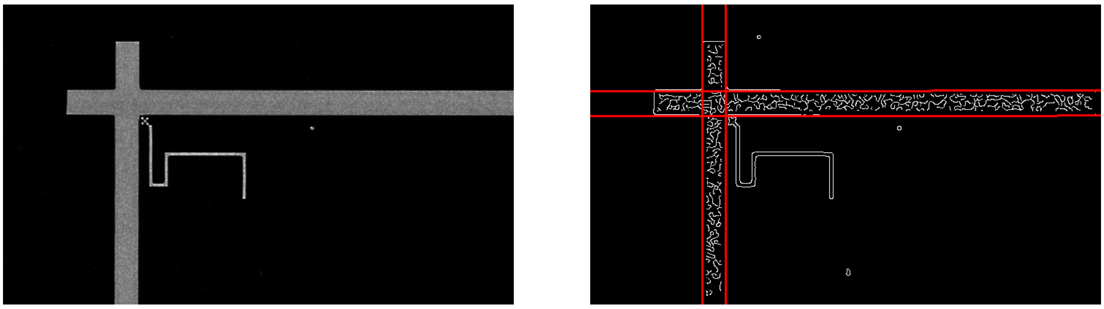


*Figure 12. Detection of lines in an image (tool-slide) (pixel size: $3\mu m \times 3\mu m$).*

## ASCLEPIOS Project-Team

# 5. New Results

## 5.1. Medical Image Analysis

### 5.1.1. Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model

**Participants:** Ezequiel Geremia [Correspondant, Inria], Nicholas Ayache [Inria], Antonio Criminisi [MSRC], Bjoern Menze [Inria,ETHZ], Marcel Prastawa [University of Utah].

*Published in the proceedings of the MCV Workshop at MICCAI 2012 [35]*

biophysiological tumor growth simulator, multi-variate regression random forests, gliomas, MRI

- A generative-discriminative framework is presented to learn model-based estimations of the tumor cell density
- The ground truth for tumor cell density is very hard to obtain
- A biophysiological tumor growth simulator is used to generate the ground truth tumor cell densities and associated MRIs
- A multi-variate regression random forests is trained to estimate the voxel-wise distribution of tumor cell density from input MR images
- The training data contains 500 synthetic cases and their associated ground truth generated by the brain tumor simulator
- The method was tested on 200 synthetic cases with excellent results
- The method also provided very promising results for estimating the tumor cell density on 16 clinical cases showing low grade gliomas from the DKFZ (German Cancer Research Center)

### 5.1.2. Automatic indexation of cardiac MR images

**Participants:** Jan Margeta [Correspondant], Nicholas Ayache, Antonio Criminisi [MSRC].

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

Machine learning, Cardiac MR, MR preprocessing

- A generic random forest framework has been implemented and its recent modifications have been applied to a fully automatic and a semisupervised image segmentation methods, and manifold learning in cardiac MRI.
- We have performed image based cardiac function quantification from preprocessed cardiac cine MRI sequences.
- An image intensity standardization in magnetic resonance images method has been proposed.

### 5.1.3. Multimodal brain tumor segmentation

**Participant:** Bjoern Menze [Correspondant].

MICCAI 2012

- Further developed the generative brain tumor segmentation model
- Developed a generative-discriminative model for multimodal brain tumor segmentation
- Developed a new regularization approach for longitudinal tumor segmentation (with Guillaume Charpiat & Yuliya Tarabalka, Inria Sophia-Antipolis)
- Initiated and co-organized an international benchmark on multimodal brain tumor segmentation as a challenge workshop during MICCAI 2012 in Nice (http://www2.imm.dtu.dk/projects/BRATS2012)

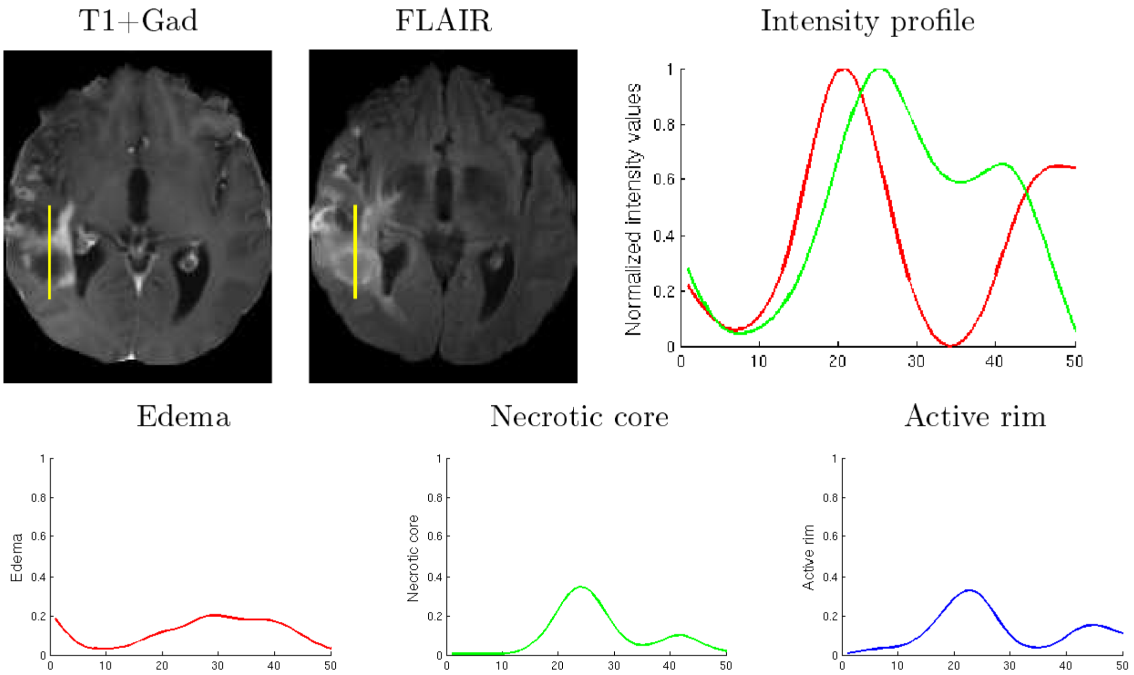


*Figure 1.* ***Prediction of the cell densities along a section of the tumor.*** *Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively.*

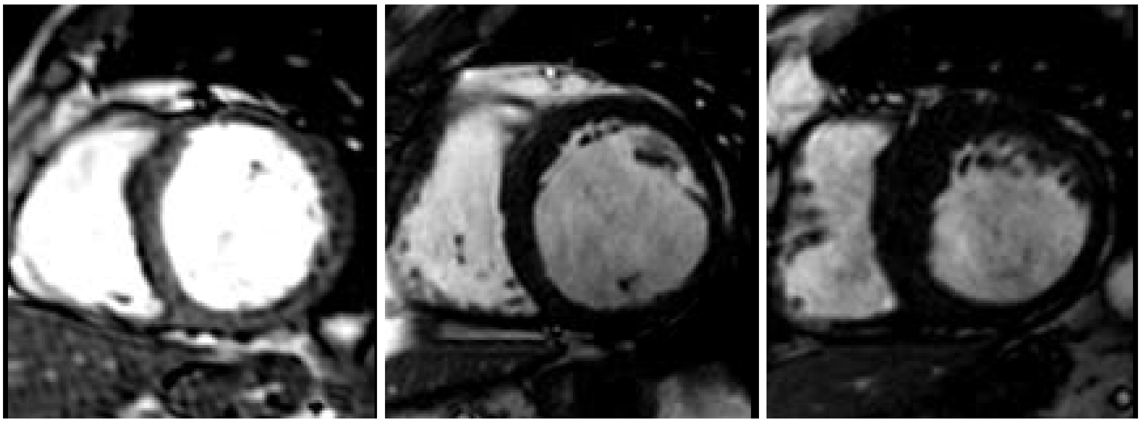


*Figure 2. Three different patients with increasing (from left to right) functional deficiency of the heart.*

### *5.1.4. Statistical Analysis of Diffusion Tensor Images of Brain*

**Participants:** Vikash Gupta [Correspondant], Xavier Pennec, Nicholas Ayache.

Diffusion Tensor Imaging of Brain, Tractography, Super-resolution, Statistical analysis

Diffusion tensor imaging (DTI) is gaining interest as a clinical tool for studying a number of brain diseases pertaining to white matter tracts and also as an aid in neuro-surgical planning. Unfortunately, in a clinical environment, the diffusion imaging is hampered by the long acquisiton times, low signal to noise ratio and a prominent partial volume effect due to thick slices. The present work aims at robustifying the analysis of clinical images by developing a super-resolution algorithm for DTI and quantifying its improvements with respect to the existing tensor estimation methods. Part of the work was presented at the 1st International Symposium on Deep Brain Connectomics [68].

### *5.1.5. 3D/2D coronary arteries registration*

**Participants:** Thomas Benseghir [Correspondant], Grégoire Malandain, Régis Vaillant [GE-Healthcare], Nicholas Ayache.

*This work is done in collaboration with GE-Healthcare (Buc).*

3D/2D registration ; coronary arteries ; Chronic Total Occlusion ; X-ray fluoroscopy / CT image fusion

The context of this work is to provide the cardiologist with an advanced guidance application, where a pre-operative 3D CT segmented image will be superimposed on the per-operative 2D live fluoroscopy. Since the relative positions of the 3D image and the 2D projective images are unknown, we are currently investigating robust pose estimation methods before using an upcoming registration algorithm.

## 5.2. Biological Image Analysis

### *5.2.1. Pre-clinical molecular imaging: breath-hold reconstruction in micro-SPECT and segmentation of IHC stomach slices*

**Participants:** Marine Breuilly [Correspondant], Grégoire Malandain, Nicholas Ayache, Jacques Darcourt [CAL], Philippe Franken [CAL], Thierry Pourcher [CEA].

*This work is jointly conducted with the Transporter in Imagery and Oncologic Radiotherapy team (TIRO, CEA-CAL-UNSA) located in Nice.*

SPECT/CT, small animal, respiratory motion, respiratory gating, 4D images, stomach, segmentation, immuno-histochemistry

Using the coupled CT and SPECT device, both the anatomy (with the CT) and physiology information targeted by a dedicated radio-pharmaceutical tracer (here the tumors, with the SPECT) can be imaged. However, tumor quantification is impaired by the respiratory motion that induces an artifical enlargement of the moving structures. Thus, the characterization of respiratory motion in dynamic images was studied.

- An ad hoc method for motion detection in dynamic image was developped and tested on two different modalities (4D-SPECT and 4D-CT).
- Image-based motion detection results were compared to the pressure signal and to lung volume variation. A temporal shift between the peak of motion in images and the ones in the pressure signal was observed (see Figure 3 ).
- The temporal shift suggested to carefully select data from the non moving phase for a motionless 3D-SPECT image reconstruction. This step was incorporated in a breath-hold like reconstruction method [65], [67], [66].

## 5.3. Computational Anatomy

### *5.3.1. Statistical Analysis of Transformations on Lie groups and longitudinal studies*

**Participants:** Xavier Pennec [Correspondant], Marco Lorenzi, Nicolas Duchateau [Hospital Clinic, Univ. Barcelona].

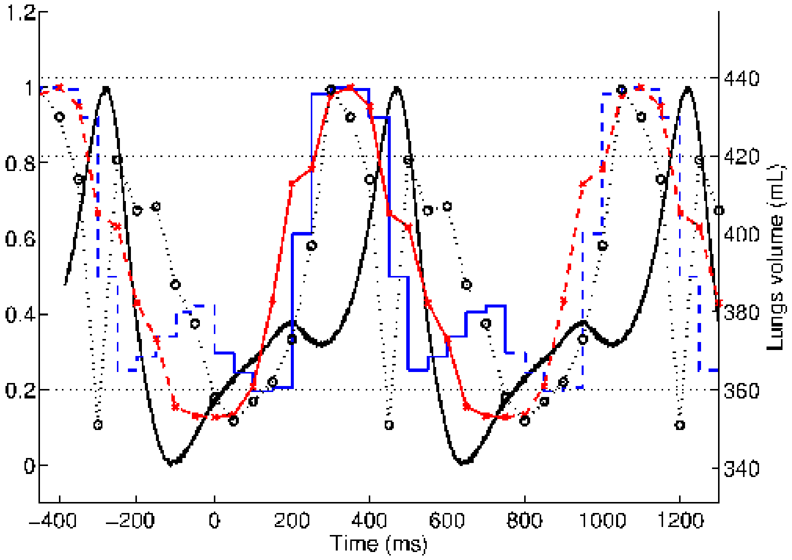


*Figure 3.* ***Comparison between pressure signal and image-based measures computed on 4D-SPECT and 4D-CT.*** *Normalized average pressure signal (continuous black curve); normalised histogram* $h_{4D-SPECT}(n)$ *computed on the 4D-SPECT image (blue staircase); normalised histogram* $h_{4D-CT}(n)$ *of motion phase computed on the 4D-CT image (red curve with cross-shaped markers); and lung volume (mL) measured on the 4D-CT image (black dotted curve with circle-shaped markers).*

Lie groups, transformations, mean value, non-linear registration

In order to perform statistics on transformations for computational anatomy purposes, we investigate alternative theoretical structures to the right- (or left-) invariant Riemannian setting usually used.

- In order to define a notion of a mean which is consistent with Lie group operations we propose in [57] to replace the Riemannian metric by an affine connection structure on the group. We show that the canonical Cartan connections of a connected Lie group provides group geodesics including one-parameter subgroups which are completely consistent with the composition and inversion. To extend statistical operations to such a non-metric structure, we propose an implicit definition as an exponential barycenter (there is no Fréchet mean like in Riemannian Manifolds) and a linearly convergent iterative fixed point algorithm to reach it. This results into naturally bi-invariant means which are unique when the dispersion of the data is small enough. In some cases including rotations and rigid-body transformations, there is even a global existence and uniqueness theorem which is similar to the Riemannian case.
- In [14], we investigate the canonical Cartan connections and their associated parallel transport for diffeomorphisms, which justifies the use of one-parameter subgroups (the flow of stationary velocity fields or SVF) for diffeomorphic image registration. In particular, we derive closed-forms for different parallel transports and we compare SVF and LDDMM approaches with experiments on longitudinal and inter-subject registration.
- In [33], we analyses with practical expriments what kind of parallel transport is needed to reorient the deformation characteristics along the time sequences of the cardiac motion. Contrarily to the case of the brain, inter-subject transformations to normalize the heart between different subjects are of the same order than deformations along the sequence.

### 5.3.2. *Statistical Analysis of Longitudinal Transformations in the LDDMM framework*

**Participants:** Stanley Durrleman [Correspondant], Xavier Pennec, Alain Trouvé [CMLA, ENS Cachan], Nicholas Ayache, José Braga [UMR 5288 CNRS-Université Toulouse Paul Sabatier].

Lie groups, transformations, mean value, non-linear registration

The work initiated the previous years with the PhD of S. Durrleman on the spatio-temporal modeling of shapes was applied with J. Braga to quantify ontogenetic differences between bonobo (Pan paniscus) and chimpanzee (Pan troglodytes) endocrania, using dental development as a timeline. We perform a temporal surface regression that estimates typical endocranial ontogenetic trajectories separately for bonobos and chimpanzees which highlights non-linear patterns of endocranial ontogenetic change and significant differences between species at local anatomical levels rather than considering the endocranium as a uniform entity. The decomposition of the spatio-temporal inter-species difference into a morphological deformation (accounting for size and shape differences independently of age) and a time warp (accounting for changes in the dynamics of development) indicates that juvenile bonobos develop much slower than juvenile chimpanzees, suggesting that inter-specific ontogenetic shifts do not only concern endocranial volume increase, but also the rate of shape changes over time. Our method provides, for the first time, a quantitative estimation of inter-specific ontogenetic shifts that appear to differentiate non-linearly. This work was pusblished in the journal of human evolution [10].

### 5.3.3. *The Kernel Bundle Framework for Diffeomorphic Image Registration*

**Participants:** Xavier Pennec [Correspondant], Stefan Sommer [Computer Science Dpt, University of Copenhagen, DK], François Lauze [Computer Science Dpt, University of Copenhagen, DK], Mads Nielsen [Computer Science Dpt, University of Copenhagen, DK].

*This work in collaboration with the Computer Science Department of the University of Copenhagen (DK) was initiated during the 6 month visit of S. Sommer at Asclepios in 2010-2011 and was continued remotely since then.*

non-rigid registration algorithm, statistics, deformations, shapes, locally affine deformations, sparsity

In order to detect small-scale deformations during longitudinal registration while allowing large-scale deformation needed for inter-subject normalization, we wish to model deformation at multiple scales and represent the deformation at the relevant scales only. We combined in [48], [26] a sparsity prior with the multi-scale Kernel Bundle framework, resulting in an algorithm allowing compact representation of deformation across scales.

In [27], we further extend the framework by introducing higher-order momentum distributions in the LDDMM registration framework. While the zeroth order moments previously used in LDDMM only describe local displacement, the first-order momenta that are proposed here represent a basis that allows local description of affine transformations. Beyond the careful mathematical construction, we show the implications for sparse image registration and we provide examples of how the parametrization enables registration with a very low number of parameters.

### 5.3.4. Spectral Correspondances in Non-linear Image Registration

**Participants:** Xavier Pennec [Correspondant], Hervé Lombaert, Nicholas Ayache, Leo Grady [SCR, Princeton, US], Farida Cheriet [Saint-Justine Hospital, Montreal, CA].

*This work was performed in collaboration with Saint-Justine Hospital in Montreal (CA) and Siemens Corporate Research in Princeton (US).*

non-rigid registration algorithm

The demons algorithm was enhanced to include spectral feature correspondences between the images [37]. This feature proves to drastically enhance the robustness of the registration algorithm, which turns out to have a major impact on the construction of atlases. This work was awarded the best paper award at the Medical Computer Vision Workshop [38] and was protected by a patent filling in the US [62]

### 5.3.5. Longitudinal Analysis of Brain Atrophy in Alzheimer's Disease

**Participants:** Marco Lorenzi [Correspondant], Xavier Pennec, Nicholas Ayache, Giovanni B. Frisoni [IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy].

*This work is done in collaboration with LENITEM, IRCCS San Giovanni di Dio Fatebenefratelli, Brescia, Italy.*

Alzheimer's Disease, non-rigid registration algorithm, longitudinal analysis.

The accurate analysis of the longitudinal structural changes in the brain plays a central role in the study of Alzheimer's disease (AD), for diagnostic purposes and for the assessment of the drugs efficacy in clinical trials. The goal of this project is to provide robust and effective instruments based on non-rigid registration of serial MR images for the modeling and the quantification of the brain atrophy evolution in AD. In 2012, our main scientific developments were the following:

- We developed a framework for the consistent definition of anatomical regions of longitudinal brain atrophy, and for the robust quantification of longitudinal regional percentage volume loss. The framework is based on the analysis of the flux associated to longitudinal deformations (see Figure 4 ), and was successfully applied to large public dataset of brain images (ADNI - http://adni.loni.ucla.edu/ ). The work was accepted for oral podium presentation at the MICCAI conference 2012 [40].
- We applied the flux analysis for the quantification of the longitudinal hippocampal and ventricular atrophy in AD. The proposed framework was presented at the NIBAD MICCAI Challenge 2012 [41] , and compared favorably with state-of-art methods in terms of accuracy and stability when applied on the challenge dataset.
- We proposed in [39] a model of the morphological changes in Alzheimer's based on the disentangling of the normal aging component from the pathological atrophy. The model was promoted and presented to the neuroscience community during international scientific conferences [71], [70].

These scientific advances were also included along with the previous ones in the PhD manuscript [1].

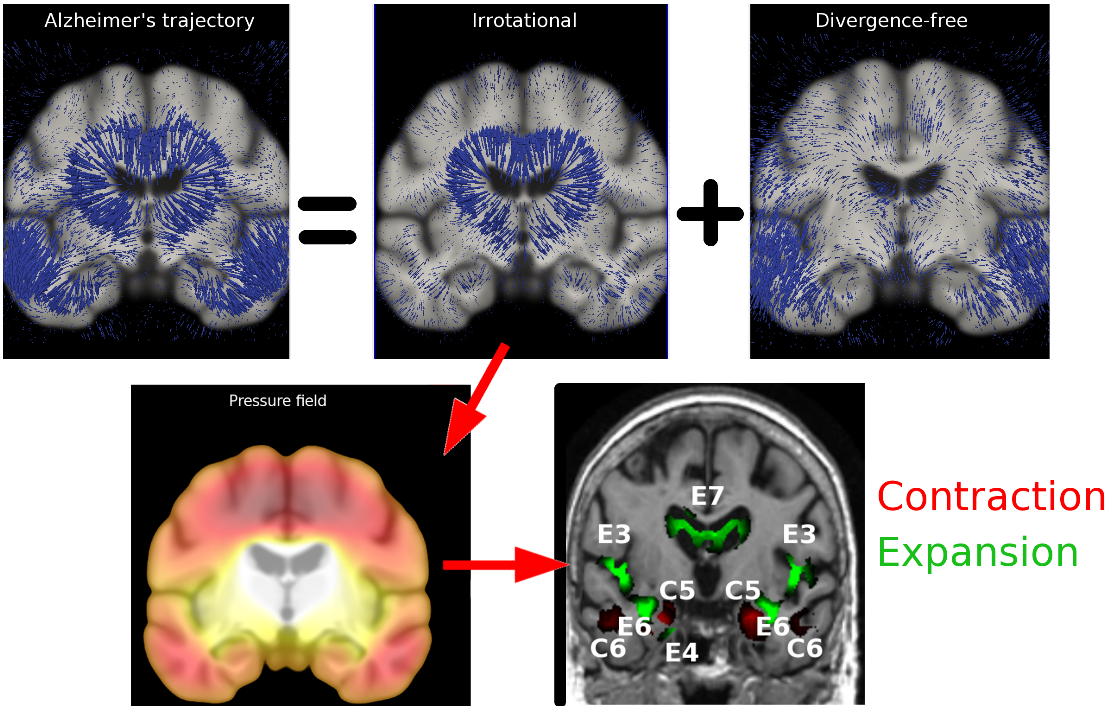


*Figure 4. Flux analysis of the longitudinal morphological changes in Alzheimer's disease. The irrotational component associated to the longitudinal deformation encodes the volume loss and gain (top figure). This component is parameterized by a pressure potential (bottom left) which is used to determine the areas of significant flux of the deformation (matter expansion and contraction) in patients affected by Alzheimer's disease (bottom right).*

### 5.3.6. Statistical Modelling of Cardiac Growth, Deformation and Blood Flow from Medical Images

**Participants:** Kristin McLeod [Correspondant], Adityo Prakosa, Christof Seiler, Maxime Sermesant, Xavier Pennec.

*This work was partially funded by the EU project Care4me ITEA2.*

Image registration, Demons algorithm, LDDMM, reduced models, CFD, polyaffine, cardiac motion tracking

This work involves developing reduced models of cardiac motion, blood flow and growth.

- Extending the 2011 motion tracking challenge [43], the iLogDemons registration algorithm was applied this year to a data-set of synthetic echocardiography sequences with a training set (provided with ground truth) and testing set to quantitavely compare this algorithms with other cardiac motion tracking algorithms [45].
- A reduced order model of cardiac motion based on a polyaffine log-demons registration was developed to represent the motion along the cardiac cycle with a smaller number of parameters compared to previously proposed methods. The method was applied to a data-set of 10 volunteers and the results were presented at the 2012 STACOM workshop at the MICCAI conference [44].
- The analysis of a statistical model for reduced blood flow simulations in the pulmonary artery proposed in the 2010 STACOM workshop is currently being extended to a journal version with an improved method and a larger data-set.
- The statistical modeling of the right ventricle growth in a population of Tetralogy of Fallot patients was extended to a full bi-ventricular growth model on different data [55]. Results confirm the previous findings which were shown to be useful in providing insights for patient treatment [13] (see Figure 5 ).

### 5.3.7. Trees on Geometrical Deformations to Model the Statistical Variability of Organs in Medical Images

**Participants:** Christof Seiler [Correspondant], Xavier Pennec, Mauricio Reyes [Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland].

*This work is performed in the context of the joint PhD of Christof Seiler at the Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland and Asclepios Inria [3].*

Parametrization of diffeomorphisms, Shape statistics, Multiscale and hierarchical trees, Log-Euclidean polyaffine transformations, Polyaffine registration, Log-Demons registration, Generative statistical model, Bayesian registration, Mandibles, Femurs

Intersubject anatomical deformations between patients can be found on coarse and fine scales. Each level of granularity has specific regions of interest in clinical applications. The challenge is to connect geometrical deformations to clinical regions across scales.

- We presented this connection by introducing structured diffeomorphic registration [24]. At the core of our method is the parametrization of geometrical deformations with trees of locally affine transformations describing intersubject variability across scales (see Figure 6 ).
- The methodology of [24] was successfully applied to mandible implant design [31] and in a clinical journal paper on allograft selection [21].
- We statistically modeled the deformation parameters in a population by formulating a generative statistical model [47]. This model allowed us to incorporate deformation statistics as a prior in a Bayesian setting and it enabled us to extend the classical sequential coarse to fine registration to a simultaneous optimization of all scales.
- We explored cell shape statistics to classify stem cells [23].
- We investigated the benefits of considering patient metadata and morphometric measures to enhance bone surface shape prediction [6].

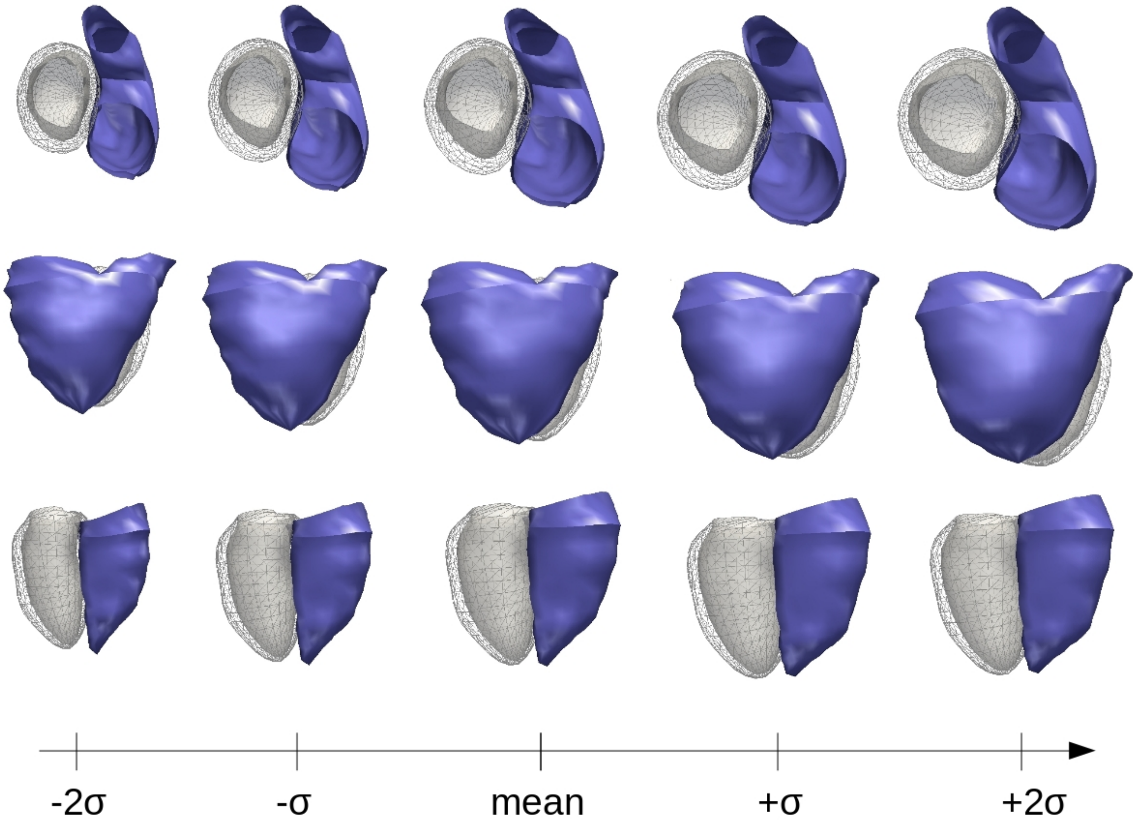


*Figure 5. Mean growth model computed from a population of 13 repaired Tetralogy of Fallot patients. Both ventricles grow as body surface area (BSA) increases.*

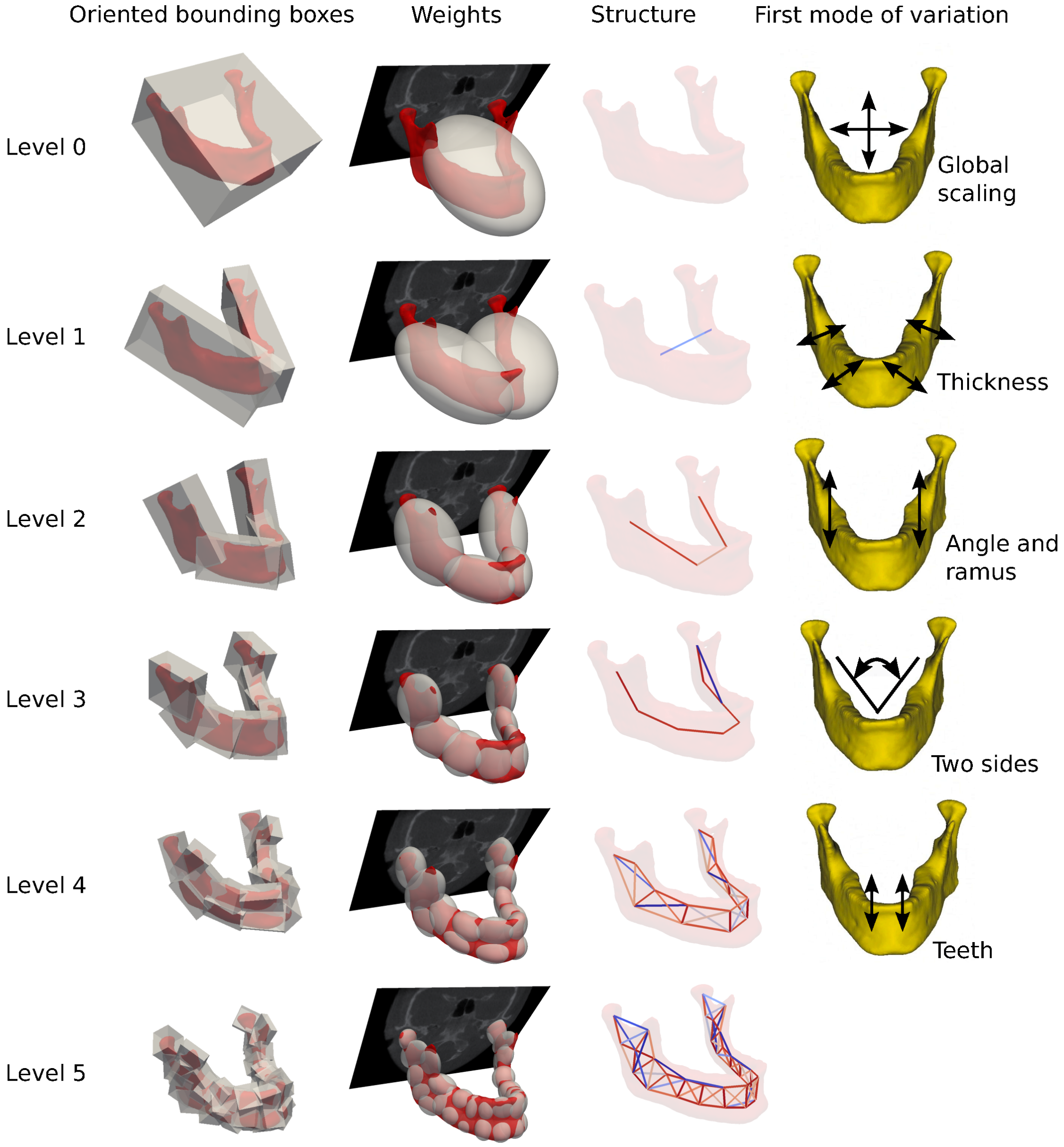


*Figure 6. The red mandible is the surface extracted from the template image. Column 1: Oriented bounding boxes computed using the algorithm presented in [24]. Column 2: Ellipsoids representing Gaussian weights derived from the oriented bounding boxes. Column 3: Structure given by the weights with correlations between regions ranging from low (blue=0:4) to high (red=1). Column 4: First PCA mode at each level showing the residual variation.*

### 5.3.8. Evaluation of iLogDemons Algorithm for Cardiac Motion Tracking in Synthetic Ultrasound Sequence

**Participants:** Adityo Prakosa [Correspondant], Kristin McLeod, Maxime Sermesant, Xavier Pennec.

*This work was partially funded by the European Research Council (ERC) through the support of the MedYMA advanced grant 291080 and the European project euHeart.*

synthetic echocardiography, iLogDemons, cardiac motion tracking

- The LogDemons and iLogDemons non-linear registration algorithms were evaluated on a dataset of synthetic cardiac ultrasound sequences [32], [45]. With these synthetic sequences, it is possible to quantify the performance of these registration algorithms since the ground truth motion was given. Therefore the LogDemons/ iLogDemons can be evaluated objectively (see Figure 7 ).

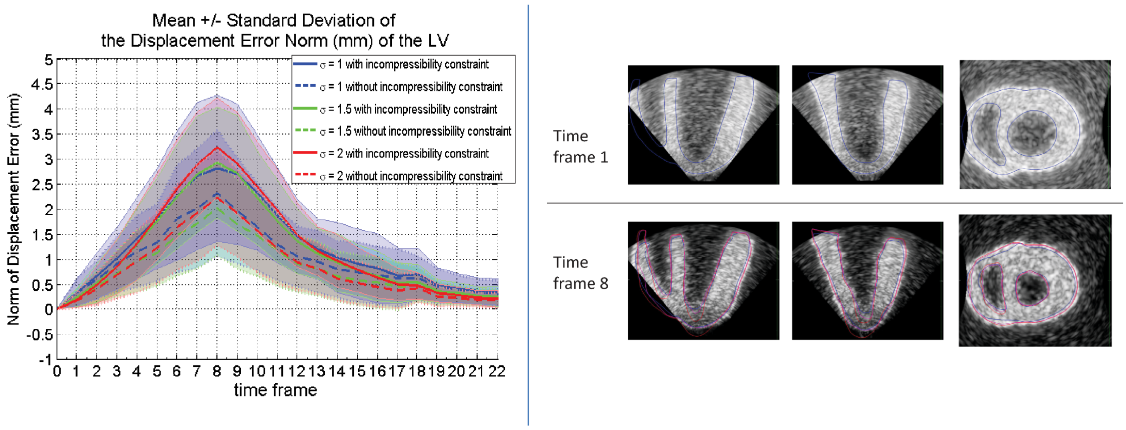


*Figure 7. **Registration Error and Motion Tracking Result** The error quantification for different registration parameter (left) and the myocardium tracking result during the maximum contraction (right): ground truth, LogDemons and iLogDemons estimation are shown in blue, purple and red respectively.*

### 5.3.9. Simulation of Atrophy in Alzheimer's disease

**Participants:** Arnaud Le Carvennec [Correspondant], Sebastien Ourselin [UCL], Nick Fox [UCL], Xavier Pennec [Inria], Nicholas Ayache [Inria].

*Thesis in collaboration between Asclepios team at Inria and Center for Medical Image Computing (CMIC)-Dementia Research Center (DRC) at University College London (UCL).*

Simulation, Alzheimer's disease, registration

- Evaluation of registration algorithm using multi-channel images.
- Simulation of atrophy based on registration.

## 5.4. Computational Physiology

### 5.4.1. Tumor Growth Modeling

**Participants:** Erin Stretton [Correspondant], Emmanuel Mandonnet, Bjoern Menze, Hervé Delingette, Nicholas Ayache.

*This work was funded by Care4me, EU program.*

DTI, MRI, simulation, clinical, tumor, brain, glioma

We aim at developing image analysis methods and biophysical models in order to guide the planning of therapies (surgical removal and radiotherapy) for brain cancer (glioma) patients. Our work is focused on those objectives :

- Predicting the location of glioma recurrence after a resection surgery [49].
- Determining the best description of tumor cell diffusion tensor in white matter (patient-based, atlas-based or isotropic) which leads to the most accurate results for predicting future tumor growth.
- Comparing tumor growth speeds on 3 patient cases. This is a work in progress and the objective is to reach 30 patients when the work is complete.

### 5.4.2. Generation of Synthetic but Visually Realistic Time Series of Cardiac Images Combining a Biophysical Model and Clinical Images

**Participants:** Adityo Prakosa [Correspondant], Maxime Sermesant, Hervé Delingette, Stéphanie Marchesseau, Eric Saloux [CHU Caen], Pascal Allain [Philips Healthcare], Nicolas Villain [Philips Healthcare], Nicholas Ayache.

*This work was done in collaboration with Medisys, Philips Healthcare Suresnes, France, and the Cardiology Department of CHU Caen, France. This work was partially supported by the European Research Council through the ERC Advanced Grant MedYMA on Biophysical Modelling and Analysis of Dynamic Medical Images and the European project euHeart.*

synthetic 4D cardiac sequences, cardiac electromechanical model, non-rigid registration

- A pipeline to create visually realistic synthetic 4D cardiac sequence using the cardiac motion simulated by an electromechanical model is developed. This pipeline combines the simulated myocardium displacement field with the estimated myocardium displacement field from a registration method. This combined displacement field is then used to warp the original images in order to create the synthetic cardiac sequence.
- In [19],we proposed a new approach based on Stationary Velocity Fields to combine the two motions (see Figure 8 ). We also proposed a new method that diffuses displacement fields in order to maintain the continuity between the simulation and the real image with minimal texture distortion. Thanks to the detailed interplay between image processing and biophysical modeling, we can fully use a complete sequence in order to generate several new ones. This method also gives better realism compared to traditional methods based on the deformation of an end-diastolic image, since the generated synthetic sequence will also contain the motion of surrounding tissues such as the motion of the mitral valve.
- The new synthetic images are similar to the original ones except for the motion of the heart which is modified to follow the motion provided by a biophysical model. The parameters of the biophysical model can be modified to create variations around this motion. This pipeline has been applied to generate different synthetic sequences from different imaging modalities. It is generic and can be used with a different biophysical model or a different image registration algorithm, and it can be extended to other organs.
- As these synthetic 4D cardiac sequences have kinematic ground truth information, those sequences represent in themselves a valuable resource to benchmark motion tracking methods or to train machine-learning algorithm.

### 5.4.3. Real-Time Cardiac Electrophysiology Computing for Training Simulator

**Participants:** Hugo Talbot [Correspondant], Hervé Delingette, Stephane Cotin, Maxime Sermesant, Christian Duriez.

*This work was funded by the ADT Sofa and is conducted in collaboration with project teams Shacra and Evasion.*

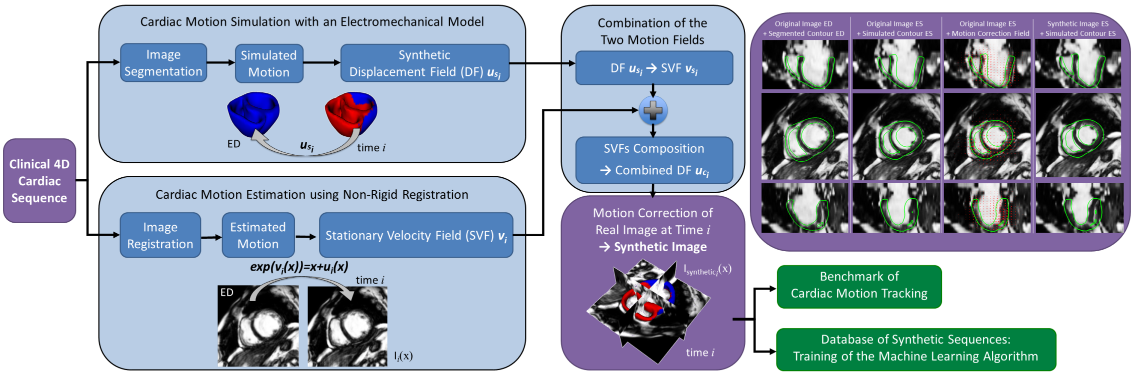


*Figure 8.* ***Synthetic 4D Cardiac Sequence Generation Pipeline*** *A clinical 4D sequence is used as an input to create a synthetic 4D sequence in which the myocardium motion follows a prescribed simulated displacement field. The combined simulated and registration motion are used to correct the motion of the real clinical images in order to create the synthetic cardiac sequence.*

Cardiac electrophysiology simulation, real-time, GPU computing, patient-specific study

Cardiac arrhythmia is a very frequent pathology related to an abnormal electrical activity in the myocardium. This work aims at developing a training simulator for interventional radiology and thermo-ablation of these arrhythmias.

- The latest improvements lead on electrophysiology simulation (see Figure 9 ) using GPU computing allowed us to reach real-time performance[51]. The issue of fast electrophysiology was a major bottleneck in the development of our simulator.

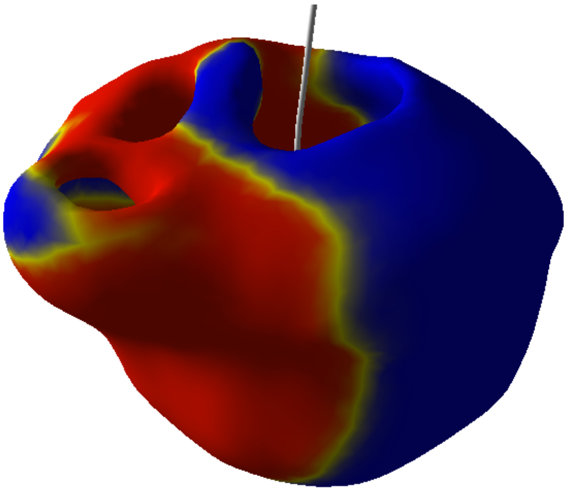


*Figure 9. Cardiac electrophysiology computed on a patient-specific geometry*

Coupling between the cardiac electrophysiology model with cardiac mechanical models has been achieved, thus leading to an interactive framework. Moreover, the electrophysiology simulation has been also coupled with a navigation simulation.

- In the context of his work on cardiac electrophysiology, we initiated two different collaborations. Joint work has been performed with the team CARMEN from Inria Bordeaux on bidomain modeling for cardiac electrophysiology. This exchange targeted the implementation in SOFA of these models. Secondly, a collaboration with the MACS team in Saclay has been initiated to personalize the cardiac electrophysiology model based on the Verdandi library.

### 5.4.4. Personalized model of the heart for cardiac therapy planning

**Participants:** Stéphanie Marchesseau [Correspondant], Hervé Delingette, Nicholas Ayache, Maxime Sermesant.

*An award has been won for this work at the MICCAI 2012 Conference. It was partially funded by the European Community's euHeart project under grant agreement 224495 and by the ERC advanced Grant MedYMA 291080.*

Cardiac simulation, sensitivity analysis, calibration algorithm, specificity study

- We implemented the full Bestel-Clement-Sorine electromechanical model of the heart in SOFA [54], [51].
- We ran a complete sensitivity analysis to check its behaviour for healthy and pathological cases [15].
- A new calibration algorithm was proposed [15] in order to initialize global mechanical parameters from the volume and pressure curves, before further personalization (see Figure 10 ).
- The application of this new method on 6 healthy and 2 pathological cases allowed to draw preliminary conclusions on specific parameters to a given pathology [42], [16].
- The model has also been used to create synthetic images in [19] and for the data of the STACOM 2012 challenge [32].

### 5.4.5. Image-based glioma modeling for radiotherapy planning

**Participants:** Bjoern Menze [Correspondant], Ender Konukoglu [MSR Cambridge], Jan Unkelbach [Harvard MGH].

- Implemented the generative tumor segmentation model together with the E. Konukoglu's tumor infiltration model for evaluation at the MGH Department of Radiation Oncology.
- Integrated tumor infiltration model with radiation therapy model.

### 5.4.6. Cardiac Arrhythmia Radio-frequency Ablation Planning

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

*This work is performed in the context of the the PhD of Rocio Cabrera Lozoya in collaboration with the IHU LIRYC Bordeaux and is funded by ERC MedYMA.*

- Biophysical model development for the prediction of radio frequency ablation sites for ventricular tachycardias.
- Target site map generation for ablation therapy guidance
- Structural and functional characterization of target sites using 3D imaging and EP measurements through machine learning algorithms (see Figure 11 ).
- Prediction validation with acquired clinical data

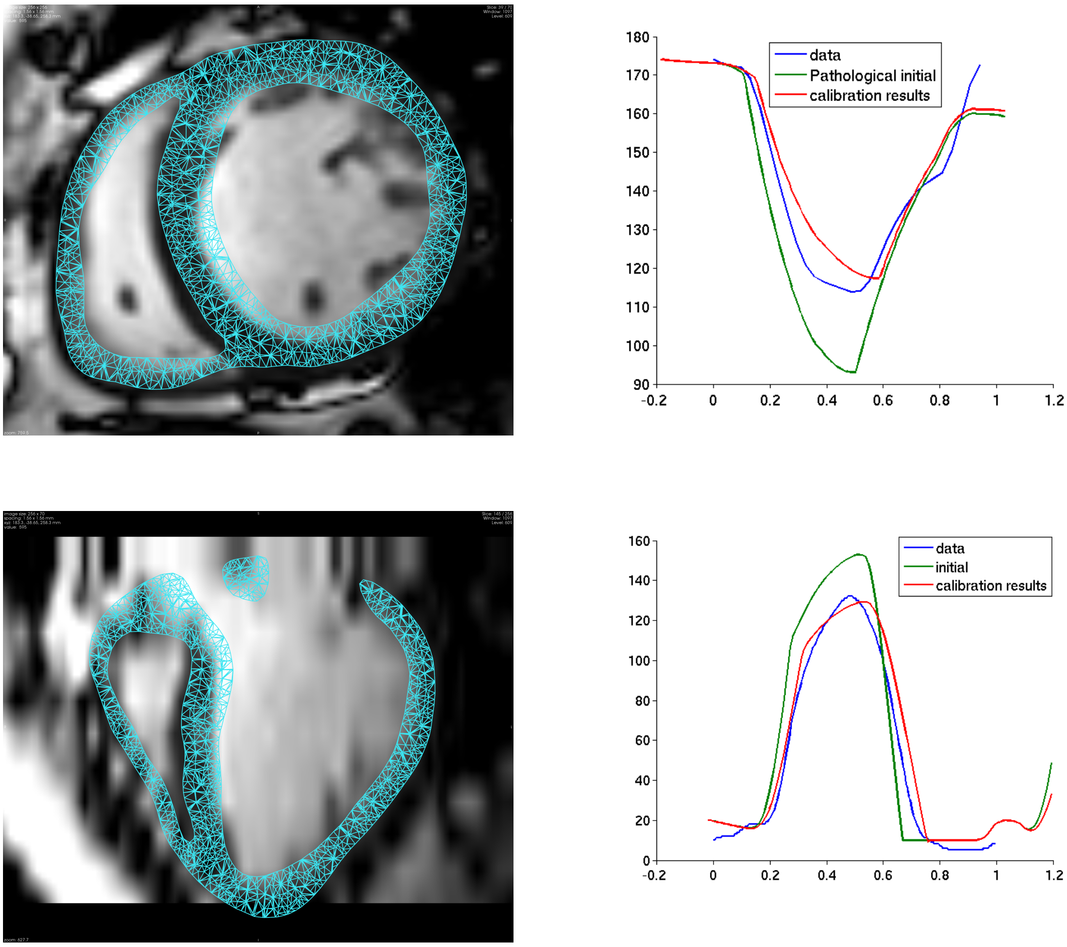


*Figure 10. Results of the calibration algorithm fro one pathological case. (Left) Simulated mesh after calibration compared to the images (at end-diastole). (Right) Resulting volume and pressure curves.*

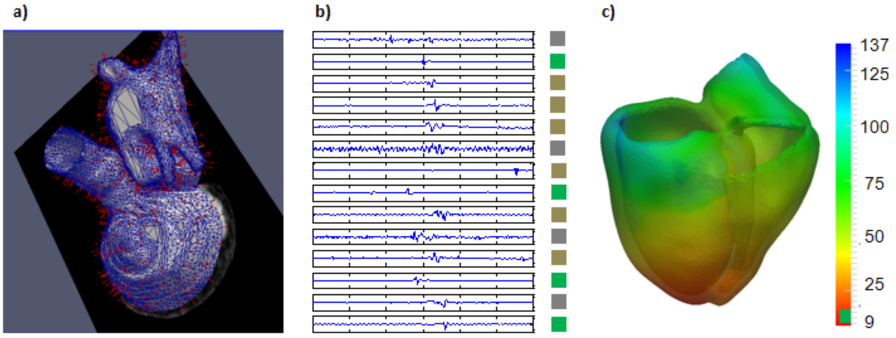


*Figure 11. a) MRI Feature Extraction b) EP Signal Feature Extraction c) EP Model Personalization*

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

**Participants:** Chloe Audigier [Correspondant], Herve Delingette, Tommaso Mansi [Siemens Corporate Research], Nicholas Ayache.

*This PhD is carried out between Asclepios research group, Inria Sophia Antipolis, France and the Image Analytics and Informatics global field, Siemens Corporate Research, Princeton, USA.*

Therapy planning, radio-frequecy ablation, Liver

The objective of this work is to develop a computational framework for patient-specific planning of radiofrequency ablation:

- A patient-specific detailed anatomical model of the liver is estimated from standard CT image and meshed to generate a tetrahedral volume mesh.
- A porous media model is used to compute the patient-specific blood flow in the hepatic circulatory system.
- Bio-heat equations have been implement in SOFA to model the heat propagation in biological tissues.
- A cell death model is included to account for the cellular necrosis.

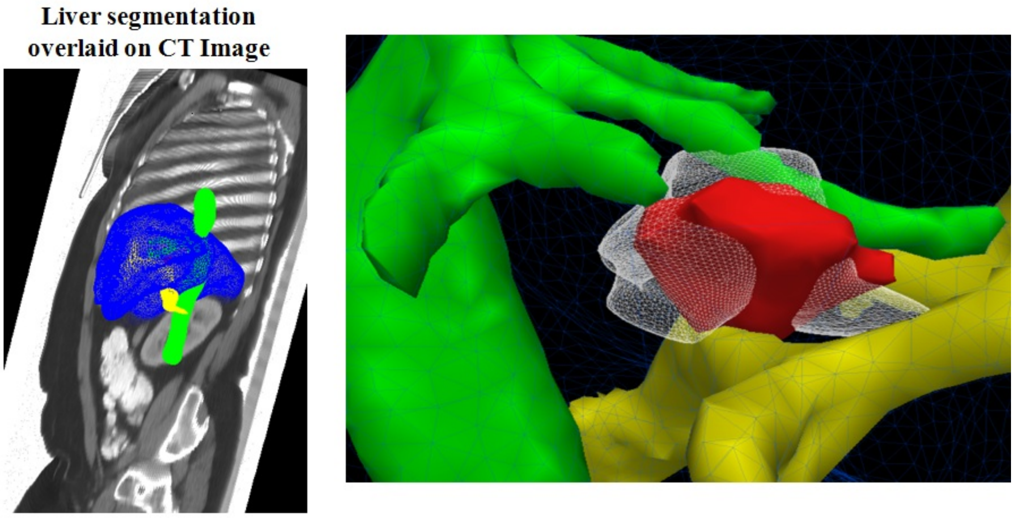


*Figure 12. (Left) Anatomical model of the liver estimated from standard clinical CT image. (Right) The predicted necrosis computed with our model compares qualitatively well with the necrosis region observed on a post-operative MRI scan.*

### 5.4.8. Tumor Growth Simulation for the creation of a database of virtual patients

**Participants:** Nicolas Cordier [Correspondant], Nicholas Ayache, Hervé Delingette, Bjoern Menze, Ezequiel Geremia.

*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, Tumor simulation.

- Synthesizing multi-channel MR images with healthy and glial tumors.
- Creating a database of synthetic images for training and validating of brain tumor segmentation algorithms.

### 5.4.9. Learning approach for the Mechanical personalization of cardiac models

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

*This work was partly funded by Microsoft Research through its PhD Scholarship Programme and by the ERC Advanced Grant MedYMA.*

Inverse problem, machine learning, patient-specific, current, kinematics

- A machine-learning framework for the mechanical personalization of the Bestel-Clement-Sorine model of the heart from patient-specific kinematics
- The computational burden is moved to an offline stage, where the inter-subject variability in motion is captured via the statistical analysis of training samples
- Towards a probabilistic framework for the personalization and therapy planning problems, to better account for significant and diverse uncertainty sources
- Published at the MICCAI 2012 Workshop on Medical Computer Vision[36]

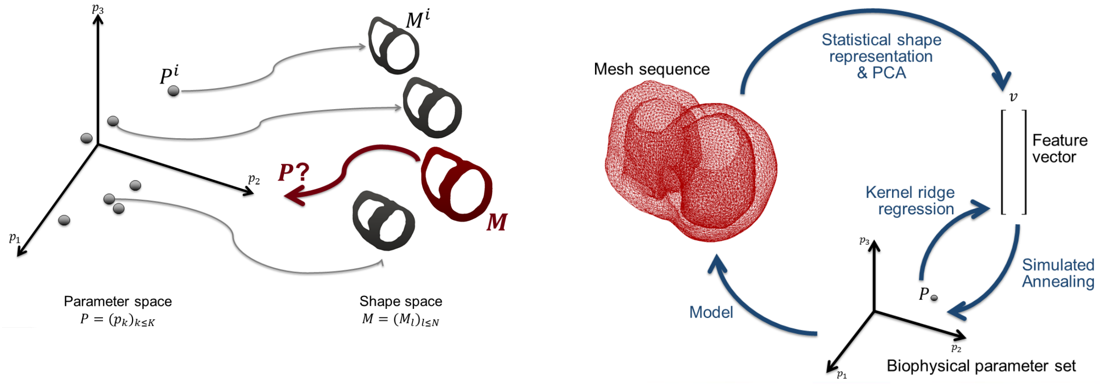


*Figure 13.*

### 5.4.10. Brain Tumor Growth Modeling

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

Gliomas simulations, reaction-diffusion, brain tumors

- In collaboration with the MC2 research team in Bordeaux, we developed a tumor growth model based on different types of cell : necrotic, proliferative and quiescent cells (see Figure 14 ). It is also based on the underlying vascularization of the brain.
- We studied the impact of the vascularization angiogenesis factor and degradation factor.

### 5.4.11. Modeling of atrophy of the brain in Alzheimer's Disease

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

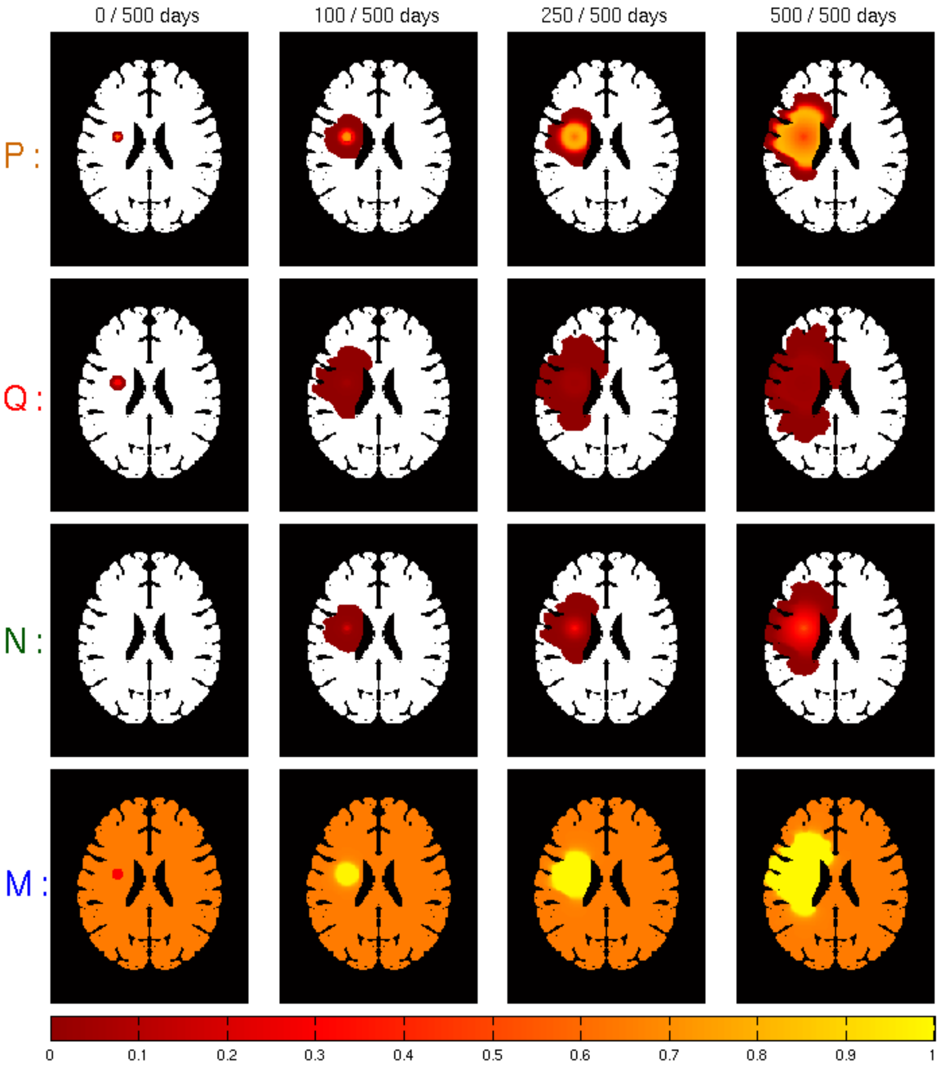


*Figure 14. Results of a glioma simulation at day 0, 50, 250 and 500. The proliferative cells are on the first row, the quiescent cells on the second row, the necrotic cells on the third one and the vascularization is on the fourth row.*

Alzheimer’s Disease (AD), modeling atrophy, biomechanical model

- The idea is to have a model which produces deformation of the brain when a known distribution of local volume change (atrophy) is prescribed to the model. The study is to understand how brain deformation evolve in time with respect to temporal and spatial variation of atrophy.
- During the masters internship period a simple model was tested in 2D square and 3D cube where high atrophy regions acted as sinks for the displacement field [69] (see Figure 15 ).

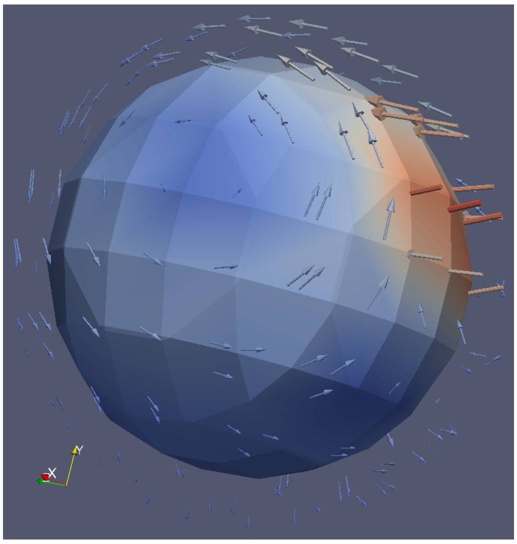


*Figure 15. Displacement field in 3D when high atrophy is prescribed in the center of a cube.*

# ATHENA Project-Team

# 6. New Results

## 6.1. Computational Diffusion MRI

### *6.1.1. Improving dMRI Signal and Acquisitions*

*6.1.1.1. Diffusion MRI Signal Reconstruction with Continuity Constraint and Optimal Regularization*

**Participants:** Emmanuel Caruyer, Rachid Deriche.

In diffusion MRI, the reconstruction of the full Ensemble Average Propagator (EAP) provides new insights in the diffusion process and the underlying microstructure. The reconstruction of the signal in the whole Q-space is still extremely challenging however. It requires very long acquisition protocols, and robust reconstruction to cope with the very low SNR at large b-values. Several reconstruction methods were proposed recently, among which the Spherical Polar Fourier (SPF) expansion, a promising basis for signal reconstruction. Yet the reconstruction in SPF is still subject to noise and discontinuity of the reconstruction. In this work, we present a method for the reconstruction of the diffusion attenuation in the whole Q-space, with a special focus on continuity and optimal regularization. We derive a modified Spherical Polar Fourier (mSPF) basis, orthonormal and compatible with SPF, for the reconstruction of a signal with continuity constraint. We also derive the expression of a Laplace regularization operator in the basis, together with a method based on generalized cross validation for the optimal choice of the parameter. Our method results in a noticeable dimension reduction as compared with SPF. Tested on synthetic and real data, the reconstruction with this method is more robust to noise and better preserves fiber directions and crossings.

This work has been published in [13]

*6.1.1.2. A Computational Framework for Experimental Design in Diffusion MRI*

**Participants:** Emmanuel Caruyer, Rachid Deriche.

In this work, we develop a computational framework for optimal design of experiment in parametric signal reconstruction. We apply this to the optimal design of one dimensional Q-space, Q-ball imaging and multiple Q-shell experimental design. We present how to construct sampling scheme leading to minimal condition number, and compare to state-of-the-art sampling methods. We show in particular a better noise performance of these scheme through Monte-Carlo simulations for the reconstruction of synthetic signal. This demonstrates the impact of this computational framework on acquisition in diffusion MRI.

This work has been published in [16]

*6.1.1.3. Parametric Dictionary Learning in Diffusion MRI*

**Participants:** Sylvain Merlet, Emmanuel Caruyer, Aurobrata Ghosh, Rachid Deriche.

*This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.*

In this work, we propose an approach to exploit the ability of compressive sensing to recover diffusion MRI signal and its characteristics from a limited number of samples. Our approach is threefold. First, we learn and design a parametric dictionary from a set of training diffusion data. This provides a highly sparse representation of the diffusion signal. The use of a parametric method presents several advantages: we design a continuous representation of the signal, from which we can analytically recover some features such as the ODF; besides, the dictionary we train is acquisition-independent. Next, we use this sparse representation to reconstruct the signal of interest, using cross-validation to assess the optimal regularization parameter for each signal reconstruction. The use of cross-validation is critical in the L1 minimization problem, as the choice of the parameter is sensitive to the noise level, the number of samples, and the data sparsity. Third, we use a polynomial approach to accurately extract ODF maxima. Finally, we motivate and describe the choice of experimental parameters for the HARDI contest.

This work has been published in [26].

#### *6.1.1.4. Diffusion and Multiple Orientations from 1.5 MR Systems with Limited Gradient Tables*

**Participants:** Sylvain Merlet, Rachid Deriche, Kevin Whittingstall [Radiology department, Université de Sherbrooke, Québec, Canada], Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

*This work has been performed within the framework of the Brain Connectivities Associate Team.*

Diffusion MRI (dMRI) enables the quantification of water diffusion, influenced by the structure of biological tissues, from the acquisition of diffusion weighted magnetic resonance images (DW-MRI). While recent advances enable to recover complex fiber geometries using diffusion measurements along various sampling schemes of high order, some older MR systems work with limited gradient tables (ex: maximum of 6 or 12 directions). These systems are designed for Diffusion Tensor Imaging (DTI). Several hospitals and research institutes in the world are limited by these fixed DTI gradient sets. Therefore, groups that want to perform state-of-the-art tractography using high angular resolution diffusion imaging (HARDI) data are penalized and can only perform DTI tractography on their old system. The Gaussian assumption of the tensor model, in DTI, is an over simplification of the diffusion phenomenon of water molecules in the brain and thus cannot resolve crossing fibers. In this work, we show that new diffusion signal modeling and processing techniques enable to capture complex angular structure of the diffusion process even from a reduced gradient direction set arising from an older MR system.

This work has been published in [27].

#### *6.1.1.5. A Robust variational approach for simultaneous smoothing and estimation of DTI*

**Participants:** Rachid Deriche, Meizhu Liu [Department of CISE, University of Florida, Gainesville, USA], Baba C. Vemuri [Department of CISE, University of Florida, Gainesville, USA].

Estimating diffusion tensors is an essential step in many applications — such as diffusion tensor image (DTI) registration, segmentation and fiber tractography. Most of the methods proposed in the literature for this task are not simultaneously statistically robust and feature preserving techniques. In this work, we propose a novel and robust variational framework for simultaneous smoothing and estimation of diffusion tensors from diffusion MRI. Our variational principle makes use of a recently introduced total Kullback–Leibler (tKL) divergence for DTI regularization. tKL is a statistically robust dissimilarity measure for diffusion tensors, and regularization by using tKL ensures the symmetric positive definiteness of tensors automatically. Further, the regularization is weighted by a non-local factor adapted from the conventional non-local means filters. Finally, for the data fidelity, we use the nonlinear least-squares term derived from the Stejskal–Tanner model. We present experimental results depicting the positive performance of our method in comparison to competing methods on synthetic and real data examples.

This work has been accepted for publication in NeuroImage [63].

### *6.1.2. Modeling in Diffusion MRI*

#### *6.1.2.1. Fast and Analytical EAP Approximation from a 4th Order Tensor*

**Participants:** Aurobrata Ghosh, Rachid Deriche.

*This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.*

Generalized Diffusion Tensor Imaging (GDTI) was developed to model complex Apparent Diffusivity Coefficient (ADC) using Higher Order Tensors (HOT) and to overcome the inherent single-peak shortcoming of DTI. However, the geometry of a complex ADC profile doesn't correspond to the underlying structure of fibers. This tissue geometry can be inferred from the shape of the Ensemble Average Propagator (EAP). Though interesting methods for estimating a positive ADC using 4th order diffusion tensors were developed, GDTI in general was overtaken by other approaches, e.g. the Orientation Distribution Function (ODF), since it is considerably difficult to recuperate the EAP from a HOT model of the ADC in GDTI. In this work, we present a novel closed-form approximation of the EAP using Hermite polynomials from a modified HOT model of the original GDTI-ADC. Since the solution is analytical, it is fast, differentiable, and the approximation converges well to the true EAP. This method also makes the effort of computing a positive ADC worthwhile, since now both the ADC and the EAP can be used and have closed forms. We demonstrate our approach with 4th order tensors on synthetic data and in vivo human data.

This work has been accepted for publication in the International Journal of Biomedical Imaging [54].

#### *6.1.2.2. A Polynomial Approach for Extracting the Extrema of a Spherical Function and its Application in Diffusion MRI*

**Participants:** Aurobrata Ghosh, Elias Tsigaridas [PolSys Project-Team, Inria, Paris Rocquencourt, France], Bernard Mourrain [Galaad Project-Team, Inria, Sophia Antipolis, Méditerranée, France], Rachid Deriche.

*This work has been partially supported by the ANR project NucleiPark and the France-Parkinson Association.*

Antipodally symmetric spherical functions play a pivotal role in diffusion MRI in representing sub-voxel-resolution microstructural information of the underlying tissue. This information is described by the geometry of the spherical function. In this work, we propose a method to automatically compute all the extrema of a spherical function. We then classify the extrema as maxima, minima and saddle-points to identify the maxima. We take advantage of the fact that a spherical function can be described equivalently in the spherical harmonic (SH) basis, in the symmetric tensor (ST) basis constrained to the sphere, and in the homogeneous polynomial (HP) basis constrained to the sphere. We extract the extrema of the spherical function by computing the stationary points of its constrained HP representation. Instead of using traditional optimization approaches, which are inherently local and require exhaustive search or re-initializations to locate multiple extrema, we use a novel polynomial system solver which analytically brackets all the extrema and refines them numerically, thus missing none and achieving high precision. To illustrate our approach we consider the Orientation Distribution Function (ODF). In diffusion MRI, the ODF is a spherical function which represents a state-of-the-art reconstruction algorithm whose maxima are aligned with the dominant fiber bundles. It is, therefore, vital to correctly compute these maxima to detect the fiber bundle directions. To demonstrate the po- tential of the proposed polynomial approach we compute the extrema of the ODF to extract all its maxima. This polynomial approach is, however, not dependent on the ODF and the framework presented in this work can be applied to any spherical function described in either the SH basis, ST basis or the HP basis.

This work has been submitted to Medical Image Analysis and has been accepted for a publication to appear early 2013 [57].

#### *6.1.2.3. 4th Order Symmetric Tensors and Positive ADC Modelling*

**Participants:** Aurobrata Ghosh, Rachid Deriche.

High Order Cartesian Tensors (HOTs) were introduced in Generalized DTI (GDTI) to overcome the limitations of DTI. HOTs can model the apparent diffusion coefficient (ADC) with greater accuracy than DTI in regions with fiber heterogeneity. Although GDTI HOTs were designed to model positive diffusion, the straightforward least square (LS) estimation of HOTs doesn't guarantee positivity. In this work, we address the problem of estimating 4th order tensors with positive diffusion profiles. Two known methods exist that broach this problem, namely a Riemannian approach based on the algebra of 4th order tensors, and a polynomial approach based on Hilbert's theorem on non-negative ternary quartics. In this work, we review the technicalities of these two approaches, compare them theoretically to show their pros and cons, and compare them against the Euclidean LS estimation on synthetic, phantom and real data to motivate the relevance of the positive diffusion profile constraint.

This work is under submission.

#### *6.1.2.4. Higher-Order Tensors in Diffusion Imaging: A Survey*

**Participants:** Thomas Schultz [MPI for Intelligent Systems, Tubingen, Germany], Andrea Fuster [Eindhoven University of Technology, The Netherlands], Aurobrata Ghosh, Luc Florack [Eindhoven University of Technology, The Netherlands], Rachid Deriche, Lek-Heng Lim [University of Chicago, USA].

Diffusion imaging is a noninvasive tool for probing the microstructure of fibrous nerve and muscle tissue. Higher-order tensors provide a powerful mathematical language to model and analyze the large and complex data that is generated by its modern variants such as High Angular Resolution Diffusion Imaging (HARDI) or Diffusional Kurtosis Imaging. This survey gives a careful introduction to the foundations of higher-order tensor algebra, and explains how some concepts from linear algebra generalize to the higher-order case. From the application side, it reviews a variety of distinct higher-order tensor models that arise in the context of diffusion imaging, such as higher-order diffusion tensors, q-ball or fiber Orientation Distribution Functions (ODFs), and fourth-order covariance and kurtosis tensors. By bridging the gap between mathematical foundations and application, it provides an introduction that is suitable for practitioners and applied mathematicians alike, and propels the field by stimulating further exchange between the two.

This work has been submitted and is under review.

#### 6.1.2.5. Nonnegative Definite EAP and ODF Estimation via a Unified Multi-Shell HARDI Reconstruction

**Participants:** Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Tianzi Jiang [LIAMA, China].

*This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.*

In High Angular Resolution Diffusion Imaging (HARDI), Orientation Distribution Function (ODF) and Ensemble Average Propagator (EAP) are two important Probability Density Functions (PDFs) which reflect the water diffusion and fiber orientations. Spherical Polar Fourier Imaging (SPFI) is a recent model-free multi-shell HARDI method which estimates both EAP and ODF from the diffusion signals with multiple b values. As physical PDFs, ODFs and EAPs are nonnegative definite respectively in their domains $S2$ and $R3$. However, existing ODF / EAP estimation methods like SPFI seldom consider this natural constraint. Although some works considered the nonnegative constraint on the given discrete samples of ODF / EAP, the estimated ODF/EAP is not guaranteed to be nonnegative definite in the whole continuous domain. The Riemannian framework for ODFs and EAPs has been proposed via the square root parameterization based on pre-estimated ODFs and EAPs by other methods like SPFI. However, there is no work on how to estimate the square root of ODF/EAP called as the wavefuntion directly from diffusion signals. In this work, based on the Riemannian framework for ODFs / EAPs and Spherical Polar Fourier (SPF) basis representation, we propose a unified model-free multi-shell HARDI method, named as Square Root Parameterized Estimation (SRPE), to simultaneously estimate both the wavefunction of EAPs and the nonnegative definite ODFs and EAPs from diffusion signals. The experiments on synthetic data and real data showed SRPE is more robust to noise and has better EAP reconstruction than SPFI, especially for EAP profiles at large radius.

This work has been published in [11] and [18].

#### 6.1.2.6. An Intrinsic Diffeomorphism Invariant Riemannian Framework for Probability Density Function Computing in diffusion MRI

**Participants:** Rachid Deriche, Jian Cheng [ATHENA and LIAMA, China], Aurobrata Ghosh, Tianzi Jiang [LIAMA, China].

*This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.*

In High Angular Resolution Imaging (HARDI),Ensemble Average Propagator (EAP) and Orientation Distribution Function (ODF) are two important Probability Density Functions (PDFs), which describe the diffusion probability respectively in 3D space and along directions. Fisher information metric has been successfully applied in Diffusion Tensor Imaging (DTI) on tensor estimation, filtering, registration, statistical analysis, etc. However, to our knowledge, existing works in HARDI mainly focus on ODF/EAP estimation, not on ODF and EAP data processing. In this work, we propose a general state-of-the-art Riemannian framework as a mathematical tool to process such PDF data, by representing the square root of the PDF, called *wavefunction* based on quantum mechanics, as a linear combination of some orthonormal basis functions. The proposed Riemannian framework is showed to be a natural extension of previous Riemannian framework for tensors. We deduced the Riemannian metric for the PDF family via orthonormal basis representation, and proved the statistical manifold to be a convex subset of a high dimensional sphere. In this framework, the exponential map, logarithmic map and geodesic have closed forms, and the weighted Riemannian mean and median uniquely exist. Moreover, we generalized the Log-Euclidean framework and the Geodesic Anisotropy (GA) form tensors to ODFs/EAPs. The theoretical results can be applied to any general PDF data under any orthonormal basis representation. Furthermore we analyzed theoretically the similarities and differences between the Riemannian frameworks for EAPs, ODFs and for tensors, and demonstrated the proposed Riemannian metric is diffeomorphism invariant, which is the natural extension of the previous affine-invariant metric for tensors. Some potential applications were proposed via the Riemannian operations on the ODF/EAP field, such as anisotropy description via GA, nonnegative definite ODF/EAP estimation, interpolation, filtering, Principal Geodesic Analysis (PGA) and atlas estimation. The Riemannian framework and its applications were validated in synthetic, phantom and real data. The experiments demonstrated that the Riemannian framework is very useful for ODF/EAP computing, although the results from Riemannian metric and Euclidean metric are similar for ODFs but much different for EAPs.

This work has been published in [11]. A longer version has been submitted and is under revision for the journal IEEE transaction on Medical Imaging.

#### *6.1.2.7. Ensemble Average Propagator Reconstruction via Compressed Sensing: Discrete or Continuous Bases ?*

**Participants:** Sylvain Merlet, Michael Paquette [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada], Rachid Deriche, Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

*This work has been partly supported within the framework of the Brain Connectivities Associate Team.*

In this work, we propose to compare the sparsity of two classes of representations for the EAP : The discrete representations, via the Haar, Daubechies-Cohen-Fauveau (DCF) 5-3, DCF 9-7 wavelets bases, and the continuous representations, via Spherical Polar Fourier (SPF) and 3D-SHORE bases.

This work has been published in [28].

#### *6.1.2.8. Parametric dictionary learning for modeling EAP and ODF in diffusion MRI*

**Participants:** Sylvain Merlet, Emmanuel Caruyer, Rachid Deriche.

In this work, we propose an original and efficient approach to exploit the ability of Compressed Sensing (CS) to recover Diffusion MRI (dMRI) signals from a limited number of samples while efficiently recovering important diffusion features such as the Ensemble Average Propagator (EAP) and the Orientation Distribution Function (ODF). Some attempts to sparsely represent the diffusion signal have already been performed. However and contrarly to what has been presented in CS dMRI, in this work we propose and advocate the use of a well adapted learned dictionary and show that it leads to a sparser signal estimation as well as to an efficient reconstruction of very important diffusion features. We first propose to learn and design a sparse and parametric dictionary from a set of training diffusion data. Then, we propose a framework to analytically estimate in closed form two important diffusion features : the EAP and the ODF. Various experiments on synthetic, phantom and human brain data have been carried out and promising results with reduced number of atoms have been obtained on diffusion signal reconstruction, thus illustrating the added value of our method over state-of-the-art SHORE and SPF based approaches.

This work has been published in [25].

#### *6.1.2.9. Constrained Diffusion Kurtosis Imaging Using Ternary Quartics and MLE*

**Participants:** Tristan Milne [Queen's University, Kingston, Ontario, Canada], Aurobrata Ghosh, Rachid Deriche.

*This work has been partly supported by the Inria International Internship Program.*

We present a ternary quartic based approach with an improved gradient based optimization scheme for diffusion kurtosis imaging to estimate constrained and physically realistic diffusion and kurtosis tensors. We account for the signal noise by considering a maximum likelihood estimation based on the Rician noise model. Diffusion kurtosis imaging (DKI) is a recent important improvement over the diffusion tensor imaging (DTI) model that quantifies the degree of non-Gaussian diffusion in a tissue using a 3D 4th order tensor. However, DKI estimation needs to consider three constraints to be physically relevant. By adopting the implicit ternary quartic parameterization which allows to elegantly impose a positivity constraint on the kurtosis tensor and by employing gradient based optimization schemes, we show dramatically improved performance in terms of estimation time and quality. We derive the mathematical framework and show results on extensive synthetic data experiments.

This work has been published in [30]. A longer version has been submitted and is under revision for the journal Magnetic Resonance in Medicine.

## 6.1.3. From DW-MRI to Fiber Pathways and Microstructures Recovery

### 6.1.3.1. From Diffusion MRI to Brain Connectomics

**Participants:** Aurobrata Ghosh, Rachid Deriche.

Diffusion MRI (dMRI) is a unique modality of MRI which allows one to indirectly examine the microstructure and integrity of the cerebral white matter in vivo and non-invasively. Its success lies in its capacity to reconstruct the axonal connectivity of the neurons, albeit at a coarser resolution, without having to operate on the patient, which can cause radical alterations to the patient's cognition. Thus dMRI is beginning to assume a central role in studying and diagnosing important pathologies of the cerebral white matter, such as Alzheimer's and Parkinson's diseases, as well as in studying its physical structure in vivo. In this work, we present an overview of the mathematical tools that form the framework of dMRI – from modelling the MRI signal and measuring diffusion properties, to reconstructing the axonal connectivity of the cerebral white matter, i.e., from Diffusion Weighted Images (DWIs) to the human connectome.

This work will be published in [55].

### 6.1.3.2. Biomarkers for HARDI : 2nd & 4th Order Tensor Invariants

**Participants:** Rachid Deriche, Aurobrata Ghosh, Théodore Papadopoulo.

*This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.*

In this paper, we explore the theory of tensor invariants as a mathematical framework for computing new biomarkers for HARDI. We present and explain the integrity basis, basic invariants and principal invariants of 2nd & 4th order tensors to expand on a recently proposed paper on 4th order tensor invariants. We present the mathematical results and compute the basic and principal invariants on a controlled synthetic dataset and an in vivo human dataset. We show how the integrity bases of these two sets of invariants can form a promising framework for developing new biomarkers for HARDI.

This work has been published in [22].

### 6.1.3.3. Generalized Invariants of a 4th order tensor: Building blocks for new biomarkers in dMRI

**Participants:** Aurobrata Ghosh, Théodore Papadopoulo, Rachid Deriche.

*This work has been partly supported by the Association France Parkinson and the ANR NucleiPark project.*

This paper presents a general and complete (up to degree 4) set of invariants of 3D 4th order tensors with respect to SO3. The invariants to SO3 for the 2nd order diffusion tensor are well known and play a crucial role in deriving important biomarkers for DTI, e.g. MD, FA, RA, etc. But DTI is limited in regions with fiber heterogeneity and DTI biomarkers severely lack specificity. 4th order tensors are both a natural extension to DTI and also form an alternate basis to spherical harmonics for spherical functions. This paper presents a systematic method for computing the SO3 invariants of 3D 4th order tensors, derives relationships between the new (generalized) invariants and existing invariants and shows results on synthetic and real data. It also presents, hitherto unknown, new invariants for 4th order tensors. Analogously to DTI, these new invariants can perhaps form building blocks for new biomarkers.

This work has been published in [23].

### 6.1.3.4. Tractography via the Ensemble Average Propagator in diffusion MRI

**Participants:** Sylvain Merlet, Anne-Charlotte Philippe, Rachid Deriche, Maxime Descoteaux [Sherbrooke Connectivity Imaging Laboratory, Computer Science Departement, Université de Sherbrooke, Québec, Canada].

*This work has been partly supported within the framework of the Brain Connectivities Associate Team.*

It's well known that in diffusion MRI (dMRI), fibre crossing is an important problem for most existing diffusion tensor imaging (DTI) based tractography algorithms. To overcome these limitations, High Angular Resolution Diffusion Imaging (HARDI) based tractography has been proposed with a particular emphasis on the the Orientation Distribution Function (ODF). In this work, we advocate the use of the Ensemble Average Propagator (EAP) instead of the ODF for tractography in dMRI and propose an original and efficient EAP-based tractography algorithm that outperforms the classical ODF-based tractography, in particular, in the regions that contain complex fibre crossing configurations. Various experimental results including synthetic, phantom and real data illustrate the potential of the approach and clearly show that our method is especially efficient to handle regions where fiber bundles are crossing, and still well handle other fiber bundle configurations such as U-shape and kissing fibers.

This work has been published in [29].

### 6.1.3.5. Using Radial NMR Profiles to Characterize Pore Size Distributions

**Participants:** Rachid Deriche, John Treilhard [Queen's University, Ontario, Canada].

*This work has been partly supported by the Inria International Internship Program.*

Extracting information about axon diameter distributions in the brain is a challenging task which provides useful information for medical purposes; for example, the ability to characterize and monitor axon diameters would be useful in diagnosing and investigating diseases like amyotrophic lateral sclerosis (ALS) or autism. In [78], three families of operators are defined, whose action upon an NMR attenuation signal extracts the moments of the pore size distribution of the ensemble under consideration; also a numerical method is proposed to continuously reconstruct a discretely sampled attenuation profile using the eigenfunctions of the simple harmonic oscillator Hamiltonian – the SHORE basis. The work we have performed here extends this method to other bases that can offer a better description of attenuation signal behaviour – in particular, we proposed the use of the radial Spherical Polar Fourier (SPF) basis. Testing is performed to contrast the efficacy of the radial SPF basis and SHORE basis in practical attenuation signal reconstruction. The robustness of the method to additive noise is tested and analyzed. We demonstrated that a low-order attenuation signal reconstruction outperforms a higher-order reconstruction in subsequent moment estimation under noisy conditions. We proposed the simulated annealing algorithm for basis function scale parameter estimation. Finally, analytic expressions are derived and presented for the action of the operators on the radial SPF basis (obviating the need for numerical integration, thus avoiding a spectrum of possible sources of error).

This work has been published [20].

### 6.1.3.6. Elliptic Fourier Features of Brain White Matter Pathways

**Participants:** Rachid Deriche, Ali Demir [Sabancy University, TU], Gozde Unal [Sabancy University, TU].

Magnetic resonance imaging provides diffusion weighted images (DMRI), which non-invasively reconstruct the brain white matter pathways. DMRI is used to study brain white matter diseases as well as aid surgical planning. As localization of different white matter pathways surrounding a pathology is crucial for surgical planning, automatic extraction and classification of different anatomical white matter pathways pre-operatively becomes an important computational tool. In this work, we propose a method for classification of brain white matter pathways based on 3D elliptic Fourier descriptors, which are extended from the 2D elliptic Fourier descriptors. We performed experiments and validation of the proposed method on a white matter atlas space and on real pathological cases.

This work has been published [41].

## 6.2. Multi-Imaging Modalities

### *6.2.1. Coupling functional and structural models*

#### *6.2.1.1. A nested cortex parcellation combining analysis of MEG forward problem and diffusion MRI tractography*

**Participants:** Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Understanding the relationship between structure and function is a major challenge in neuroscience. Diffusion MRI (dMRI) in the only non-invasive modality allowing to have access to the neural structure. Magnetoencephalography (MEG) is another non-invasive modality that allows a direct access to the temporal succession of cognitive processes. Functional cortex parcellation being one of the most important ways to understanding structure-function relationship, we propose an innovative method merging MEG and dMRI to parcellate the cortex. The combination of MEG forward problem and connectivity information reveals cortical areas generating a similar magnetic field at sensors while having a similar connectivity. Results show suitable clusters that forecast interesting studies for inter- and intra- subjects comparisons of the cortex parcellations. The automatic nested cortex parcellation we propose could be a first step to analyse sources that are seeds of long or short range connectivity and to differentiate these connectivities in the white matter

This work has been published in [31].

#### *6.2.1.2. dMRI tractography of WM fibers to recover the anatomical connectivity supporting a MEG epileptic network*

**Participants:** Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Cerebral organization is determined by segregated and integrated regions both functionally and anatomically. These cerebral networks are the foundations of the execution of major part of cognitive processes. Information about the structure of the white matter (WM) and the functionality of networks are both needed to understand these cerebral networks.

This work proposes an efficient method to inform a given functional network on its anatomical support: how many anatomical connections exist between functionally connected regions and what are their geometries . Diffusion MRI being the only non invasive method allowing to have access to the micro-structure of the WM, we used diffusion information to underline the degree of connectivity between functionally connected regions while taking advantage of WM fibers reconstruction to determine the way taken by the anatomical network supporting the functional network.

Due to the complex dynamical alteration of epilepsy, the study of large-scale functional connectivity is difficult. But diffusion imaging studies have shown alterations of the WM between epileptic zones and connected areas. This methodology allows to add qualitative (degree of connectivity) and geometrical (WM fibers reconstruction) information on the anatomical network supporting an epileptic network mostly determined by magneto-encephalography (MEG).

This work has been published in [35].

#### *6.2.1.3. Whole cortex parcellation combining analysis of MEG forward problem, structural connectivity and Brodmann's atlas*

**Participants:** Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

Functional cortex parcellation is one of the most important ways to understand the link between structure and function in the brain. Brodmann's atlas remains a fundamental pillar to understand this relationship because its areas are defined by similar cytoarchitecture and functional imaging notably had revealed that they correspond, entirely or in part, to functional areas. So, its integration to diffusion MRI (dMRI) data is pertinent, dMRI being the only non invasive and in-vivo imaging modality able to have access to a detailed geometric description of the anatomical connectivity between brain areas. In this work, we propose to define a new connectivity profile of cortical sources based on the Brodmann's atlas. After its registration to T1 and diffusion weighted images of the same subject, we reconstructed the brain surfaces and considered the cortical sources to be the vertices of the white matter/ grey matter boundary mesh. We performed a probabilistic tractography taking each cortical sources as seeds and the L Brodmann's areas as L targets. Thus, we obtained the connectivity profile of a cortical source: a vector v of size L where v (l) is the degree of connectivity of the source to the l th Brodmann's area. Then, we developped a cortical parcellation method jointly analyzing the MEG forward problem and the connectivity profiles based on Brodmann's atlas of cortical sources. We computed the leadfield matrix that relates the sources to the MEG sensors. We applied a k-means algorithm to the leadfield matrix to cluster sources having a close magnetic field to the MEG sensors. Then, in each leadfield-based cluster, we clustered sources via their connectivity profile based on Brodmann's atlas. This automatic parcellation is an efficient preprocessing to compute a MEG inverse problem on functional data informed by its structural connectivity.

This work has been published in [32].

#### *6.2.1.4. Study of the brain connectivity in an Immersive Space*

**Participants:** Anne-Charlotte Philippe, Jean-Christophe Lombardo [Dream Project-Team, Inria, Sophia Antipolis, Méditerranée, France].

Virtual reality is a powerful tool for scientific visualization. When the amount and complexity of the visualized data grows, standard visualization applications on desktop computers become inefficient. In this work, we present the use of a CAVE like VR facility in a neuroscientific context. The aim is to have a better understanding of the brain connectivity. Both anatomical and functional data are attached to a mesh representing the brain surface.

Specific tools developed for this study and the way we used them are presented in [36] emphasizing drawbacks and advantages of virtual reality in a scientific visualization context.

This work has been published in [36].

#### *6.2.1.5. Cortex parcellation via diffusion data as prior knowledge for the MEG inverse problem*

**Participants:** Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo, Rachid Deriche.

In this work, we present a new approach to the recovery of dipole magnitudes in a distributed source mo-del for magnetoencephalographic (MEG) imaging. This method consists in introducing prior knowledge regarding the anatomical connectivity in the brain to this ill-posed inverse problem. Thus, we perform cortex parcellation via structural information coming from diffusion MRI (dMRI), the only non-invasive modality allowing to have access to the structure of the WM tissues. Then, we constrain, in the MEG inverse problem, sources in the same diffusion parcel to have close magnitude values. Results of our method on MEG simulations are presented and favorably compared with classical source reconstruction methods.

This work is currently under submission.

#### *6.2.1.6. Fractality in the neuron axonal topography of the human brain based on 3-D diffusion MRI*

**Participants:** Panayotis Katsaloulis [Institute of Physical Chemistry "Demokritos" (IPC),National Center for Scientific Research "Demokritos", Greece], Aurobrata Ghosh, Anne-Charlotte Philippe, Astero Provata [Institute of Physical Chemistry "Demokritos" (IPC),National Center for Scientific Research "Demokritos", Greece], Rachid Deriche.

In this work, we conduct a group study, with 18 subjects, to validate the computational robustness of the fractal dimension of the neuron axonal topography in the human brain that is derived from diffusion MRI (dMRI) acquisitions. We extend the work done in a previous paper by some of the current authors where the fractal dimension of the neuron axonal topography from dMRI data was computed from 2-D regions of interest. The fractal dimensions Df of the entire 3-D volume of the brain is here estimated via the Box Counting, the Correlation DImension and the Fractal Mass Dimension methods. 3-D neuron axon data are obtained using tractography algorithms on Diffusion Tensor Imaging of the brain. We find that all three calculations of Df give consistent results across subjects, namely, they demonstrate fractal characteristics in the short and medium length scales: different fractal exponents prevail at different length scales, an indication of multifractality. We surmise that this complexity stems as a collective property emerging when many local brain units performing different functional tasks and having different local topologies are recorded together.

This work has been published in [15].

## 6.3. Forward and Inverse Problems

### 6.3.1. Source localization using rational approximation on plane sections

**Participants:** Maureen Clerc, Théodore Papadopoulo, Juliette Leblond [Apics Project-Team, Inria, Sophia Antipolis, Méditerranée, France], Jean-Paul Marmorat [CMA, Ecole des Mines Paristech, Sophia Antipolis, France].

In functional neuroimaging, a crucial problem is to localize active sources within the brain non-invasively, from knowledge of electromagnetic measurements outside the head. Identification of point sources from boundary measurements is an ill-posed inverse problem. In the case of electroencephalography (EEG), measurements are only available at electrode positions, the number of sources is not known in advance and the medium within the head is inhomogeneous. This work presents a new method for EEG source localization, based on rational approximation techniques in the complex plane. The method is used in the context of a nested sphere head model, in combination with a cortical mapping procedure. Results on simulated data prove the applicability of the method in the context of realistic measurement configurations.

This work has been published in the journal Inverse Problems [14].

### 6.3.2. The adjoint method of OpenMEEG for EEG and MEG with large source space

**Participants:** Maureen Clerc, Théodore Papadopoulo, Alexandre Gramfort [Telecom Paristech], Emmanuel Olivi [Forner member of the Athena Project-Team].

In EEG or MEG, a lead field is the linear operator which associates unitary dipolar sources to the resulting set of sensor measurements. In practise, the source space often includes over 10 000 dipoles, which sometimes causes memory problems. The adjoint approach considers the forward problem from the viewpoint of sensors instead of sources: this drops down the number of linear systems to solve by two orders of magnitude. The adjoint approach is here proposed in the context of the Boundary Element Method, and its implementation is provided by the OpenMEEG library.

This work was presented at the BIOMAG conference [38].

### 6.3.3. Comparison of Boundary Element and Finite Element Approaches to the EEG Forward Problem

**Participants:** Maureen Clerc, Carsten Wolters [Institute for Biomagnetism and Biosignal Analysis, University of Münster], Johannes Vorwerk [Institute for Biomagnetism and Biosignal Analysis, University of Münster], Martin Burger [Institut für Numerische und Angewandte Mathematik, Fachbereich Mathematik und Informatik, Westfälische Wilhelms Universität (WWU) Münster], Jan de Munck [Vrije Universiteit Medical Centre (VUMC), The Netherlands].

The accurate simulation of the electric fields evoked by neural activity is crucial for solving the inverse problem of EEG. Nowadays, boundary element methods (BEM) are frequently applied to achieve this goal, usually relying on the simpli-fication of approximating the human head by three nested compartments with isotropic conductivities (skin, skull, brain). Here, including the highly-conducting cerebrospinal fluid (CSF) is a difficult task due to the complex geometrical structure of the CSF, demanding a high number of additional nodes for an accurate modeling and thus a strongly increased computational effort. Though, CSF conductivity is well-known and nearly not varying inter-individually and its significant influence on EEG forward simulation has been shown. The CSF can be included at negligible computational costs when applying finite element (FE) forward approaches. In this study we compare the accuracy and performance of state-of-the-art BE and FE approaches in both artificial and realistic three layer head models, showing that all approaches lead to high numerical accuracies. Furthermore, we demonstrate the significant influence of modeling the CSF compartment as disregarding this compartment leads to model errors that lie clearly above the observed numerical errors.

A book chapter on BEM and FEM models has been published in the Handbook for Neural Activity Measurement [40]. The comparison was presented at the BIOMAG conference [19].

### 6.3.4. Domain Decomposition to handle versatile conductivity models

**Participants:** Maureen Clerc, Théodore Papadopoulo, Emmanuel Olivi [Forner member of the Athena Project-Team].

Source localization from external data such EEG or MEG, requires a good understanding of the electromagnetic behavior of the patient head. Several models can been used, representing more or less complex geometrical shapes, and conductivity profiles. Different numerical methods allow to cope with different types of models: the Finite Element Method (FEM) can handle very general conductivity models, whereas the Boundary Element Method (BEM) is limited to piecewise constant conductivity. On the other hand, it is easier with BEM than with FEM to accurately represent sources in isotropic media. Thanks to domain decomposition, we propose to solve a EEG forward problem using BEM where the sources are (the brain) and FEM for other tissues (with notably inhomogeneities in the skull).

This work was presented at the BIOMAG conference [37].

## 6.4. Brain Computer Interfaces

### 6.4.1. Combining ERD and ERS features to create a system-paced BCI

**Participants:** Maureen Clerc, Théodore Papadopoulo, Joan Fruitet, Eoin Thomas.

An important factor in the usability of a brain computer interface (BCI) is the setup and calibration time required for the interface to function accurately. Recently, brain-switches based on the rebound following motor imagery of a single limb effector have been investigated as basic BCIs due to their good performance with limited electrodes, and brief training session requirements. Here, a BCI is proposed which expands the methodology of brain-switches to design an interface composed of multiple brain-buttons. The algorithm is designed as a system paced interface which can recognise 2 intentional-control tasks and a no-control state based on the activity during and following motor imagery in only 3 electroencephalogram channels. An online experiment was performed over 6 subjects to validate the algorithm, and the results show that a working BCI can be trained from a single calibration session and that the post motor imagery features are both informative and robust over multiple sessions.

This work, which was partially presented at the EMBS conference [33], is currently under revision for the Journal of Neuroscience Methods.

### 6.4.2. Bandit algorithms for faster task selection in BCI

**Participants:** Maureen Clerc, Joan Fruitet, Alexandra Carpentier [Sequel Project-Team, Inria Lille, France], Rémi Munos [Sequel Project-Team, Inria Lille, France].

BCIs based on sensorimotor rhythms use a variety of motor tasks, such as imagining moving the right or left hand, the feet or the tongue. Finding the tasks that yield best performance, specifically to each user, is a time consuming preliminary phase to a BCI experiment. This study presents a new adaptive procedure to automatically select, online, the most promising motor task for an asynchronous brain-controlled button.

We develop for this purpose an adaptive *Upper Confidence Bound* algorithm based on the stochastic bandit theory, and design an EEG experiment to test our method. We compare (offline) the adaptive algorithm to a naive selection strategy which uses uniformly distributed samples from each task. We also run the adaptive algorithm online to fully validate the approach.

By not wasting time on inefficient tasks, and focusing on the most promising ones, this algorithm results in a faster task selection and a more efficient use of the BCI training session. More precisely, the offline analysis reveals that the use of this algorithm can reduce the time needed to select the most appropriate task by almost half without loss in precision, or alternatively, allow to investigate twice the number of tasks within a similar time span. Online tests confirm that the method leads to an optimal task selection.

This study is the first one to optimize the task selection phase by an adaptive procedure. By increasing the number of tasks that can be tested in a given time span, the proposed method could contribute to reducing “BCI illiteracy”.

This work is the result of the collaboration between Sequel and Athena within the ANR CoAdapt. It has been published in NIPS [21] and is accepted in the Journal of Neural Engineering [52].

### 6.4.3. An analysis of performance evaluation for motor-imagery based BCI

**Participants:** Maureen Clerc, Matthew Dyson [Laboratoire de Neurosciences Cognitives, Aix-Marseille Université, France], Eoin Thomas.

In recent years, numerous brain computer interfaces (BCIs) have been proposed which incorporate features such as adaptive classification, error detection and correction, fusion with auxiliary signals and shared control capabilities. Due to the added complexity of such algorithms, the evaluation strategy and metrics used for analysis must be carefully chosen to accurately represent the performance of the BCI. In this article, metrics are reviewed and contrasted using both simulated examples and experimental data. Furthermore, a review of the recent literature is presented to determine how BCIs are evaluated, in particular focusing on the correlation between how the data are used relative to the BCI subcomponent under investigation. From the analysis performed in this study, valuable guidelines are presented regarding the choice of metrics and evaluation strategy dependent upon any chosen BCI paradigm.

This work was supported by the ANR Co-Adapt and is currently under revision for the Journal of Neural Engineering.

# CORTEX Project-Team

# 6. New Results

## 6.1. Spiking neurons

**Participants:** Hana Belmabrouk, Dominique Martinez, Thierry Viéville, Thomas Voegtlin.

### *6.1.1. Mathematical modeling*

In order to understand the dynamics of spiking neural networks under the influence of a modified synaptic dynamics of single neurons, we study the effect of tonic inhibition on the population activity in spiking neural networks. The aim is to derive mathematical relations of the population activity and some statistics estimated numerically from the simulation of networks [4], [8].

### *6.1.2. Biophysical modeling*

Our understanding of the computations that take place in the human brain is limited by the extreme complexity of the cortex, and by the difficulty of experimentally recording neural activities, for practical and ethical reasons. The Human Genome Project was preceded by the sequencing of smaller but complete genomes. Similarly, it is likely that future breakthroughs in neuroscience will result from the study of smaller but complete nervous systems, such as the insect brain or the rat olfactory bulb. These relatively small nervous systems exhibit general properties that are also present in humans, such as neural synchronization and network oscillations. Our goal is therefore to understand the role of these phenomena by combining biophysical modelling and experimental recordings, before we can apply this knowledge to humans. In the last year, we have studied new aspects of our models of the insect olfactory system [7], [14].

### *6.1.3. Using event-based metric for event-based neural network weight adjustment*

The problem of adjusting the parameters of an event-based network model is addressed here at the programmatic level. Considering temporal processing, the goal is to adjust the network units weights so that the outcoming events correspond to what is desired. The work of [18] proposes, in the deterministic and discrete case, a way to adapt usual alignment metrics in order to derive suitable adjustment rules. At the numerical level, the stability and unbiasness of the method is verified.

The key point, here, is the non-learnability of even-based , since it is proved that this problem is NP-complete, when considering the estimation of both weights in the general case, except for exact simulation. We show that we can “elude” this caveat and propose an alternate efficient estimation mechanism, inspired by alignment metrics used in spike train analysis, thus providing a complement of other estimation approaches, beyond usual convolution metric. At last, the proposed mollification *is* a series of convolution metric, but that converges towards the expected alignment metric.

### *6.1.4. Predictive learning*

In collaboration with Sander Bohte (CWI, Netherlands) and Nicolas Fourcaud-Trocme (CNRS, Lyon), we are developing a model of predictive learning using oscillations in a population of spiking neurons. The model is based on previous work performed in the Cortex group. Our previous model suggested a possible role for neuronal synchronization in unsupervised, predictive-type learning. However, that model was not compatible with sustained oscillations observed in biological networks. We are extending our initial approach in order to allow the network to learn during a stable, steady-state oscillatory regime. This extension involves using type-2 neurons and two distinct types of inhibition.

## 6.2. Dynamic Neural Fields

**Participants:** Frédéric Alexandre, Yann Boniface, Laurent Bougrain, Georgios Detorakis, Hervé Frezza-Buet, Bernard Girau, Axel Hutt, Mathieu Lefort, Nicolas Rougier, Wahiba Taouali.

The work reported this year represents both extensions of previous works and new results linked to the notion of neural population, considered at (i) a formal level (theoretical studies of neural fields), (ii) a numerical level (study of functioning and learning rules) and (iii) a more embodied one (implementations of specific functions).

### *6.2.1. Formal Level*

To study the effect of external stimuli on nonlinear neural population dynamics involving constant delays, the work aims to apply the center manifold theorem and derive expressions of the time-dependent centre manifold. It is observed that additive noise and external quasi-periodic driving change the stability of neural populations dependent on the delay [9], [10].

### *6.2.2. Numerical Level*

At the numerical level, specific developments were carried out to assess our software platform, to master functioning rules and to study the performances of new learning rules:

- Adaptation of the BCM rule to multi-modality by adapting the dynamics of the threshold by the use of a feed-back signal generated by a neural field map [1], [26]
- We investigate the formation and maintenance of ordered topographic maps in the primary somatosensory cortex as well as the reorganization of representations after sensory deprivation or cortical lesion. We consider both the critical period (postnatal) where representations are shaped and the post-critical period where representations are maintained and possibly reorganized. We hypothesize that feed-forward thalamocortical connections are an adequate site of plasticity while cortico-cortical connections are believed to drive a competitive mechanism that is critical for learning. We model a small skin patch located on the distal phalangeal surface of a digit as a set of 256 Merkel ending complexes (MEC) that feed a computational model of the primary somatosensory cortex (area 3b). This model is a two-dimensional neural field where spatially localized solutions (a.k.a. bumps) drive cortical plasticity through a Hebbian-like learning rule. Simulations explain the initial formation of ordered representations following repetitive and random stimulations of the skin patch. Skin lesions as well as cortical lesions are also studied and results confirm the possibility to reorganize representations using the same learning rule and depending on the type of the lesion. For severe lesions, the model suggests that cortico-cortical connections may play an important role in complete recovery [6].

### *6.2.3. Embodied Level*

#### *6.2.3.1. Motion detection*

We develop bio-inspired neural architectures to extract and segment the direction and speed components of the optical flow from sequences of images. Following this line, we have built additional models to code and distinguish different visual sequences. The structure of these models takes inspiration from the course of visual movement processing in the human brain, such as in area MT (middle temporal) that detects patterns of movement, or area FBA where neurons have been found to be sensitive to single spatio-temporal patterns. This work has been extended to complex movements: to fight, to wave, to clap, using real-world video databases [5].

#### *6.2.3.2. Anticipatory mechanisms in neural fields*

We have defined first models of neural fields that include anticipatory mechanisms through the integration of spatiotemporal representations into the lateral interactions of a dynamic neural field. In [20], the case of multiple anticipated trajectories is studied.

#### 6.2.3.3. Action selection

Within the context of enaction and a global approach to perception, we focused on the characteristics of neural computation necessary to understand the relationship between structures in the brain and their functions. We first considered computational problems related to the discretization of differential equations that govern the studied systems and the synchronous and asynchronous evaluation schemes. Then, we investigated a basic functional level : the transformation of spatial sensory representations into temporal motor actions within the visual-motor system. We focused on the visual flow from the retina to the superior colliculus to propose a minimalist model of automatic encoding of saccades to visual targets. This model, based on simple local rules (CNFT and logarithmic projection) in a homogeneous population and using a sequential processing, reproduces and explains several results of biological experiments. It is then considered as a robust and efficient basic model. Finally, we investigated a more general functional level by proposing a computational model of the basal ganglia motor loop. This model integrates sensory, motor and motivational flows to perform a global decision based on local assessments. We implemented an adaptive process for action selection and context encoding through an innovative mechanism that allows to form the basic circuit for other cortico-basal loops. This mechanism allows to create internal representations according to the enactive approach that opposes the computer metaphor of the brain. Both models have interesting dynamics to study from whether a biological point of view or a computational numerical one [2], [12].

## 6.3. Higher level functions

**Participants:** Frédéric Alexandre, Laurent Bougrain, Octave Boussaton, Axel Hutt, Maxime Rio, Carolina Saavedra, Christian Weber.

Our activities concerned information analysis and interpretation and the design of numerical distributed and adaptive algorithms in interaction with biology and medical science. To better understand cortical signals, we choose a top-down approach for which data analysis techniques extract properties of underlying neural activity. To this end several unsupervised methods and supervised methods are investigated and integrated to extract features in measured brain signals. More specifically, we worked on Brain Computer Interfaces (BCI).

### *6.3.1. Using Neuronal States for Transcribing Cortical Activity into Muscular Effort*

We studied the relations between the activity of corticomotoneuronal (CM) cells and the forces exerted by fingers. The activity of CM cells, located in the primary motor cortex is recorded in the thumb and index fingers area of a monkey. The activity of the fingers is recorded as they press two levers. The main idea of this work is to establish and use a collection of neuronal states. At any time, the neuronal state is defined by the firing rates of the recorded neurons. We assume that any such neuronal state is related to a typical variation (or absence of variation) in the muscular effort. Our forecasting model uses a linear combination of the firing rates, some synchrony information between spike trains and averaged variations of the positions of the levers [17].

### *6.3.2. From the decoding of cortical activities to the control of a JACO robotic arm: a whole processing chain*

We realized a complete processing chain for decoding intracranial data recorded in the cortex of a monkey and replicates the associated movements on a JACO robotic arm by Kinova. We developed specific modules inside the OpenViBE platform in order to build a Brain-Machine Interface able to read the data, compute the position of the robotic finger and send this position to the robotic arm. More precisely, two client/server protocols have been tested to transfer the finger positions: VRPN and a light protocol based on TCP/IP sockets. According to the requested finger position, the server calls the associated functions of an API by Kinova to move the fingers properly. Finally, we monitor the gap between the requested and actual fingers positions. This chain can be generalized to any movement of the arm or wrist [22].

### *6.3.3. Wavelet-based Semblance for P300 Single-trial Detection*

Electroencephalographic signals are usually contaminated by noise and artifacts making difficult to detect Event-Related Potential (ERP), specially in single trials. Wavelet denoising has been successfully applied to ERP detection, but usually works using channels information independently. This paper presents a new adaptive approach to denoise signals taking into account channels correlation in the wavelet domain. Moreover, we combined phase and amplitude information in the wavelet domain to automatically select a temporal window which increases class separability. Results on a classic Brain-Computer Interface application to spell characters using P300 detection show that our algorithm has a better accuracy with respect to the VisuShrink wavelet technique and XDAWN algorithm among 22 healthy subjects, and a better regularity than XDAWN [21].

### *6.3.4. Filter for P300 detection*

According to recent literature, the most appropriate preprocessing to improve P300 detection is still unknown or at least there is no consensus about it. Research papers refer to different low-pass filters, high-pass filters, baseline, subsampling or feature selection. Using a database with 23 healthy subjects we compared the effect on the letter accuracy (single-trial detection) provided by a linear support vector machine of a high-pass filter with cutoff frequencies from 0.1 to 1 Hz and a low-pass filter with cutoff frequencies from 8 to 60 Hz. According to this study, the best combination is for a band-pass filter of 0.1 to 15 Hz [16].

### *6.3.5. Processing Stages of Visual Stimuli and Event-Related Potentials*

Event-evoked potentials (ERP) in electroencephalograms reflect various visual processing stages according to their latencies and locations. Thus, ERP components such as the N100, N170 and the N200 which appears 100, 170 and 200 ms after the onset of a visual stimulus correspond respectively to a selective attention, the processing of color, shape and rotation (e.g. processing of human faces) and a degree of attention [24].

### *6.3.6. Exploring the role of the thalamus in visuomotor tasks implicating non-standard ganglion cells*

Non-standard ganglion cells in the retina have specific loci of projection in the visuomotor systems and particularly in the thalamus and the superior colliculus. In the thalamus, they feed the konio pathway of the LGN. Exploring the specificities of that pathway, we discovered it could be associated to the matrix system of thalamo-cortical projections, known to allow for diffuse patterns of connectivity and to play a major role in the synchronization of cortical regions by the thalamus.

An early model [23] led to the design of the corresponding information flows in the thalamo-cortical system, that we are now expanding, in the framework of the Keops project § 7.2 , to be applied to real visuomotor tasks.

### *6.3.7. Formalization of input/output retinal transformation regarding non-standard gangion cells behavior*

We propose to implement the computational principles raised by the study on the K-cells of the retina using a variational specification of the visual front-end, with an important consequence: In such a framework, the GC are not to be considered individually, but as a network, yielding a mesoscopic view of the retinal process.

Given natural image sequences, fast event-detection properties appears to be exhibited by the mesoscopic collective non-standard behavior of a subclass of the so-called dorsal and ventral konio-cells (K-cells) that correspond to specific retinal output.

We consider this visual event detection mechanism to be based on image segmentation and specific natural statistical recognition, including temporal pattern recognition, yielding fast region categorization. We discuss how such sophisticated functionalities could be implemented in the biological tissues as a unique generic two-layered non-linear filtering mechanism with feedback. We use computer vision methods to propose an effective link between the observed functions and their possible implementation in the retinal network.

The available computational architecture is a two-layers network with non-separable local spatio-temporal convolution as input, and recurrent connections performing non-linear diffusion before prototype based visual event detection.
The numerical robustness of the proposed model has been experimentaly checked on real natural images. Finally, model predictions to be verified at the biological level are discussed [25].

## 6.4. Embodied and embedded systems

**Participants:** Yann Boniface, Hervé Frezza-Buet, Bernard Girau, Mathieu Lefort.

### *6.4.1. InterCell*

Our research in the field of dedicated architectures and connectionist parallelism mostly focuses on embedded systems (*cf.* §3.5 ). Nevertheless we are also involved in a project that considers coarse-grain parallel machines as implementation devices. The core idea of this InterCell project (*cf.*http://intercell.metz.supelec.fr) is to map fine grain computation (cells) to the actual structure of PC clusters. The latter rather fit coarse grain processing, using relatively few packed communication, which a priori contradicts neural computing. Another fundamental feature of the InterCell project is to promote interaction between the parallel process and the external world. Both features, cellular computing and interaction, allow to consider the use of neural architectures on the cluster on-line, for the control of situated systems, as robots.

### *6.4.2. Hardware implementations of neural models*

In the field of dedicated embeddable neural implementations, we use our expertise in both neural networks and FPGAs so as to propose efficient implementations of applied neural networks on FPGAs, as well as to define hardware-friendly neural models.

- We currently intend to minimize the topological constraints of FPGA-embedded spiking neural fields using reduced neighborhoods but randomly propagating spikes. A preliminary result has been obtained so as to implement massively distributed pseudo-random number generators based on cellular automata that use minimal areas though they produce random streams that pass most randomness tests [19]. These results have also been applied to cellular aurtomata using randomnes in their transition rules [13].
- Researchers have proposed the concept of Central Pattern Generators (CPGs) as a neural mechanism for generating an efficient control strategy for legged robots based on biological locomotion principles. We have developed a reconfigurable hardware implementation of a CPG-based controller which is able to generate several gaits for quadruped and hexapod robots [3].

### *6.4.3. Towards brain-inspired hardware*

Our activities on dedicated architectures have strongly evolved in the last years. We now focus on the definition of brain-inspired hardware-adapted frameworks of neural computation. Our current works aim at defining hardware-compatible protocols to assemble various perception-action modalities that are implemented and associated by different bio-inspired neural maps.

#### *6.4.3.1. Multimodal learning through joint dynamic neural fields*

This work relates to the development of a coherent multimodal learning for a system with multiple sensory inputs. We have modified the BCM synaptic rule, a local learning rule, to obtain the self organization of our neuronal inputs maps and we use a CNFT based competition to drive the BCM rule. In practice, we introduce a feedback modulation of the learning rule, representing multimodal constraints of the environment, and we introduce an unler-arning term in the BCM equation to solve the problem of the different temporalities between the raise of the activity within modal maps and the multimodal learning of the organization of the maps [1], [26].

#### *6.4.3.2. Randomly spiking dynamic neural fields*

We have defined a new kind of spiking neural field that is able to use only local links while transmitting spikes through the map by succesive random propagations. Such a model is able to be mapped onto FPGAs, while maintaining most properties of neural fields. Early results will be soon published.

# DEMAR Project-Team

# 6. New Results

## 6.1. Modelling and Identification

### 6.1.1. Subject-specific Center of Mass estimation in human subjects

**Participants:** Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

Center of mass position (CoM) in humanoid robots can be used to generate a joint trajectory suitable for walking and standing. Oscillations of the CoM while maintaining a standing posture have been observed in older patients. These oscillations are thought to occur due a change in the subject's balancing strategy. This is why the motion of the CoM in generally considered as good metric to be used while diagnosing pathologies which affect gait. With this in mind we propose the use of the statically equivalent serial chain (SESC) to provide a subject specific estimate of CoM position. The state of the art techniques for CoM estimation in humans involve the use of expensive equipment in a laboratory setting, making it difficult to use as a clinical tool or inside the home environment. Current work for diagnosing a subject's balance moves away from this, focusing on wearable and minimally invasive sensors that obtain information during the subject's daily activities. We propose the use of widely available sensors like the Kinect camera, for tracking the subject's movements, and the Wii balance board during the calibration of the SESC.

a) b)

*Figure 1. Center of mass estimation can be done using the statically equivalent serial chain. We assume a 9 rigid-link model (a) with spherical joints, capable of three dimensional movements. The SESC estimation can be performed in real time and is driven by joint angular measurements (b).*

We have focused in improving the reliability of the identification during a study of the CoM trajectory in the sagital plane [14]. We studies small value for the condition number of the used data as well as of the parameter relative standard deviation ($\sigma_{\widehat{\mathbf{R}}_r}\%$) are useful to determine the validity of the estimate. Subsequent works have extended the human model to three-dimensional motion (Fig. 1 .a). In order to observe the tracking of the center of mass, we have developed a 3D visualization tool which represents the subject's skeleton and SESC in real time (Fig. 1 .b) [16]; it is also possible to observe the CoM history. Finally we are now comparing the performance of the inexpensive Kinect sensor and the traditional video based motion capture Vicon system, with reasonably good results [15]. Current work is also focused in improving the speed of the identification phase by simplifying the assumed human model and using physical constraints to reduce the complexity of the SESC. Also the development of a simple to follow identification protocol which takes into account multiple supporting surfaces is desired.

### 6.1.2. Recursive estimation of SESC parameters for human Center of Mass estimation

**Participants:** Alejandro González, Mitsuhiro Hayashibe, Philippe Fraisse.

A human's center of mass (CoM) trajectory is useful to evaluate the dynamic stability during daily life activities such as walking and standing up. To estimate the subject-specific CoM position in the home environment, we make use of a statically equivalent serial chain (SESC) developed with a portable mesurement system. In order to adapt to the subject's physical capacities we implement a constrained Kalman filter to achieve an online parameter estimation of the SESC parameters. By accounting for the human body bilateral symmetry we hope to reduce the identification time. This results in constrainig SESC parameters to be consistent with the human skeletal model used. The Kinect camera is used as a markerless motion capture system for measuring limb orientations while the Wii board is used to measure the subject's center of pressure (CoP) during the identification phase. For his experiment CoP measurements and Kinect data were recorded for five able-bodied subjects. The data was then given to the proposed recursive algorithm to identify the parameters of the SESC online (Fig.2 ). This method of online identification allows the subject or the therapist to know the quality of the on-going CoM identification while giving postures; by doing this the time needed to perform the identification can be reduced. A cross-validation was performed to verify the identification performance.

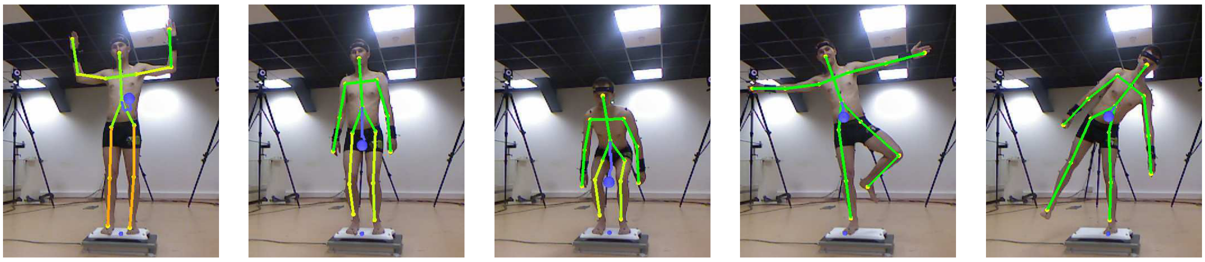


*Figure 2. Online statically equivalent serial chain (SESC) parameter estimation. The length of each of the SESC's link is updated when a static pose has been found and can be observed in real time. The color of the skeleton can be updated as a cue to the subject and/or therapist.*

### 6.1.3. FES-Induced Torque Prediction with Evoked EMG Synthesized by Recurrent Neural Network

**Participants:** Zhan Li, Mitsuhiro Hayashibe, David Guiraud.

A NARX-type recurrent neural network (NARX-RNN) model is proposed for identification and prediction of FES-induced muscular dynamics with eEMG. Such NARX-RNN model is with a novel architecture for prediction, with robust prediction performance. To make fast convergence for identification of such NARXRNN, directly-learning pattern is exploited during the learning phase. Due to difficulty of choosing a proper forgetting factor of Kalman filter for predicting time-variant torque with eEMG, such NARX-RNN may be considered to be a better alternative as torque predictor. Data gathered from two SCI patients is used to evaluate the proposed NARX-RNN model. The NARX-RNN model shows promising estimation and prediction performance only based on eEMG [23].

### 6.1.4. Inverse Estimation of Muscle Activations with Weights Optimization

**Participants:** Zhan Li, Mitsuhiro Hayashibe, David Guiraud.

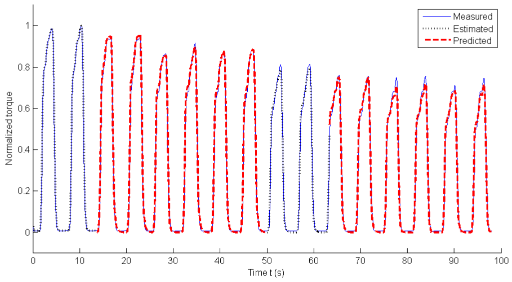


*Figure 3. Performance of NARX-RNN model with eEMG for prediction with identification on periodically*

Inverse estimation of activations of muscle groups at human lower leg in random movement condition is investigated with merely the ankle joint torque used. Optimization technique for the relationship between muscle activations and torque is exploited. Such optimization is able to rebuilt the relationship between muscle activations and torque inversely based on experimental data obtained from five healthy subjects, and the optimal weight matrix can indicate each muscle's contribution for producing the torque. Further cross-validation on prediction of muscle activations with joint torque with optimal weights shows such approach may possess promising performance [22].

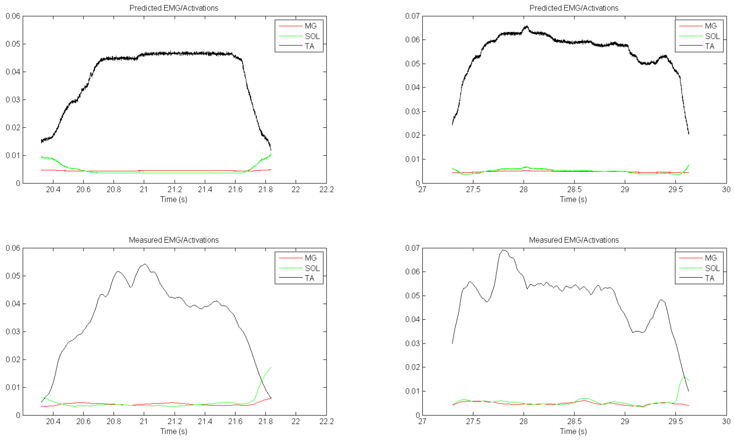


*Figure 4. Prediction of muscle activations*

### 6.1.5. 3D Volumetric Muscle Modeling For Real-time Deformation Analysis With FEM

**Participants:** Yacine Berranen, Mitsuhiro Hayashibe, Benjamin Gilles, David Guiraud.

Computer simulators are promising numerical tools to study muscle volumetric deformations but most models are facing very long computation time and thus are based on simplified Hill model versions. The purpose of this study is to develop a real-time three-dimensional biomechanical model of fusiform muscle based on modified Hill model for the active stress which is controlled from EMG recordings. Finite element model is used to estimate the passive behavior of the muscle and tendons during contraction. We show that this 3D model implementation is very cost effective with respect to the computation time and the simulation gives good results compared to real measured data. Thus, this effective implementation will allow implementing much more complex and realistic models, with moderate computation time.

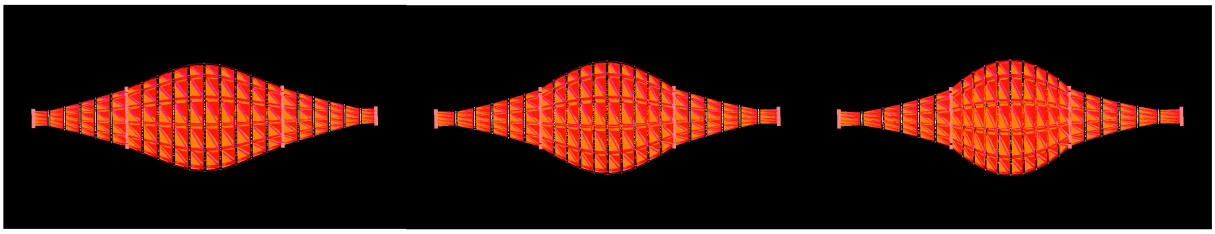


*Figure 5. Real-time Simulation of Volumetric Muscle Deformation.*

### 6.1.6. Principal Geodesic Dynamics

**Participants:** Maxime Tournier, Lionel Reveret.

This paper presents a new integration of a data-driven approach using dimension reduction and a physically-based simulation for real-time character animation. We exploit Lie group statistical analysis techniques (Principal Geodesic Analysis, PGA) to approximate the pose manifold of a motion capture sequence by a reduced set of pose geodesics. We integrate this kinematic parametrization into a physically-based animation approach of virtual characters, by using the PGA-reduced parametrization directly as generalized coordinates of a Lagrangian formulation of mechanics. In order to achieve real-time without sacrificing stability, we derive an explicit time integrator by approximating existing variational integrators. Finally, we test our approach in task-space motion control. By formulating both physical simulation and inverse kinematics time stepping schemes as two quadratic programs, we propose a features-based control algorithm that interpolates between the two metrics. This allows for an intuitive trade-off between realistic physical simulation and controllable kinematic manipulation.

### 6.1.7. An improved kinematic model of the spine for three-dimensional motion analysis in the Vicon system

**Participants:** Pawel Maciejasz, Wieslaw Chwala (University School of Physical Education, Krakow), Miroslawa Dlugosz, Daria Panek, Witold Alda (AGH University of Science and Technology, Krakow).

The mechanism of creation and pathomechanics of lateral spinal deformation is still not fully explained. Modern medical imaging techniques give scientists possibility to understand some aspects, but vast majority of those techniques is based on static trials. A motion capture system belongs to techniques which enable visualization of a spine during dynamic trials; however, due to lack of appropriate computational model, it is unsuitable for scoliosis imaging.

A few years ago a kinematic model of the spine has been proposed to be used with Vicon Motion Capture System (Master thesis of P. Maciejasz). This model was based on Bézier curves and allowed for much more precise investigation of spinal kinematics during dynamic trials as compared with other computational models. However, it did not allowed to restrict only selected movements for particular segments of the spine (e.g. axial rotation for lumbar spine). The aim of the current work is to improve the proposed model in order to be able to restrict selected movements according to the knowledge concerning spinal anatomy and spinal range of motion. The new kinematic model of the spine was written in BodyBuilder for Biomechanics Language.

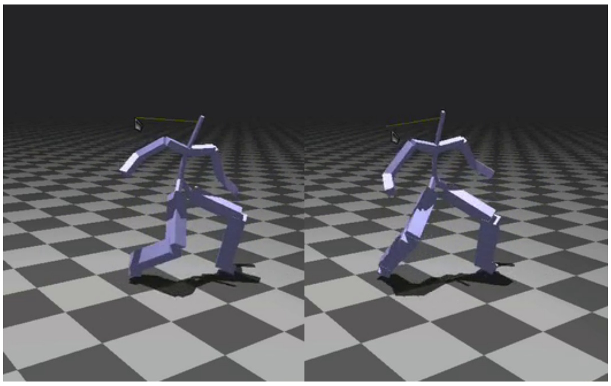


*Figure 6. Real-time user interaction with a virtual character maintaining balance, animated using our approach: based on a break-dance motion capture sequence, our dimension reduction process allows one to compute both dynamics and task-space control in a low-dimensional, data-driven subspace.*

For the purpose of visualization also an accurate graphical representation of each vertebra (polygon mesh) was computed and adapted to be compatible with the kinematic model. Using a new version of the model it is possible to perform precise analysis of movement of all vertebrae during such dynamic activities as e.g. gait and forward or lateral bending, as well as to present the results not only on the charts, but also as a 3D animation of movements of a realistically looking spine [12].

### 6.1.8. Methodology of automated detection and classification of action potentials in nerve fiber based on multichannel recordings

**Participants:** Thomas Guiho, Pawel Maciejasz, David Guiraud.

For some neuroprosthetic applications it would be beneficial to be able to automatically detect when particular nerve fibers (axons) are in "firing" (i.e. when an action potential is propagating along them). Due to limitations of currently available recording electrodes on one side, and the attempt to be as little invasive as possible, at the moment in practical application it is not possible to record signals coming from a single nerve fiber. In signal recorded using typical electrodes placed close to nerve fibers, action potentials coming from various nerve fibers, as well as noise coming from outside of the nerve, may be detected. One of the possibilities allowing to distinguish action potentials coming from various nerve fibers in such a case is to record signals at a few places along the nerve and compare them.

We have proposed an algorithm that allows to automatically detect and classify evoked action potentials in a simple earthworm model. The signals were recorded concurrently at 2 places along the giant nerve fibers. In the first step the algorithm tries to identify only does recordings in which action potentials generated by various nerve fibers can be easily distinguished. Afterwards, the most significant features (such as amplitude, duration, propagation velocity, etc.) to distinguish between different populations of fibers are identified. Finally, the action potentials in all signals are identified and classified using the features determined in the previous step.

The proposed method was implemented using MATLAB software and tested on the file containing almost 200 signals record in response to various stimuli. The same data ware later inspected manually and the action potentials were manually classified. More than 99% of action potentials were classified to the same nerve fiber when performing automatic and manual classification.

## 6.2. Function control and synthesis

### 6.2.1. FES assisted sitting pivot transfer

**Participants:** Jovana Jovic, Christine Azevedo Coste, Philippe Fraisse, Sebastien Lengagne, Charles Fattal.

Transferring from a wheelchair to a treatment table, bed, tub/shower bench, toilet seat, car seat and vice versa represent typical Sitting Pivot Transfer (SPT) realized by individuals with Spinal Cord Injury (SCI). Individuals with SCI, perform this postural task around fifteen times a day using upper extremities. In the chronic stage after SCI, soft tissue structures are exposed to overuse in activities of daily living, such as, transfer task in which the shoulder becomes a weight-bearing joint. Therefore, the risk of shoulder pain and musculoskeletal disorders is higher in persons with paraplegia compared with an able-bodied population. A lot of scientific effort has been focused on experimental studies in which the kinetic and the kinematic of the SPT movement have been analyzed. To our best knowledge, the scientists have focused their attention only on the performance of SPT; the influence of Functional Electrical Stimulation (FES) on SPT maneuver has not been investigated so far.

Therefore, we investigate the influence of FES on SPT motion of a paraplegic person. First, we develop dynamic optimization method in order to predict SPT motion of an able-bodied subject. This approach have been validated by comparing computed SPT trajectories with the ones measured during the experiment with an able-bodied subject (see Fig. 7 ). After validating our method, we used the optimization tool for analyzing the influence of FES on SPT maneuver in paraplegic persons. Our results suggest that FES can decrease arm participations during the transfer motion of a paraplegic person. [6], [21].

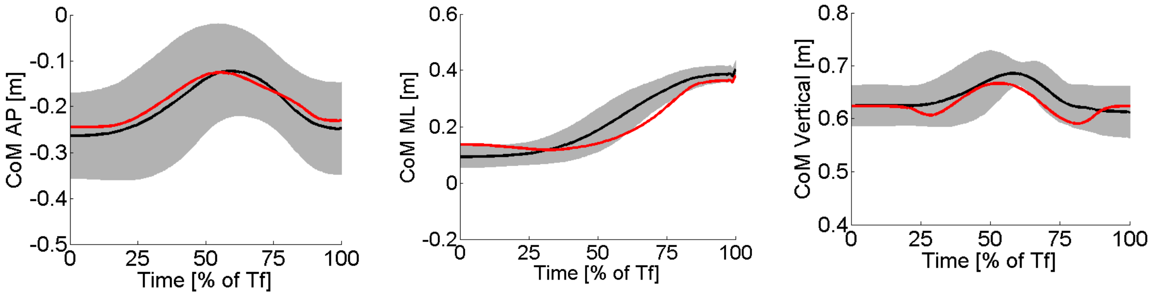


*Figure 7. Computed (red line) CoM position, mean value of CoM positions estimated from experimental data (black line) and its plus/minus standard deviation (gray line) in AP, ML, and vertical direction in able-bodied subject.*

### 6.2.2. FES assisted prolonged standing

**Participants:** Jovana Jovic, Philippe Fraisse, Christine Azevedo Coste, Charles Fattal.

Prolonged immobilization which occurs after spinal cord injury results in many physiological problems. Standing therapy can ameliorate many of those problems. The approaches proposed in the literature for restoration of standing in paraplegic population based on Functional Electrical Stimulation (FES) focuse on the control of each individual joint, i.e. joint space control. In those cases the balance of the postural system is not directly controlled. This could be problematic especially when only the lower limbs are controlled. During paraplegic's quiet standing two concurrent controllers are acting in parallel, the physiological system under control of Central Nervous System (CNS), and artificial FES system. Upper part of the paraplegic's body is under voluntary control, therefore artificial controllers should be designed in the way to take into account actions of the intact part of the body and to assist users in their task.

For human beings the Center of Mass (CoM) provides an indicator of stability and it is an essential parameter in human postural stability. By controlling CoM position in paraplegic person the voluntary motions under CNS control are taken into account. Therefore, in we propose a whole body controller based on control of the CoM position. The goal was to develop a simple balance controller which would, by means of FES, enable quiet standing of individuals suffering from SCI while taking into account the voluntary motion of the upper limbs. The controller should enable prolonged standing by simulating the behavior of an able-bodied subject during the standing task, i.e. by imposing posture switching and in that way allowing the stimulated muscles to relax. The proposed approach is based on a 10 DoF biomechanical model and Proportioan Integral (PI) controller (see Fig. 8 ). The validity of the approach is tested, in computer simulations, using human CoM trajectories estimated from experimental data and by applying perturbations in simulation during quiet standing in order to simulate voluntary upper body movements.The results show that proposed controller is able to track desired CoM position with sufficient precision and to maintain stability even in the presence of simulated movements of the upper body [20].

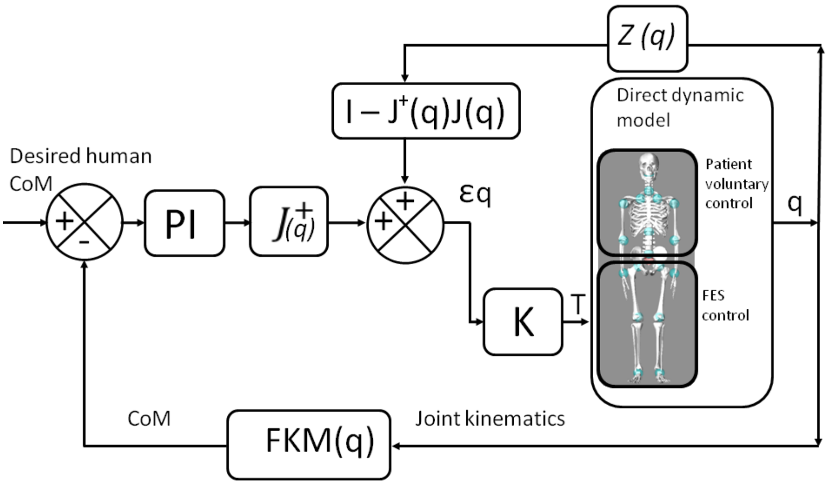


*Figure 8. Block diagram of the proposed postural controller. Controller follows desired 3D CoM positions and controls the lower limbs by applying torque at ankle, knee and hip joints (10 DoF).*

### 6.2.3. Bimanual reeducation assisted by FES in post-stroke patients

**Participants:** Nader Rouis, Christine Azevedo Coste, Philippe Fraisse, Isabelle Laffont, Denis Mottet.

This starting project will investigate the possibility to stimulate the deficient upper limb of a post-stroke patient in order to reproduce the movements observed on the valid upper limb or in order to achieve a bimanual task in cooperation with the valid upper limb. The aim is to improve the bimanual training tasks classically used in fictional rehabilitation. Both embedded sensors and kinect type systems will be investigated as possible ways to observe the valid upper limbs.

### 6.2.4. Freezing detection in Parkinson Disease patients

**Participants:** Maud Pasquier, Christine Azevedo Coste, Christian Geny, Bernard Espiau.

This work intends to apply the results of Maud Pasquier thesis about data segmentation and locomotion analysis to the detection as soon as possible of freezing episodes in Parkinson Disease (PD) patients. PD is a chronic degenerative disease of the central nervous system. One of the consequence is walking troubles and increased postural instability and falling risks. Freezing concerns at least half of PD patients, it is characterized by the transitory incapacity to make a step. It classically occurs at the gait initiation, turn around and passing doors. This freezing of gait (FOG) strongly impacts patient's mobility. As an example, in figure 9 , data recorded by an inertial sensor placed at the ankle is presented.

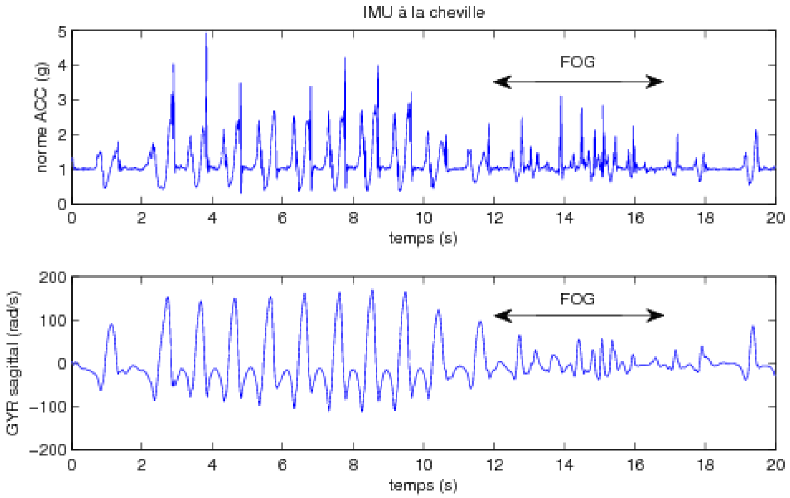


*Figure 9. Signals recorded with an inertial sensor placed at the ankle during a walking trial. Top: acceleration norm, Bottom: sagittal plane rotation velocity*

Assistive devices have been proposed, it has been shown that providing the patient with an auditory metronome or visual lines on the ground allows to reduce FOG occurrence but the effects are not maintained within the time. An approach would be to present this signals only when a freezing episode occurs and this implies to be able to detect it. Several authors have shown that FOG are in general associated to walking rhythm variability. Moore et al have used an accelerometer placed on lower limv in order to detect the presence of high frequencies. Indeed, the tremor pattern which can be observed during freezing is located at 3 et 8Hz, whereas normal locomotion is around 3Hz. These authors are able to detect a large part of the FOG but only those presenting high frequency patterns which is not systematic. Furthermore the detection delay (FFT) is very high and cannot be compatible with assistive device control constraints. We have proposed to observe stride properties in an online manner and compute a criterion which value informs about the FOG occurrence. The criterion is based on two variables: the stride length and the cadence. When stride length diminishes and cadence increases a FOG may be upcoming. 3 patients have been involved in this study and the criterion proposed has been shown to be as efficient than Moore's method in terms of number of detected FOG but the detection time is strongly improve with our method.

### 6.2.5. "Awake surgery" of slow-growing tumors and cortical excitability measured by EEG recordings.

**Participants:** François Bonnetblanc, Guillaume Herbet, Pom Charras, Mitsuhiro Hayashibe, David Guiraud, Hugues Duffau, Bénédicte Poulin-Charronnat.

Using direct electrical stimulation, real-time functional mapping of the brain can be used to perform resections of slow-growing infiltrative tumors in awake patients and to prevent the resection of essential areas near the tumor. To investigate interhemispheric imbalance following " awake surgeries" of slow-growing tumors we recorded EEG in a visuo- manual RT paradigm. Increase of cortical excitability within the ipsilesional hemisphere was signed by increased event related potentials (ERPs) amplitude for two patients. The cortical excitability in the lesioned hemisphere may be increased to maintain performances and cerebral plasticity.

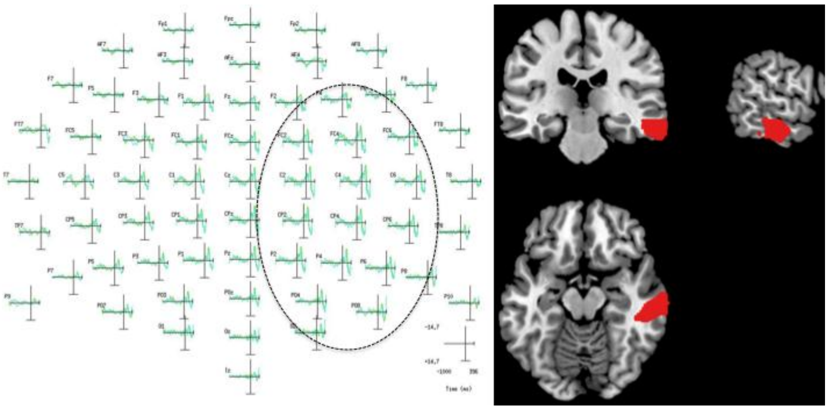


*Figure 10. Left : Post-operative Event-Related Potentials (ERPs) for Patient 1. Increased ERPs amplitude can be seen in the right ipsilesional hemisphere (ellipse) in comparison to the contralesional hemisphere and homologous recording sites. The patient had to respond to visual stimuli occurring in the right or left hemifield with his right hand. The vertical line on each ERP indicates the occurrence of the visual go- signal Right : Lesion mapping for the same patient after the surgery.*

## 6.3. Neuroprostheses

### *6.3.1. Distributed Measurement Unit for Closed-Loop Functional Electrical Stimulation: Prototype for Muscular Activity Detection*

**Participants:** Guillaume Coppey, David Andreu, David Guiraud.

One way to face centralized Functional Electrical Stimulation (FES) architecture limitations is to distribute electronics close to electrodes. These Distributed Stimulation and Measurement Units (DSU and DMU) are interconnected by a network. Different DSU have been designed and prototyped. We started the design and prototyping of a DMU dedicated to ElectroMyoGramm (EMG) activity reading.

To validate both the digital architecture of the DMU and the digital processing it performs, we prototyped a DMU in charge of muscular activity detection. This DMU is able to detect a threshold crossing on an EMG input signal. The experimental setup is schematically represented on figure 11 , showing also the digital architecture of the prototyped DMU.

This DMU is able to accurately detect EMG activity after filtering and then processing a rolling average. Figure 12 shows intermediate signals: (a) is the absolute value of the filtered EMG signal, (b) is the rolling average, and (c) is the threshold used for activity detection.

This DMU prototype showed that digital processing chain dedicated to EMG activity detection can be embedded within a distributed measurement unit using a programmable logical device (FPGA), like we did for distributed stimulation unit. The embedded architecture of this unit is designed according to a Petri Net based methodology. This allows to exploit effective parallelism offered by FPGA devices, and to reach expected

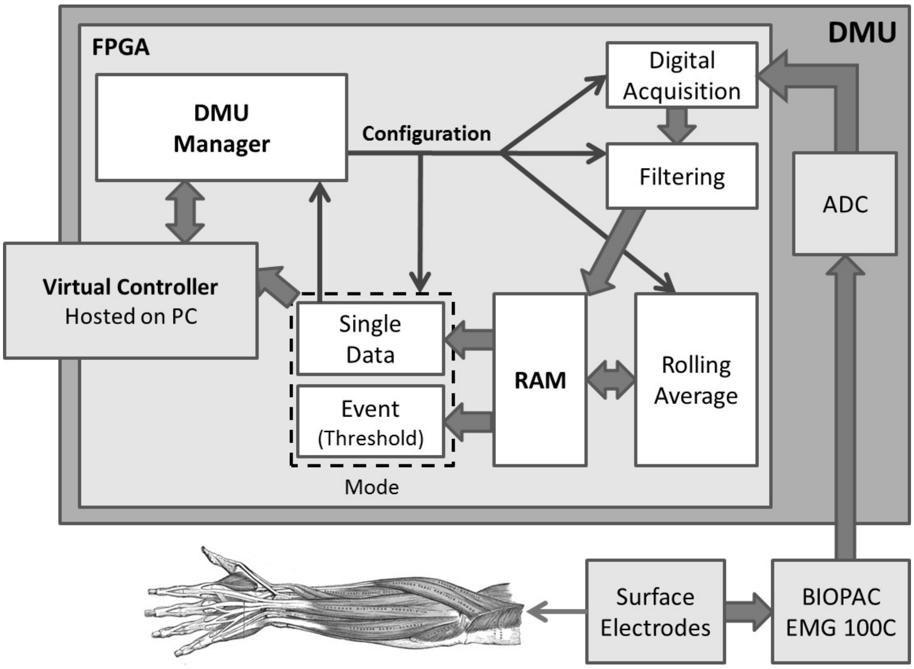


Figure 11. Experimental setup

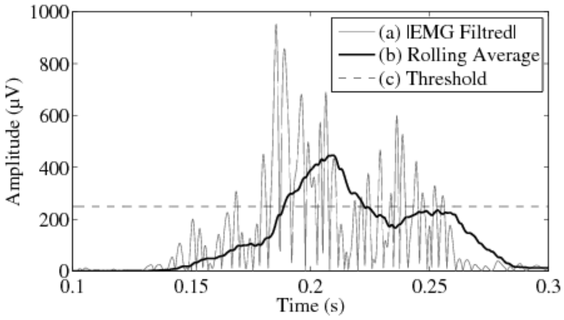


Figure 12. Intermediate signals for muscular activity detection on extensor digitorum communis

performances even at low frequency. The embedded processing chain is configurable and parameters can be adjusted, in order to optimize performance.

Future works will consist in adding the protocol stack to the digital architecture of the DMU, allowing integrating it within our distributed FES architecture. This will allow us to measure effective latencies and other performances from a closed-loop point of view. This work is necessary to ensure that such a distributed EMG activity detection is adequate with FES requirements. After that, we will investigate the trade-off between the global performances versus the implantable device constraints, like its size and power consumption.

### 6.3.2. Abstraction and composition for formal design of neuroprotheses

**Participants:** Hélène Leroux, David Andreu, Karen Godary.

In the framework of specification and implementation of complex digital systems on FPGA, we have developped an approach based on components whose behavior and composition are specified by generalized interpreted T-time Petri nets. One of the inherent difficulties for designer is to take into account, on the behavioral part, exceptions. This leads often to a complex modeling and is a source of human errors. Indeed, it is intricate to express all the possible situations (i.e. current state of model). We have defined a way to model exception handling by integrating the well-know concept of macroplace into the formalism. The analysability of the model and the efficiency of the implementation on FPGA (reactivity and surface, ie number of logic blocks) have been preserved. An example of macroplace is given in figure 13 ; it contains a sub-net (set of places of its refinement) from which exception handling is simply described by a dedicated output transition (transition $t_e$ on fig. 13 ), whatever is the current state of the sub-net.

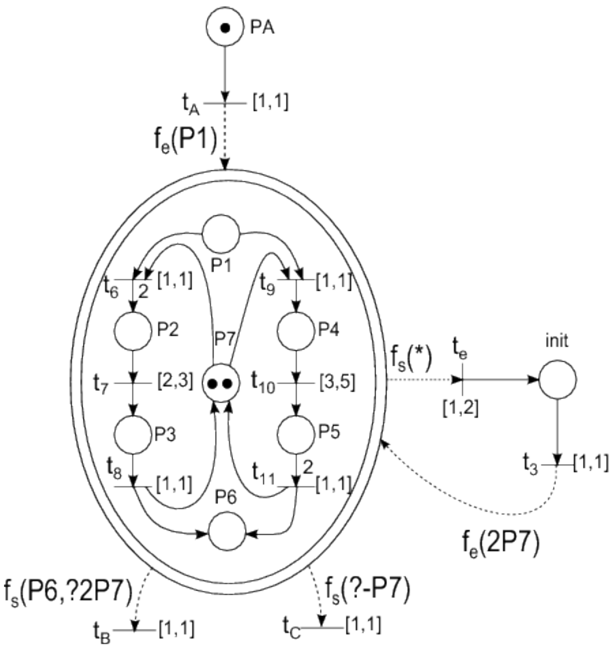


*Figure 13. An example of macroplace and exception catching*

### 6.3.3. Increasing stimulation selectivity using the Transversal Intrafascicular Multichannel Electrode (TIME)

**Participants:** Pawel Maciejasz, David Andreu, David Guiraud, Xavier Navarro, Jordi Badia (Universitat Autònoma de Barcelona), Winnie Jensen, Kristian Rauhe Harreby, Aritra Kundu, Bo Geng (Aalborg University), Thomas Stieglitz, Tim Boretius (University of Freiburg), Ken Yoshida (Purdue University Indianapolis).

The electrical stimulation of nerve fibres may allow to restore or augment some body functions lost due to disease or injury. However, in typical peripheral nerves there are thousands of nerve fibres innervating various organs. Therefore, it is necessary to develop interfaces and methods allowing for selective activation of only desired population of nerve fibres. Various neural interfaces have been already proposed for that purpose, including multipolar cuff electrodes, longitudinal intrafascicular electrodes (LIFE) and the Utah Slanted Electrode Array (USEA), all with different selectivity and invasiveness ratios. Recently a new electrode concept of a transversal intrafascicular multichannel electrode (TIME) has been proposed (Fig. 14 ). This electrode has been developed in frame of the European Project TIME in which the DEMAR team is participating (grant CP-FP-INFSO 224012 from the European Union). It is intended to be implanted transversally in the nerve and address several fascicles or subgroups of nerve fibres with one device. It has longitudinal shape and has several independent stimulation sites equally spread on both sides of the electrode.

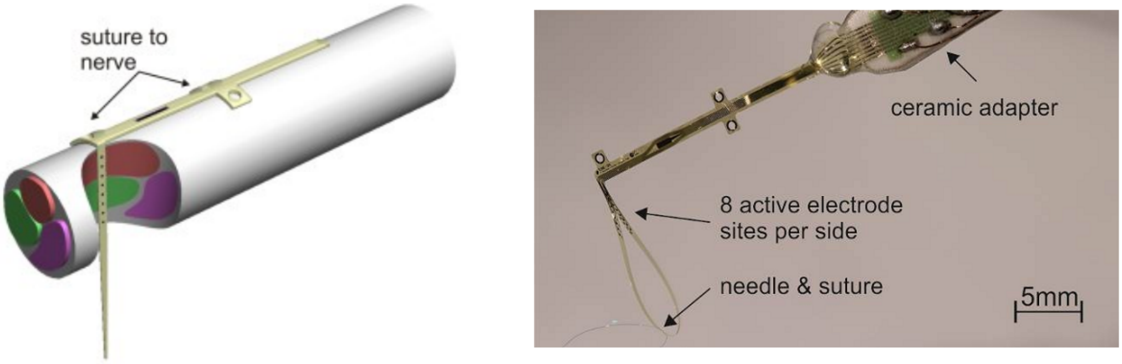


*Figure 14. Left: The schematic representation of the TIME electrode implanted into the peripheral nerve. Right: detailed photograph of an TIME-3H electrode with pre-attached loop-thread (Source: Boretius et al. Proc. IEEE Biomedical Robotics and Biomechatronics Conference, Rome 2012)*

It has been already shown that TIME allows to achieve high selectivity of stimulation when using monopolar configuration, i.e. when current is delivered through one of the sites of the TIME against small needle electrode placed in the proximity of the nerve. We have performed investigation in the sciatic nerve of rat to verify if the use of bipolar configuration, i.e. when current is delivered through one of the TIME sites against an other site of the same electrode, could allow to further enhance selectivity of stimulation. The results of our studies suggest that using bipolar configuration do allow to increase selectivity of stimulation. However, higher charge of the stimulation may be necessary to achieve similar level of muscle activation, as compared to the monopolar configuration [24].

When applied in the rat model the Transverse Intrafascicular Multi-channel Electrode (TIME) showed selective nerve fascicle recruitment. But results from the larger and poly-fasicular median nerves in pigs indicated that a single TIME could not reach the entire nerve and could only selectively recruit a subset of the nerve fasicles. The use of multiple TIME structures could offer a means to achieve highly selective fascicular stimulation while reaching a larger percentage of the fascicles in the nerve. Therefore we have investigated the use of pairs of TIMEs implanted in the median nerves of anesthesized pigs. TIME structures were implanted at different angles relative to each other or in parallel with one another. Electrical stimuli was passed through each contact of each TIME and the resulting electromyograms were recorded from seven muscles innervated by the median nerve. The ability to recruit these muscles was used to assess the stimulation selectivity of each contact using a selectivity index comparing the root-mean-square of the the evoked EMG of individul muscles. Results showed a signifcant increase in the selectivity index, when using two TIMEs compared to one. The optimal improvement was obeserved when TIMEs were placed in parallel to each other in such a way that they interfaced non-overlapping nerve regions [18].

### 6.3.4. Nerve model for ENG recording

**Participants:** Olivier Rossel, Guy Cathébras, Fabien Soulier, Serge Bernard.

In the context of selective electroneurogram recording, we showed last year the efficiency of a *small tripole* filtering (the distance between contacts is $375\,\mu\mathrm{m}$) thanks to simulated signals. This recording is locally sensitive and greatly increases the selectivity of the electrode. This year, we realized an experiment to verify the simulated results. Theoritical study of the *small tripole* sensitivity was realized for a single-fiber action potential (SFAP). We wanted to proceed in the same way for the experiment by trying to measure a SFAP.

However, actual biological SFAP would be hardly measurable by a *small tripole* in an in-vitro experiment. So, we decided to choose an approach based on an artificial axon. In this artificial model, every parameter is under control. The position of the fiber, the nodes of Ranvier, the position of the measuring electrode, as well as the involved currents are perfectly known. This allows us to implement and to estimate with accuracy the filtering realized by the *small tripole* with exactly the same configurations as in simulations. And by performing measures for several radial distances, we can verify the influence of this distance, in order to estimate the sensibility the *small tripole*. Moreover, we can increase the activity amplitude to be higher than in real fiber and then achieve a beneficial signal to noise ratio.

#### 6.3.4.1. Method

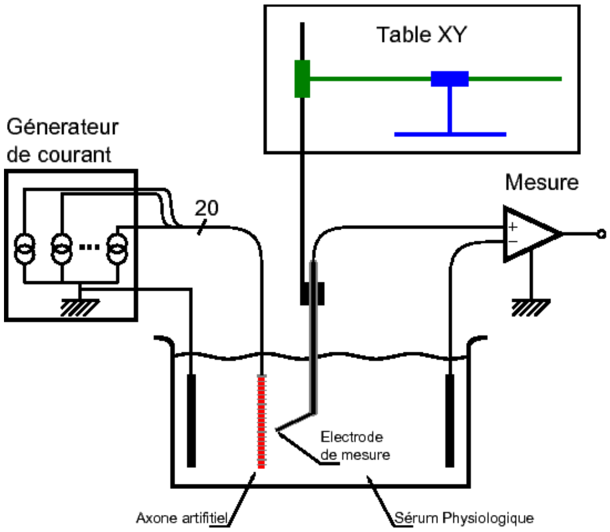


*Figure 15. Principle of the experimental setup using the artificial axon.*

The experimental setup consists in an artificial model emulating an axon and in a system measuring the action potential across the surrounding medium, as sketched in Fig. 15 . Biological tissues are modeled by a saline solution having a conductivity close to the human body. The space and time behavior of the current generated by a natural axon is reproduced on the artificial axon. The latter is emulated by an cochlear electrode. It exhibits 20 contacts, being an accurate image of the nodes of Ranvier on an axon of $8.7\,\mu\mathrm{m}$ diameter regarding the spatial periodicity.

In order to generate electric activity on several contacts that can be compared to the one of several nodes of Ranvier during the conduction of an action potential, we realized a custom multi-current generator with asynchronous outputs. The chosen amplitude was $35 \cdot 10^4$ times that of an human axon.

The measure of the SFAP in the space, is done using a punctual monopolar electrode. The position of the measuring electrode relative to the fiber is automatically set by a programmable micromanipulator. Then, the measure is repeated every $50\,\mu$m along $2\,mm$ on the radial axis.

#### 6.3.4.2. Monopolar and tripolar sensitivity

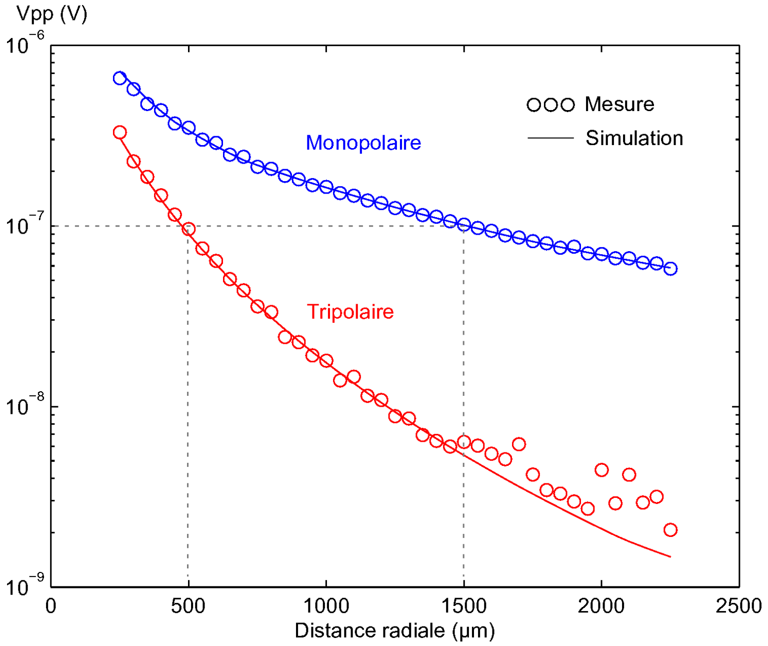


*Figure 16. Comparison between monopolar and small-tripolar recording sensitivities. Measures are plotted with circles and simulation results are solid lines. The monopolar data are in blue and tripolar are in red.*

To verify quantitatively the sensibility results obtained by simulation for monopolar and tripolar electrodes, we have first estimated the amplitude of a SFAP measured by a monopolar electrode. The measured SFAP amplitude directly gives the monopolar recording that varies according to the radial distance between the artificial axon and the electrode. Then, the *small-tripole* recording results from an off-line process combining three points of measure spaced out of $375\,\mu$m. The two kind of recordings are represented in Fig. 16 .

By comparing the recording sensitivities either estimated or measured, we can conclude that measures perfectly fit the theoretical results. We can also notice that the attenuation relative to the radial distance is more important for the *small tripole* than for the monopolar measure. This confirms that the measure performed by one *small tripole* electrode is exclusively sensitive to the closest fibers.

# GALEN Team

# 6. New Results

## 6.1. Machine Learning & Optimization

**Participants:** Andreas Argyriou, Matthew Blaschko, Pawan Kumar.

- **Sparse Prediction & Convex Optimization Decomposition** [*Andreas Argyriou*]
  In [36], we have introduced a new regularization penalty for sparse prediction, the $k$-support norm. This norm corresponds to the tightest convex relaxation of sparsity combined with an $\ell_2$ penalty. We have shown that this new norm provides a tighter relaxation than the elastic net, and is thus a good replacement for the Lasso or the elastic net in sparse prediction problems. In [41], motivated by learning problems we proposed a novel optimization algorithm for minimizing a convex objective which decomposes into three parts: a smooth part, a simple non-smooth Lipschitz part, and a simple non-smooth non-Lipschitz part.
- **Learning Optimization for NP-complete Inference** [*Matthew Blaschko*]
  In [14] an optimization strategy for learning to optimize boolean satisfiability (SAT) solvers is given. Applications to real-world SAT problems show improved computational performance as a result of the learning algorithm.
- **Max-Margin Min-Entropy Models & Dissimilarity Coefficient based Learning** [*Pawan Kumar*]
  In [22] we proposed the family of max-margin min-entropy (M3E) models, which predicts a structured output for a given input by minimizing the Renyi entropy. The parameters of M3E are learned by minimizing an upper bound on a user-defined loss. We demonstrated the efficacy of M3E on two problems using publicly available datasets: motif finding and image classification. In [19] we proposed a novel structured prediction framework for weakly supervised datasets. The framework minimizes a dissimilarity coefficient between the predictor and a conditional distribution over the missing information. We demonstrated the efficacy of our approach on two problems using publicly available datasets: object detection and action detection.

## 6.2. Computational Vision & Perception

**Participants:** Matthew Blaschko, Iasonas Kokkinos, Pawan Kumar, Nikos Paragios.

- **Structured Output Ranking & Detailed Understanding of Objects in Computer Vision** [*Matthew Blaschko*]
  In [23] we proposed a novel method for efficiently optimizing an objective that ranks structured outputs by their loss. Based on the observation that structured output spaces [9] in computer vision problems can be well-modeled by a small number of loss values, our algorithm is able to optimize a quadratic number of pairwise constraints in linear time. In [38] we detail the research activities of a summer workshop hosted by Johns Hopkins University on learning a detailed understanding of objects and scenes in natural images. We worked on automatic verification of annotations provided through Amazon Mechanical Turk [35], texture categorization, and dependence modeling for bottom up proposals.
- **Efficient inference and learning for structured probabilistic models of deformable objects** [*Iasonas Kokkinos*, Haithem Boussaid & Stavros Tsogkas]

We have developed novel features to describe surface points intrinsically through the Intrinsic Shape Context (ISC) descriptor published in [17]. This method has delivered state-of-the-art results in surface point matching and we will explore its use for surface correspondence. The implementation of these descriptors is publicly available. In [32] we proposed a learning-based approach to symmetry detection by fusing multiple cues related to image intensity, color and texture, which delivered state-of-the-art results. We intend to extend this approach to 3D image analysis, and in particular for medical images. The implementation of these detectors is publicly available. In [27] we introduce a grouping-based method to learn and detect action classes in spatio-temporal data. Our method can both classify actions and indicate the spatio-temporal structures which provide support for the decision. The implementation of our front-end is publicly available. In [40] we have extended our work on efficient algorithms for object detection to accommodate fast methods for computing the part scores in a principled optimization framework, while he have thoroughly presented it in [40] and made the implementation publicly available.

- **Multi-view Image Segmentation & Parsing** [*Nikos Paragios*]
  In [28] a method for image matching was proposed that exploits hierarchical image representations through higher order graphs. The matching was achieved through a graph-based theoretical framework where the similarity and spatial consistency of the image semantic objects is encoded in a graph of commute times that is also endowed with singleton terms through shape descriptors. Many-to-many matching of regions are specially challenging due to the instability of the segmentation under slight image changes, and we explicitly handle it through high order potentials. These ideas were further explored in the context of co-segmentation [29] where a method to determine a consistent partition of multiple images was introduced through a multi-scale multiple-image generative model based on region matching that exploits inter-image information and establishes correspondences between the common objects that appear in the scene. Last, but not least in [24] a method that combines bottom up (visual information, visual descriptors, elements detection) information and top-town models (hierarchical shape grammars) was considered towards automatic facade parsing though reinforcement learning while in [30] a method for 3D image parsing was proposed based on a hierarchical grammar that was performing explicit 3D modeling of the scene through a combination of multi-image segmentation and a depth reconstruction process. The problem optimal combination of these two concurrent terms was addressed trough a pareto-driven criterion while the optimization was addressed through an evolutionary computation algorithm.

## 6.3. Biomedical Image Analysis

**Participant:** Nikos Paragios.

- **Image Reconstruction** [*Nikos Paragios* & Hellene Langet]
  In [21] a novel iterative reconstruction algorithm based on compressed sensing was proposed for Digital Subtraction Rotational Angiography (DSRA) that exploits both spatial and temporal sparsity through a proximal implementation that accommodates multiple$L-1$-penalties. These ideas was further explored in [20] where we introduced a three-dimensional reconstruction of tomographic acquisitions in C-arm-based rotational angiography was proposed that was able to deal with the temporal variations due to intra-arterial injections through a compressed-sensing approach leading to significant motion artifacts reduction in spite of the cone-beam geometry, the short-scan acquisition, and the truncated and subsampled data.
- **Image Segmentation** [*Nikos Paragios*, Pierre-Yves Baudin, Xiang Bo & Sarah Parisot]
  In [11] the problem of human skeletal muscle segmentation was considered through a graph-based approach (random walker). An automatic seed placement framework was introduced through a graph-theoretic formulation. Towards accounting for anatomical constraints, the Random Walker algorithm was endowed with a liner sub-space statistical prior towards improving segmentation robustness on missing and incomplete data [12]. The same formulation was extended to cope with non-linear priors through a Gaussian-like local prior model penalizing the deviations of the

coefficients of the random walker diffusion matrix from the ones learned from the training data [13]. In [25] a novel graph-based prior was considered towards modeling the distribution of low-glioma brain tumors and spatially characterizing them through a sparse hierarchical graph. Such a prior model was integrated to an image-driven voxel-like segmentation framework where image separation was achieved through a machine learning method towards automatic detection, characterization and segmentation of brain tumors. Furthermore, towards encoding pose invariance in the context of knowledge-based segmentation in [33] where a higher order graph-based implicit pose invariant formulation was introduced for cardiac segmentation. The formulation was endowed with higher order cliques allowing (i) the estimation of boundary and regional image support and (ii) the implicit modeling of local deformations with respect to a prior statistical model while being invariant to linear transformations.

- **Image Registration** [*Nikos Paragios*, Nicolas Honnorat & Sarah Parisot]
  In [15] the problem of organ-driven registration was addressed through simultaneous combined fusion of multi-modal images in the context of guide-wire segmentation through fluoroscopic and contrast enhanced images. To this end, a graphical model was considered that was segmenting and registering the guide-wire in the two modalities while establishing correspondences between the associated curves as well. Similar philosophy was used in the [26] where a method for one shot deformable brain registration and tumor segmentation was proposed between a healthy anatomical atlas and a diseased patient. Both tasks were addressed through a discrete formulation (pair-wise MRF using grid-like deformation models and machine learning discriminative frameworks for the separation of healthy versus diseased tissues) while interconnections between the two graphs were used to alleviate the registration requirement on tumor areas. The problem of symmetric registration was studied in [31] through a common grid deforming in both directions according to a symmetric manner towards minimizing the image similarity criterion between the source and the target image while guaranteeing the expected diffeomorphic nature of the deformation field.
- **Computational Anatomy** [*Nikos Paragios*]
  In [16] we introduced a novel approach for detecting the presence of white matter lesions in periventricular areas of the brain using manifold-constrained embeddings. The proposed method uses locally linear embedding (LLE) to create ”normality” distributions of the brain where deviations from the manifolds are estimated by calculating geodesic distances along locally linear planes in the embedding. Experiments highlight the need of nonlinear techniques to learn the studied data leading to outstanding detection rates when comparing individuals to a specific pathological pattern.

## MNEMOSYNE Team

# 6. New Results

## 6.1. Introduction

2012 is the year of birth of the Mnemosyne team; it is also a year of transition, since most of its members were previously in the Cortex project-team in Nancy. Accordingly, this year was partly devoted to ongoing projects in the Cortex team and to the initiation of the first activities in Mnemosyne. Apart from the results listed below, corresponding to the extension of current projects from the Cortex team to our new topics, we have also begun to study models of episodic and semantic memory at the interface between the hippocampus and the cortex and interaction between models of the basal ganglia and the prefrontal cortex.

## 6.2. Systemic view of visuomotor transformations

Visuomotricity was an important topic in our activities in the Cortex team. We are pursuing these activities in ongoing projects (particularly the ANR Keops project, cf. § 7.1 ), and extend them according to a systemic view, integrating new information flows from the external world and from other neuronal structures:

- We consider the role of non-standard cells in the retina [11], reported as responsible for fast event-detection in the visual flow.
- We have initiated a modeling study of the retina-thalamus-cortex information flow and particularly of its non-specific pathway [9] that can account for attentional mechanisms.

# NEUROMATHCOMP Project-Team

# 5. New Results

## 5.1. Neural Networks as dynamical systems

### 5.1.1. Dynamics and spike trains statistics in conductance-based Integrate-and-Fire neural networks with chemical and electric synapses

**Participants:** Rodrigo Cofré, Bruno Cessac [correspondent].

We investigate the effect of electric synapses (gap junctions) on collective neuronal dynamics and spike statistics in a conductance-based Integrate-and-Fire neural network, driven by a Brownian noise, where conductances depend upon spike history. We compute explicitly the time evolution operator and show that, given the spike-history of the network and the membrane potentials at a given time, the further dynamical evolution can be written in a closed form. We show that spike train statistics is described by a Gibbs distribution whose potential can be approximated with an explicit formula, when the noise is weak. This potential form encompasses existing models for spike trains statistics analysis such as maximum entropy models or Generalized Linear Models (GLM). We also discuss the different types of correlations: those induced by a shared stimulus and those induced by neurons interactions. This work has been presented in several conferences [43], [45], [46], [47], [31], [48] and submitted to Chaos, Solitons and Fractals [13].

### 5.1.2. Parameter estimation in spiking neural networks: a reverse-engineering approach

**Participants:** Horacio Rostro-Gonzalez [Holistic Electronics Research Lab, University of Cyprus], Bruno Cessac [correspondent], Thierry Viéville [Inria Mnemosyne].

This work presents a reverse engineering approach for parameter estimation in spiking neural networks (SNNs). We consider the deterministic evolution of a time-discretized network with spiking neurons, where synaptic transmission has delays, modeled as a neural network of the generalized integrate and fire type. Our approach aims at by-passing the fact that the parameter estimation in SNN results in a non-deterministic polynomial-time hard problem when delays are to be considered. Here, this assumption has been reformulated as a linear programming (LP) problem in order to perform the solution in a polynomial time. Besides, the LP problem formulation makes explicit the fact that the reverse engineering of a neural network can be performed from the observation of the spike times. Furthermore, we point out how the LP adjustment mechanism is local to each neuron and has the same structure as a 'Hebbian' rule. Finally, we present a generalization of this approach to the design of input–output (I/O) transformations as a practical method to 'program' a spiking network, i.e. find a set of parameters allowing us to exactly reproduce the network output, given an input. Numerical verifications and illustrations are provided. This work has been published in Journal of Neural Engineering [24].

## 5.2. Mean field approaches

### 5.2.1. Noise-induced behaviors in neural mean field dynamics

**Participants:** Olivier Faugeras [correspondent], Geoffroy Hermann, Jonathan Touboul [Inria Bang].

The collective behavior of cortical neurons is strongly affected by the presence of noise at the level of individual cells. In order to study these phenomena in large-scale assemblies of neurons, we consider networks of firing-rate neurons with linear intrinsic dynamics and nonlinear coupling, belonging to a few types of cell populations and receiving noisy currents. Asymptotic equations as the number of neurons tends to infinity (mean field equations) are rigorously derived based on a probabilistic approach. These equations are implicit on the probability distribution of the solutions which generally makes their direct analysis difficult. However, in our case, the solutions are Gaussian, and their moments satisfy a closed system of nonlinear ordinary differential equations (ODEs), which are much easier to study than the original stochastic network equations, and the statistics of the empirical process uniformly converge towards the solutions of these ODEs. Based on this description, we analytically and numerically study the influence of noise on the collective behaviors, and compare these asymptotic regimes to simulations of the network. We observe that the mean field equations provide an accurate description of the solutions of the network equations for network sizes as small as a few hundreds of neurons. In particular, we observe that the level of noise in the system qualitatively modifies its collective behavior, producing for instance synchronized oscillations of the whole network, desynchronization of oscillating regimes, and stabilization or destabilization of stationary solutions. These results shed a new light on the role of noise in shaping collective dynamics of neurons, and gives us clues for understanding similar phenomena observed in biological networks. This work has been published in the SIAM Journal on Applied dynamical Systems [25].

### *5.2.2. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley neurons*

**Participants:** Javier Baladron, Diego Fasoli, Olivier Faugeras [correspondent], Jonathan Touboul [Inria Bang].

We derive the mean-field equations arising as the limit of a network of interacting spiking neurons, as the number of neurons goes to infinity. The neurons belong to a fixed number of populations and are represented either by the Hodgkin-Huxley model or by one of its simplified version, the FitzHugh-Nagumo model. The synapses between neurons are either electrical or chemical. The network is assumed to be fully connected. The maximum conductances vary randomly. Under the condition that all neurons' initial conditions are drawn independently from the same law that depends only on the population they belong to, we prove that a propagation of chaos phenomenon takes place, namely that in the mean-field limit, any finite number of neurons become independent and, within each population, have the same probability distribution. This probability distribution is a solution of a set of implicit equations, either nonlinear stochastic differential equations resembling the McKean-Vlasov equations or non-local partial differential equations resembling the McKean-Vlasov-Fokker-Planck equations. We prove the well posedness of the McKean-Vlasov equations, i.e. the existence and uniqueness of a solution. We also show the results of some numerical experiments that indicate that the mean-field equations are a good representation of the mean activity of a finite size network, even for modest sizes. These experiments also indicate that the McKean-Vlasov-Fokker-Planck equations may be a good way to understand the mean-field dynamics through, e.g. a bifurcation analysis. This work has appeared in the Journal of Mathematical Neuroscience [11].

## 5.3. Neural fields theory

### *5.3.1. Localized radial bumps of a neural field equation on the Euclidean plane and the Poincaré disk*

**Participants:** Grégory Faye [correspondent], James Rankin, David, J.B. Lloyd [Department of Mathematics at the University of Surrey].

We analyze radially symmetric localized bump solutions of an integro-differential neural field equation posed in Euclidean and hyperbolic geometry. The connectivity function and the nonlinear firing rate function are chosen such that radial spatial dynamics can be considered. Using integral transforms, we derive a PDE for the neural field equation in both geometries and then prove the existence of small amplitude radially symmetric spots bifurcating from the trivial state. Numerical continuation is then used to path-follow the spots and their bifurcations away from onset in parameter space. It is found that the radial bumps in Euclidean geometry are linearly stable in a larger parameter region than bumps in the hyperbolic geometry. We also find and path-follow localized structures that bifurcate from branches of radially symmetric solutions with D6-symmetry and D8-symmetry in the Euclidean and hyperbolic cases, respectively. Finally, we discuss the applications of our results in the context of neural field models of short term memory and edges and textures selectivity in a hypercolumn of the visual cortex. This work has been accepted for publication in Nonlinearity [57].

### 5.3.2. Center manifold for delayed neural fields equations

**Participants:** Olivier Faugeras [correspondent], Romain Veltz [Salk Institute,San Diego,USA].

We develop a framework for the study of delayed neural fields equations and prove a center manifold theorem for these equations. Specific properties of delayed neural fields equations make it impossible to apply existing methods from the literature concerning center manifold results for functional differential equations. Our approach for the proof of the center manifold theorem uses the original combination of results from Vanderbauwhede and colleagues together with a theory of linear functional differential equations in a history space larger than the commonly used set of time-continuous functions. This work has been submitted to the SIAM Journal on Applied Mathematics and is under revision [27].

### 5.3.3. Reduction method for localized solutions

**Participant:** Grégory Faye [correspondent].

We present a reduction method to study localized solutions of an integrodifferential equation defined on the Poincaré disk. This equation arises in a problem of texture perception modeling in the visual cortex. We first derive a partial differential equation which is equivalent to the initial integrodifferential equation and then deduce that localized solutions which are radially symmetric satisfy a fourth order ordinary differential equation. This work has appeared in the “Comptes Rendus Mathematique” [15].

### 5.3.4. Spatially localized solutions

**Participants:** Pascal Chossat, Grégory Faye [School of Mathematics,University of Minnesota, correspondent], James Rankin.

The existence of spatially localized solutions in neural networks is an important topic in neuroscience as these solutions are considered to characterize working (short-term) memory. We work with an unbounded neural network represented by the neural field equation with smooth firing rate function and a wizard hat spatial connectivity. Noting that stationary solutions of our neural field equation are equivalent to homoclinic orbits in a related fourth order ordinary differential equation, we apply normal form theory for a reversible Hopf bifurcation to prove the existence of localized solutions; further, we present results concerning their stability. Numerical continuation is used to compute branches of localized solution that exhibit snaking-type behaviour. We describe in terms of three parameters the exact regions for which localized solutions persist. This work has appeared in the Journal of Mathematical Biology [16].

### 5.3.5. Bumps in the Poincaré disk

**Participants:** Grégory Faye [School of Mathematics,University of Minnesota, correspondent], David, J.B. Loyd, James Rankin.

We analyze radially symmetric localized bump solutions of an integro-differential neural field equation posed in Euclidean and hyperbolic geometry. The connectivity function and the nonlinear firing rate function are chosen such that radial spatial dynamics can be applied. Using integral transforms, we derive a PDE of the neural field equation in both geometries and then prove the existence of small amplitude radially symmetric spots bifurcating from the trivial state. Numerical continuation is then used to path-follow the spots and their bifurcations away from onset in parameter space. It is found that the radial bumps in Euclidean geometry are linearly stable in a larger parameter region than bumps in the hyperbolic geometry. We also find and path follow localized structures that bifurcate from branches of radially symmetric solutions with D6-symmetry and D8-symmetry in the Euclidean and hyperbolic cases, respectively. Finally, we discuss the applications of our results in the context of neural field models of short term memory and edges and textures selectivity in a hypercolumn of the visual cortex. This work has been submitted to Nonlinearity.

## 5.4. Spike trains statistics

### 5.4.1. *Natural image identification from spike train analysis*

**Participants:** Geoffrey Portelli, Olivier Marre [Institution de la Vision, Paris, France], Marc Antonini [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Michael Berry II [Princeton Neuroscience Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA], Pierre Kornprobst [correspondent].

We started a new activity to analyse how natural images are encoded in retinal output. This work is related to [67], [72] where synthetic stimuli are used. Here, we recorded a population of 100-200 ganglion cells of a salamander retina, while flashing 720 natural images from the Torralba database [71] plus one control image, each repeated 10 times. We characterized the response of each cell by two parameters : the latency of the first spike after the stimulus onset, and the firing rate. A distribution of these two features was then estimated for each neuron and natural image. Pooling the information across all the neurons, a discriminability coefficient between pairs of image is proposed, using either the rate or the latency, or both. We also provide a way to identify a given image among others based on the rate–latency distributions. Preliminary results have been presented in [40]. Results showed that, on average, the discriminability was better based on the latency than on the rate. The most discriminable pairs were different using the rate or the latency, so these two features conveyed complementary information. In addition, we observe a similar evolution of the identification performance when the rate, or the latency, or both are used.

### 5.4.2. *Spike train statistics from empirical facts to theory: the case of the retina*

**Participants:** Bruno Cessac [correspondent], Adrian Palacios [CINV-Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso].

This work focuses on methods from statistical physics and probability theory allowing the analysis of spike trains in neural networks. Taking as an example the retina we present recent works attempting to understand how retina ganglion cells encode the information transmitted to the visual cortex via the optical nerve, by analyzing their spike train statistics. We compare the maximal entropy models used in the literature of retina spike train analysis to rigorous results establishing the exact form of spike train statistics in conductance-based Integrate-and-Fire neural networks. This work has been published in Mathematical Problems in Computational Biology and Biomedicine, F. Cazals and P. Kornprobst, Springer [55].

### 5.4.3. *Gibbs distribution analysis of temporal correlations structure in retina ganglion cells*

**Participants:** Juan-Carlos Vasquez, Olivier Marre [Institution de la Vision, Paris, France], Adrian Palacios [CINV-Centro Interdisciplinario de Neurociencia de Valparaiso, Universidad de Valparaiso], Michael Berry II [Princeton Neuroscience Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA], Bruno Cessac [correspondent].

We present a method to estimate Gibbs distributions with spatio-temporal constraints on spike trains statistics. We apply this method to spike trains recorded from ganglion cells of the salamander retina, in response to natural movies. Our analysis, restricted to a few neurons, performs more accurately than pairwise synchronization models (Ising) or the 1-time step Markov models (Marre et al. (2009)) to describe the statistics of spatio-temporal spike patterns and emphasizes the role of higher order spatio-temporal interactions. This work has been presented in several conferences [29], [30], [28] and pusblished in J. Physiol. Paris [26].

### 5.4.4. *Spatio-temporal spike trains analysis for large scale networks using maximum entropy principle and Monte-Carlo method*

**Participants:** Hassan Nasser, Olivier Marre [Institut de la Vision, Paris, France], Bruno Cessac [correspondent].

Understanding the dynamics of neural networks is a major challenge in experimental neuroscience. For that purpose, a modelling of the recorded activity that reproduces the main statistics of the data is required. We present a review on recent results dealing with spike train statistics analysis using maximum entropy models (MaxEnt). Most of these studies have been focusing on modelling synchronous spike patterns, leaving aside the temporal dynamics of the neural activity. However, the maximum entropy principle can be generalized to the temporal case, leading to Markovian models where memory effects and time correlations in the dynamics are properly taken into account. We also present a new method based on Monte-Carlo sampling which is suited for the fitting of large-scale spatio-temporal MaxEnt models. The formalism and the tools presented here will be essential to fit MaxEnt spatio-temporal models to large neural ensembles. This work has been presented in several conferences [54], [51], [53], [52] and accepted in Journal of Statistical Mechanics [22].

### *5.4.5. Spike train statistics and Gibbs distributions*

**Participants:** Rodrigo Cofré, Bruno Cessac [correspondent].

We introduce Gibbs distribution in a general setting, including non stationary dynamics, and present then three examples of such Gibbs distributions, in the context of neural networks spike train statistics: (i) Maximum entropy model with spatio-temporal constraints; (ii) Generalized Linear Models; (iii) Conductance based Inte- grate and Fire model with chemical synapses and gap junctions. This leads us to argue that Gibbs distributions might be canonical models for spike train statistics analysis. This work has been presented in several conferences [43], [31] and submitted to J. Physiol. Paris [12].

## 5.5. Visual Neuroscience

### *5.5.1. Neural fields models for motion integration: Characterising the dynamics of multi-stable visual motion stimuli*

**Participants:** Olivier Faugeras, Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], Andrew Meso [Institut de Neurosciences de la Timone, UMR 6193, CNRS, Marseille, France], James Rankin.

We are investigating the temporal dynamics of the neural processing of a multi-stable visual motion stimulus with two complementary approaches: psychophysical experiments and mathematical modelling. The so called “barber pole” stimulus is considered with an aperture configuration that supports horizontal (H), diagonal (D) or vertical (V) perceived directions for the same input. The phenomenon demonstrates an interesting variable and dynamic competition for perceptual dominance between underlying neural representations of the three directions. We study the temporal dynamics of this phenomenon with a neural fields, population-level representation of activity in MT, a cortical area dedicated to motion estimation. Numerical tools from bifurcation analysis are used to investigate the model’s behaviour in the presence of different types of input; this general approach could be applied to a range of neural fields models that are typically studied only in terms of their spontaneous activity. The model reproduces known multistable behaviour in terms of the predominant interpretations (percepts) of the barber pole stimulus.

We probe the early processing from stimulus presentation to initial perceived direction (before perceptual reversals). The basic dynamic properties of the early transition from D to H/V are well predicted by the model. This work has been presented in the European Conference on Vision Perception (ECVP) [38], [41] and it has been published in [23].

We are extending this work to investigate the longer term dynamics for which perceptual reversals are known to occur, due to competition between 1D motion cues aligned with the grating’s motion direction and 2D motion cues aligned with aperture edges. This work has been presented in the Vision Sciences Society 12th Annual Meeting (VSS) [39], [42].

## 5.6. Neuromorphic Vision

**Participants:** Khaled Masmoudi [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Marc Antonini [Laboratoire I3S, UMR CNRS, Universite´ Nice Sophia Antipolis, France], Pierre Kornprobst.

In the scope of Khaled Masmoudi's PhD [9], we have developed bio-inspired schemes for image coding. This is a new area of research on which very few teams are committed. We have proposed schemes for encoding/decoding images directly using the functional architecture of the retina and the properties of its spiking output (e.g., using Laplacian pyramids model and the rank-order coding [69], [68][20], [21] and the Virtual Retina simulator [37] from Adrien Wohrer during his PhD [79], [78]).

## PARIETAL Project-Team

# 6. New Results

## 6.1. Randomized cluster-based predictive model

**Participants:** Gaël Varoquaux [Correspondant], Bertrand Thirion, Alexandre Gramfort.

Functional neuroimaging can measure the brain's response to an external stimulus. It is used to perform brain mapping: identifying from these observations the brain regions involved. This problem can be cast into a linear supervised learning task where the neuroimaging data are used as predictors for the stimulus. Brain mapping is then seen as a support recovery problem. On functional MRI (fMRI) data, this problem is particularly challenging as i) the number of samples is small due to limited acquisition time and ii) the variables are strongly correlated. We propose to overcome these difficulties using sparse regression models over new variables obtained by clustering of the original variables. The use of randomization techniques, e.g. bootstrap samples, and clustering of the variables improves the recovery properties of sparse methods. We demonstrate the benefit of our approach on an extensive simulation study as well as two fMRI datasets.

More details can be found in [32].

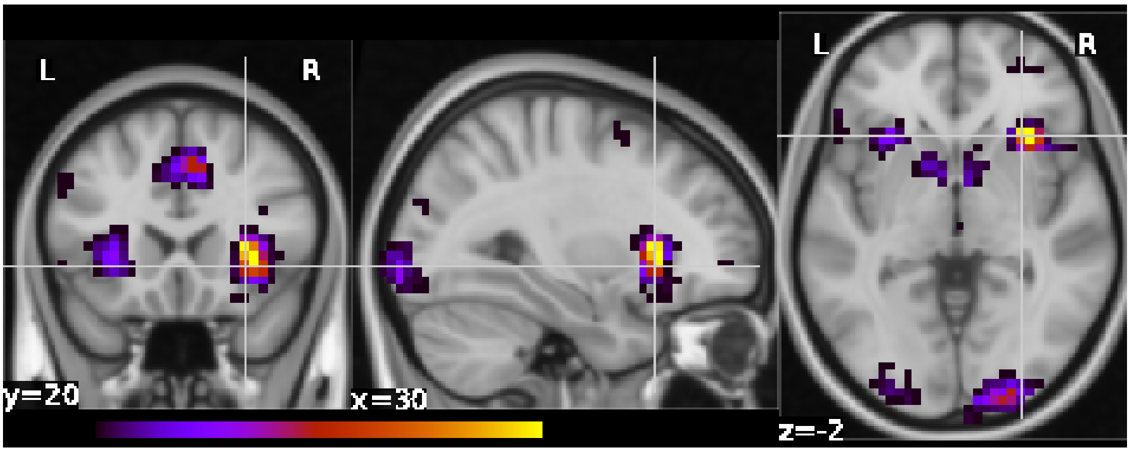


*Figure 3. The randomized cluster-based predictive model can be used to predict the behavior of the subject, such as the gain in a gambling task [33]. More importantly, the support of the resulting classifier is indeed sparse and provides a reliable definition of the truly involved regions.*

## 6.2. Random Projections for Outlier Detection

**Participants:** Gaël Varoquaux, Bertrand Thirion, Jean-Baptiste Poline, Virgile Fritsch [Correspondant].

Medical imaging datasets often contain deviant observations, the so-called outliers, due to acquisition or preprocessing artifacts or resulting from large intrinsic inter-subject variability. These can undermine the statistical procedures used in group studies as the latter assume that the cohorts are composed of homogeneous samples with anatomical or functional features clustered around a central mode. The effects of outlying subjects can be mitigated by detecting and removing them with explicit statistical control. With the emergence of large medical imaging databases, exhaustive data screening is no longer possible, and automated outlier detection methods are currently gaining interest. The datasets used in medical imaging are often high-dimensional and strongly correlated. The outlier detection procedure should therefore rely on high-dimensional statistical multivariate models. However, state-of-the-art procedures are not well-suited for such high-dimensional settings. In this work, we introduce regularization in the Minimum Covariance Determinant framework and investigate different regularization schemes. We carry out extensive simulations to provide backing for practical choices in absence of ground truth knowledge. We demonstrate on functional neuroimaging datasets that outlier detection can be performed with small sample sizes and improves group studies.

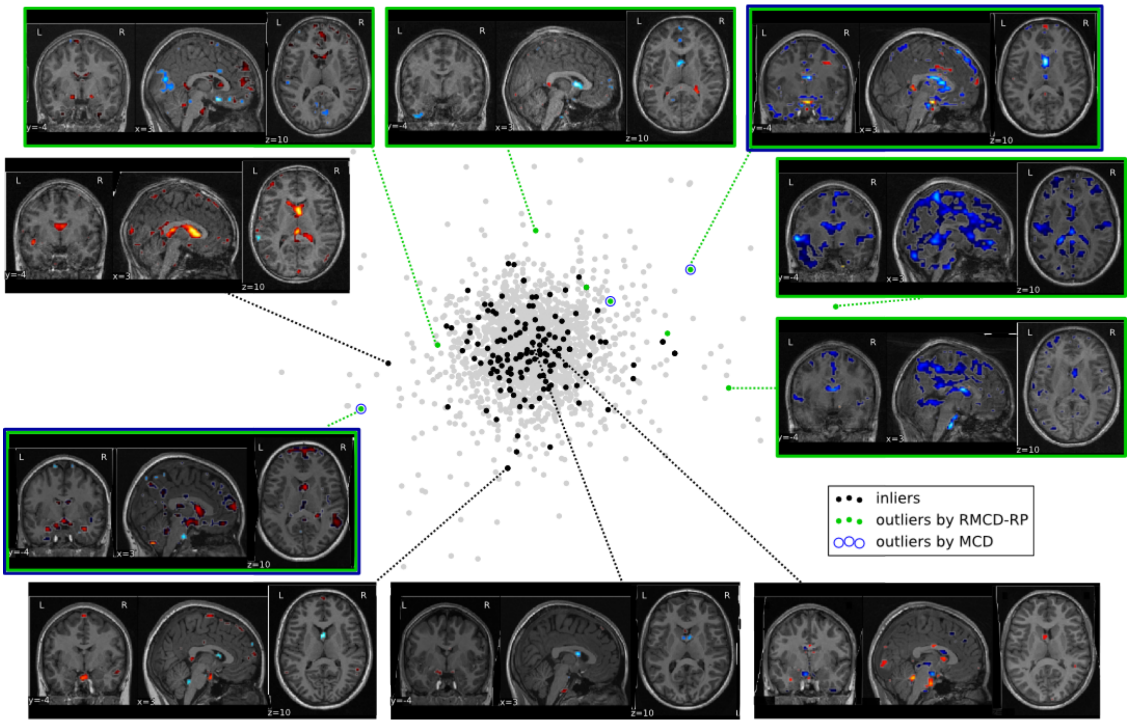


*Figure 4. A large set of images can be mined for structures using the regularized MCD framework, which reveals both standard and unusual patterns in these images.*

More details can be found in [11].

## 6.3. Registration of brain images based on Currents

**Participants:** Pierre Fillard, Bertrand Thirion, Viviana Siless [correspondant].

We present an extension of the diffeomorphic Geometric Demons algorithm which combines the iconic registration with geometric constraints. Our algorithm works in the log-domain space, so that one can efficiently compute the deformation field of the geometry. We represent the shape of objects of interest in the space of currents which is sensitive to both location and geometric structure of objects. Currents provide a distance between geometric structures that can be defined without specifying explicit point-to-point correspondences. We demonstrate this framework by registering simultaneously T1 images and 65 fiber bundles consistently extracted in 12 subjects and compare it against non-linear T1, tensor, and multi-modal T1+ Fractional Anisotropy (FA) registration algorithms. Results show the superiority of the Log-domain Geometric Demons over their purely iconic counterparts.

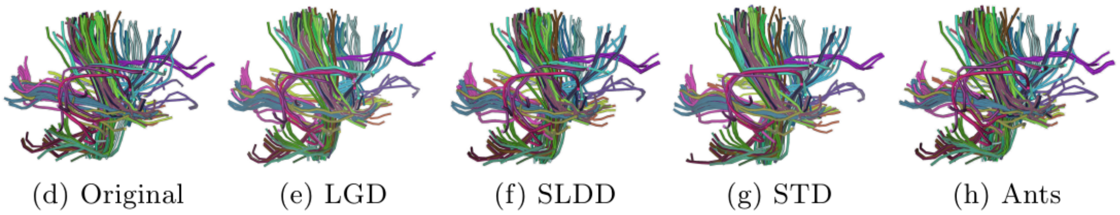

(d) Original (e) LGD (f) SLDD (g) STD (h) Ants

*Figure 5. Comparison of the fiber registration through various algorithms. We display a moving and a reference fiber for 29 selected bundles. The proposed approach (SLDD) outperforms state-of-the art alternatives that do not take into account the fiber geometry explicitly.*

More details can be found in [31].

## 6.4. Structured Sparsity for brain mapping

**Participants:** Gaël Varoquaux [Correspondant], Bertrand Thirion, Alexandre Gramfort.

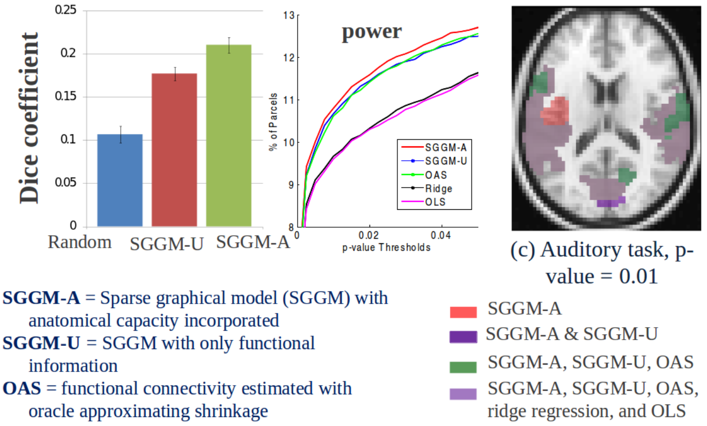
Reverse inference, or *brain reading*, is a recent paradigm for analyzing functional magnetic resonance imaging (fMRI) data, based on pattern recognition and statistical learning. This approach aims at decoding brain activity by predicting some cognitive variables related to brain activation maps. Reverse inference takes into account the multivariate information between voxels and is currently the only way to assess how precisely some cognitive information is encoded by the activity of neural populations within the whole brain. However, it relies on a prediction function that is plagued by the curse of dimensionality, since there are far more features than samples, i.e., more voxels than fMRI volumes. To address this problem, different methods have been proposed, such as, among others, univariate feature selection, feature agglomeration and regularization techniques. In this work, we consider a sparse hierarchical structured regularization. Specifically, the penalization we use is constructed from a tree that is obtained by spatially-constrained agglomerative clustering. This approach encodes the spatial structure of the data at different scales into the regularization, which makes the overall prediction procedure more robust to inter-subject variability. The regularization used induces the selection of spatially coherent predictive brain regions simultaneously at different scales. We test our algorithm on real data acquired to study the mental representation of objects, and we show that the proposed algorithm not only delineates meaningful brain regions but yields as well better prediction accuracy than reference methods.

More details can be found in [15].

## 6.5. A Novel Sparse Graphical Approach for Multimodal Brain Connectivity Inference

**Participants:** Bertrand Thirion, Jean-Baptiste Poline, Gaël Varoquaux [Correspondant], Bernard Ng.

Despite the clear potential benefits of combining fMRI and diffusion MRI in learning the neural pathways that underlie brain functions, little methodological progress has been made in this direction. In this work, we propose a novel multimodal integration approach based on sparse Gaussian graphical model for estimating brain connectivity. Casting functional connectivity estimation as a sparse inverse covariance learning problem, we adapt the level of sparse penalization on each connection based on its anatomical capacity for functional interactions. Functional connections with little anatomical support are thus more heavily penalized. For validation, we showed on real data collected from a cohort of 60 subjects that additionally modeling anatomical capacity significantly increases subject consistency in the detected connection patterns. Moreover, we demonstrated that incorporating a connectivity prior learned with our multimodal connectivity estimation approach improves activation detection.



(c) Auditory task, p-value = 0.01

*Figure 6. The information conveyed by anatomical connectivity improves the estimation of functional connectivity, as it makes it more reproducible. It also enhances the power of fMRI activation detection studies when used as a prior on these activation maps.*

More details can be found in [26].

## 6.6. Transfer learning for met-analyses of functional neuroimaging datasets

**Participants:** Bertrand Thirion, Jean-Baptiste Poline, Gaël Varoquaux, Yannick Schwartz [Correspondant].

Typical cohorts in brain imaging studies are not large enough for systematic testing of all the information contained in the images. To build testable working hypotheses, investigators thus rely on analysis of previous work, sometimes formalized in a so-called meta-analysis. In brain imaging, this approach underlies the specification of regions of interest (ROIs) that are usually selected on the basis of the coordinates of previously detected effects. In this work, we propose to use a database of images, rather than coordinates, and frame the problem as transfer learning: learning a discriminant model on a reference task to apply it to a different but related new task. To facilitate statistical analysis of small cohorts, we use a sparse discriminant model that selects predictive voxels on the reference task and thus provides a principled procedure to define ROIs. The benefits of our approach are twofold. First it uses the reference database for prediction, i.e. to provide potential biomarkers in a clinical setting. Second it increases statistical power on the new task. We demonstrate on a set of 18 pairs of functional MRI experimental conditions that our approach gives good prediction. In addition, on a specific transfer situation involving different scanners at different locations, we show that voxel selection based on transfer learning leads to higher detection power on small cohorts.

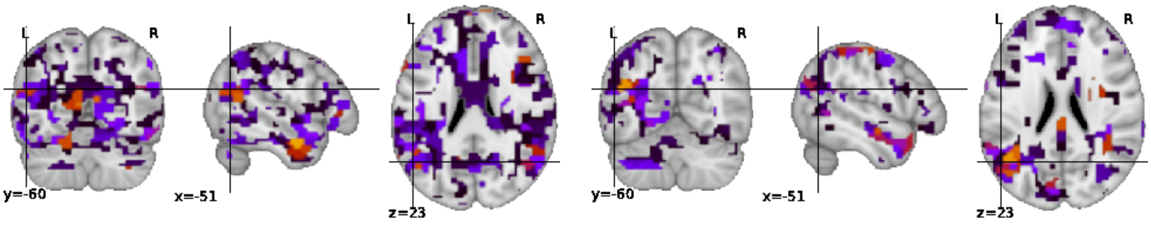


*Figure 7. The brain regions that reliably predict that the subject is listening to Korean versus native (french) language (left) are similar to those that can be used to predict that the subject is listening an unintelligible language (jabberwoky) as opposed to their native (french) language (right).*

More details can be found in [29] and [30].

## 6.7. Learning to rank medical images

**Participants:** Bertrand Thirion, Gaël Varoquaux, Alexandre Gramfort, Fabian Pedregosa [Correspondant].

Medical images can be used to predict a clinical score coding for the severity of a disease, a pain level or the complexity of a cognitive task. In all these cases, the predicted variable has a natural order. While a standard classifier discards this information, we would like to take it into account in order to improve prediction performance. A standard linear regression does model such information, however the linearity assumption is likely not be satisfied when predicting from pixel intensities in an image. In this work we address these modeling challenges with a supervised learning procedure where the model aims to order or rank images. We use a linear model for its robustness in high dimension and its possible interpretation. We show on simulations and two fMRI datasets that this approach is able to predict the correct ordering on pairs of images, yielding higher prediction accuracy than standard regression and multi-class classification techniques.

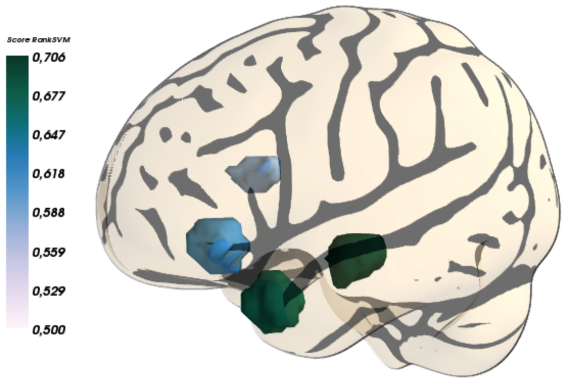


*Figure 8. Based on a ranking procedure, the information present in different regions of interest of the brain volume can be used to predict a cognitive feature, in that case the level of complexity of sentences heared by the subject.*

More details can be found in [27] and [28].

## 6.8. Decoding four letter words from brain activations

**Participants:** Bertrand Thirion, Alexandre Gramfort [Correspondant].

Word reading involves multiple cognitive processes. To infer which word is being visualized, the brain first processes the visual percept, deciphers the letters, bigrams, and activates different words based on context or prior expectation like word frequency. In this contribution, we use supervised machine learning techniques to decode the first step of this processing stream using functional Magnetic Resonance Images (fMRI). We build a decoder that predicts the visual percept formed by four letter words, allowing us to identify words that were not present in the training data. To do so, we cast the learning problem as multiple classification problems after describing words with multiple binary attributes. This work goes beyond the identification or reconstruction of single letters or simple geometrical shapes and addresses a challenging estimation problem, that is the prediction of multiple variables from a single observation, hence facing the problem of learning multiple predictors from correlated inputs.

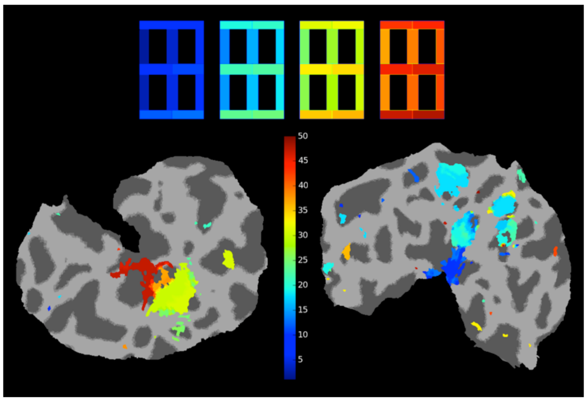


*Figure 9. The bars of a word presented on a fixed visual brain activate specific domains of the visual field, and thus can be decoded through this marked. This makes it possible to identify a four letters word presented on a screen.*

More details can be found in [22].

# SHACRA Project-Team

# 6. New Results

## 6.1. Non-Rigid Augmented Reality for Hepatic Surgery

Hepatic resection and tumors removal approaches remains a major challenge. Despite the use of new minimally invasive techniques which has several advantages such as precision, decreased blood loss, quicker healing time and less pain, the lack of informations due to poor depth perception and direct contact lost leads the surgeons and the research groups to use Augmented Reality to overcome these issues. Augmented Reality is the visual overlay of computers-generated images over real world images. This technique can be used to overlay vessels, tumors and cutting planes performed on the pre-operative data (3D reconstruction from CT or MR scan) onto the laparoscopic video per-operatively. However, current techniques are limited to a rigid registration of the pre-operative liver anatomy onto the intra-operative image, and often this registration is not performed automatically. Our objective is to develop a real-time, non-rigid registration and tracking of the intra and pre-operative liver data.

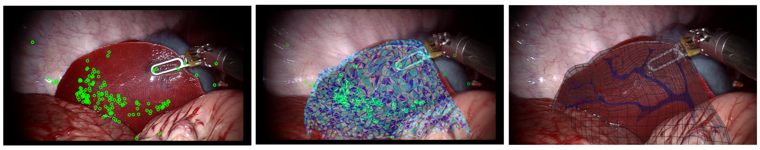


*Figure 5. Non-rigid augmentation of a vascular network of a porcine liver : (left) The liver tracking. (Middle) Biomechanical model of the liver under deformation. (Right) Overlaid vascular network.*

## 6.2. Implicit Modeling of Vascular Trees

Many clinical applications require a vessel segmentation process that is able to both extract the centerline and the surface of the blood vessels. However, noise and topology issues (such as kissing vessels) prevent existing algorithms from being able to easily retrieve such a complex system as the brain vasculature. We propose a new blood vessel tracking algorithm that 1) detect the vessel centerline; 2) provide a local radius estimate; and 3) extracts a dense set of points at the blood vessel surface. This algorithm is based on a RANSAC-based robust fitting of successive cylinders along the vessel. Our method was validated against the Multiple Hypothesis Tracking (MHT) algorithm on 10 3DRA patient data of the brain vasculature. Over 30 blood vessels of various sizes were considered for each patient. Our results demonstrated a greater ability of our algorithm to track small, tortuous and touching vessels (96% success rate), compared to MHT (65% success rate). The computed centerline precision was below 1 voxel when compared to MHT. Moreover, our results were obtained with the same set of parameters for all patients and all blood vessels, except for the seed point for each vessel, also necessary for MHT. The proposed algorithm is thereafter able to extract the full intracranial vasculature with little user interaction.

In the context of computer-based simulation, contact management requires an accurate, smooth, but still efficient surface model for the blood vessels. A new implicit model is proposed, consisting of a tree of local implicit surfaces generated by skeletons (*blobby models*). The surface is reconstructed from data points by minimizing an energy, alternating with an original blob selection and subdivision scheme. The reconstructed models are very efficient for simulation and were shown to provide a sub-voxel approximation of the vessel surface on 5 patients.

## 6.3. Riskmaps in DBS

As discussed in previous sections, Deep Brain Stimulation is a neurosurgical treatment that provides remarkable benefits in neurological movement and affective disorders. It consists in the implantation of a wired electrode deep into the brain. However, the accuracy of the placement is difficult due to brain shifts occuring during the procedure. Due to a potential risk of hemorrhage during the implantation, we specially investigated the brain shift induced motion of the vascular structures. We proposed a method to estimate this motion, based on a physics simulation that consider brain deformation, cerebrospinal fluid and multiple interactions, such as brain-skull contacts etc. The aim is to take it into account during the pre-operative planification step. Thus, we developped a brain-shift aware risk map. It estimate the risk for a trajectory to dissect a vessel. It could help surgeons to choose a safer trajectory for the electrode, and then avoid hemorrhages. The next steps is the use of more complex deformation models.

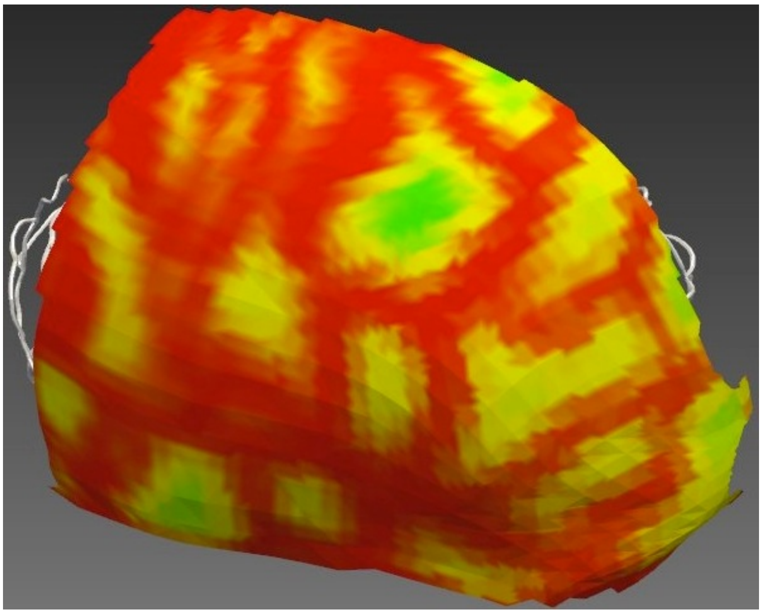


*Figure 6. Brain-shift aware risk map*

## 6.4. Electro Physiology

Cardiac arrhythmia is a very frequent pathology that comes from an abnormal electrical activity in the myocardium. This Ph.D. aims at developing a training simulator for interventional radiology and thermo-ablation of these arrhythmias. The latest improvements lead on electrophysiology simulation (using GPU computing) allowed us to reach real-time performance. The issue of fast electrophysiology was a major bottleneck in the development of our simulator.

This new result enabled us to couple the cardiac eletrophysiology with cardiac mechanical models, thus leading to an interactive framework. Our tractable simulation can therefore simulate a patient-specific electrophysiology and then compute the associated cardiac motion using an electromechanical model.

Moreover, the electrophysiology simulation has been also coupled with a navigation simulation. This is still a work in progress. The implementation of more complex models, such as bidomain models, is also in progress.

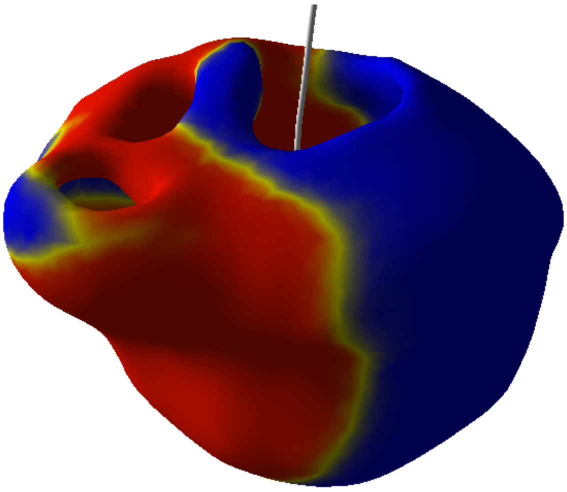


*Figure 7. Cardiac electrophysiology computed on a patient-specific geometry*

## 6.5. Shells

Many tissues in human body have thin structure and may be seen as surfaces or at least be modeled as such. Deformation modeling of surfaces is a topic with wide area of applications especially in computer graphics. However, many of the previously presented techniques are not applicable to the area of surgical simulations where a more physically based approach is desired.

To address this problem we present a new model of shell elements based on the formulation of Bézier triangles. To reduce the number of necessary degrees of freedom a kinematic link between nodes inside the element is defined. Furthermore, using implicit integration scheme allows us to achieve interactive frame rate of the simulation.

The applicability of the model has been validated on a prototype of simulator for preoperative planning of surgery of congenital heart diseases.

## 6.6. Interaction simulation between fluid film and deformable solids

Body fluids are a major constituent of the human body as well by their volume as by their functions. Besides the blood and the lymphatic liquid, many other liquids are present in the body and they have important functions such as lubrication or shock absorption. In this work, we are more particularly interested in the fluids being in the interface between two anatomical structures. We present a method making it possible to simulate the phenomena of interaction between a fluid film and surfaces between which it is forced. The approach that we propose is based on a fluid model and its mechanical coupling with deformable surfaces. According to the pressure of the fluid and the stiffness of the deformable solids in contact with the fluid, various behaviours are expected. Our preliminary results show that it is possible to simulate the main features of these behaviours. Furthermore, the approaches chosen for the fluid model, the deformable model and the coupling between both, are compatible with real time simulations.

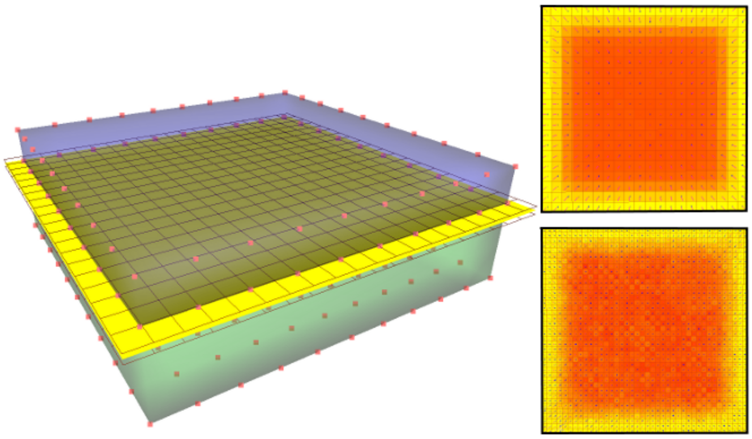


*Figure 8. The fluid is between a rigid solid (green) and a deformable solid (blue). The deformable solid is constraint at the edges. Right: the height map of the fluid (yellow minimum and red maximum height).*

# VISAGES Project-Team

# 6. New Results

## 6.1. Image Segmentation, Registration and Analysis

### *6.1.1. Estimating A Reference Standard Segmentation with Spatially Varying Performance Parameters: Local MAP STAPLE*

**Participant:** Olivier Commowick.

We present a new algorithm, called local MAP STAPLE, to estimate from a set of multi-label segmentations both a reference standard segmentation and spatially varying performance parameters. It is based on a sliding window technique to estimate the segmentation and the segmentation performance parameters for each input segmentation. In order to allow for optimal fusion from the small amount of data in each local region, and to account for the possibility of labels not being observed in a local region of some (or all) input segmentations, we introduce prior probabilities for the local performance parameters through a new maximum a posteriori formulation of STAPLE. Further, we propose an expression to compute confidence intervals in the estimated local performance parameters. We carried out several experiments with local MAP STAPLE to characterize its performance and value for local segmentation evaluation. First, with simulated segmentations with known reference standard segmentation and spatially varying performance, we show that local MAP STAPLE performs better than both STAPLE and majority voting. Then we present evaluations with data sets from clinical applications. These experiments demonstrate that spatial adaptivity in segmentation performance is an important property to capture. We compared the local MAP STAPLE segmentations to STAPLE, and to previously published fusion techniques and demonstrate the superiority of local MAP STAPLE over other state-of-the-art algorithms.

This work was done in collaboration with Alireza Akhondi-Asl and Simon K. Warfield [15].

### *6.1.2. Voxel-based quantitative analysis of brain images from F-18 Fluorodeoxyglucose Positron Emission Tomography with a Block-Matching algorithm for spatial normalization*

**Participant:** Olivier Commowick.

Statistical Parametric Mapping (SPM) is widely used for the quantitative analysis of brain images from F-18 fluorodeoxyglucose positron emission tomography (FDG PET). SPM requires an initial step of spatial normalization to align all images to a standard anatomic model (the template), but this may lead to image distortion and artifacts, especially in cases of marked brain abnormalities. This study aimed at assessing a block-matching (BM) normalization algorithm, where most transformations are not directly computed on the overall brain volume but through small blocks, a principle that is likely to minimize artifacts. Large and/or small hypometabolic areas were artificially simulated in initially normal FDG PET images to compare the results provided by statistical tests computed after either SPM or BM normalization. Results were enhanced by BM, compared with SPM, with regard to (i) errors in the estimation of large defects volumes (about 2-fold lower) because of a lower image distortion, and (ii) rates of false-positive foci when numerous or extended abnormalities were simulated. These observations were strengthened by analyses of FDG PET examinations from epileptic patients. Results obtained with the BM normalization of brain FDG PET appear more precise and robust than with SPM normalization, especially in cases of numerous or extended abnormalities.

This work was done in collaboration with Christophe Person, Valérie Louis-Dorr, Sylvain Poussier, Grégoire Malandain, Louis Maillard, Didier Wolf, Nicolas Gilet, Véronique Roch, Gilles Karcher and Pierre-Yves Marie [19].

### 6.1.3. Block-matching strategies for rigid registration of multimodal medical images

**Participants:** Olivier Commowick, Sylvain Prima.

We propose and evaluate a new block-matching strategy for rigid-body registration of multimodal or multisequence medical images. The classical algorithm first matches points of both images by maximizing the iconic similarity of blocks of voxels around them, then estimates the rigid-body transformation best superposing these matched pairs of points, and iterates these two steps until convergence. In this formulation, only discrete translations are investigated in the block-matching step, which is likely to cause several problems, most notably a difficulty to tackle large rotations and to recover subvoxel transformations. We propose a solution to these two problems by replacing the original, computationally expensive, exhaustive search over translations by a more efficient optimization over rigid-body transformations. The optimal global transformation is then computed based on these local blockwise rigid-body transformations, and these two steps are iterated until convergence. We evaluate the accuracy, robustness, capture range and run time of this new block-matching algorithm on both synthetic and real MRI and PET data, demonstrating faster and better registration than the translation-based block-matching algorithm [28].

### 6.1.4. Automated diffeomorphic registration of anatomical structures with rigid parts: Application to dynamic cervical MRI

**Participants:** Olivier Commowick, Sylvain Prima.

We propose an iterative two-step method to compute a diffeomorphic non-rigid transformation between images of anatomical structures with rigid parts, without any user intervention or prior knowledge on the image intensities. First we compute spatially sparse, locally optimal rigid transformations between the two images using a new block matching strategy and an efficient numerical optimiser (BOBYQA). Then we derive a dense, regularised velocity field based on these local transformations using matrix logarithms and M-smoothing. These two steps are iterated until convergence and the final diffeomorphic transformation is defined as the exponential of the accumulated velocity field. We show our algorithm to outperform the state-of-the-art log-domain diffeomorphic demons method on dynamic cervical MRI data [27].

### 6.1.5. Computer-assisted paleoneurology

**Participant:** Sylvain Prima.

In collaboration with Antoine Balzeau and colleagues at the MNHN (http://www.mnhn.fr), we made the first ever description of the “digital” endocranial cast of the Cro-Magnon 1 specimen, discovered in 1868 at the Eyzies-de-Tayac, Dordogne, France [13]. Together with Benoit Combès (Géosciences Rennes, UMR 6118), we were especially involved in the assessment of its endocranial asymmetries, using an algorithm previously developed at VisAGeS [51] in the context of the ARC 3D-MORPHINE coordinated by Sylvain Prima (http://3dmorphine.inria.fr).

## 6.2. Image processing on Diffusion Weighted Magnetic Resonance Imaging

### 6.2.1. Non-Local Robust Detection of DTI White Matter Differences with Small Databases

**Participants:** Olivier Commowick, Aymeric Stamm.

Diffusion imaging, through the study of water diffusion, al- lows for the characterization of brain white matter, both at the population and individual level. In recent years, it has been employed to detect brain abnormalities in patients suffering from a disease, e.g. from multiple sclerosis (MS). State-of-the-art methods usually utilize a database of matched (age, sex, ...) controls, registered onto a template, to test for differences in the patient white matter. Such approaches however suffer from two main drawbacks. First, registration algorithms are prone to local errors, thereby degrading the comparison results. Second, the database needs to be large enough to obtain reliable results. However, in medical imaging, such large databases are hardly available. In this paper, we propose a new method that addresses these two issues. It relies on the search for samples in a local neighborhood of each pixel to increase the size of the database. Then, we propose a new test based on these samples to perform a voxelwise comparison of a patient image with respect to a population of controls. We demonstrate on simulated and real MS patient data how such a framework allows for an improved detection power and a better robustness and reproducibility, even with a small database [26].

### 6.2.2. Registration and Analysis of White Matter Group Differences with a Multi-Fiber Model

**Participant:** Olivier Commowick.

Diffusion magnetic resonance imaging has been used extensively to probe the white matter in vivo. Typically, the raw diffusion images are used to reconstruct a diffusion tensor image (DTI). The incapacity of DTI to represent crossing fibers leaded to the development of more sophisticated diffusion models. Among them, multi-fiber models represent each fiber bundle independently, allowing the direct extraction of diffusion features for population analysis. However, no method exists to properly register multi-fiber models, seriously limiting their use in group comparisons. This paper presents a registration and atlas construction method for multi-fiber models. The validity of the registration is demonstrated on a dataset of 45 subjects, including both healthy and unhealthy subjects. Morphometry analysis and tract-based statistics are then carried out, proving that multi-fiber models registration is better at detecting white matter local differences than single tensor registration.

This work was done in collaboration with Maxime Taquet, Benoit Scherrer, Jurriaan Peters, Mustafa Sahin, Benoît Macq and Simon K. Warfield [44].

### 6.2.3. Automated delineation of white matter fiber tracts with a multiple region-of-interest approach

**Participant:** Olivier Commowick.

White matter fiber bundles of the brain can be delineated by tractography utilizing multiple regions-of-interest (MROI) defined by anatomical landmarks. These MROI can be used to specify regions in which to seed, select, or reject tractography fibers. Manual identification of anatomical MROI enables the delineation of white matter fiber bundles, but requires considerable training to develop expertise, considerable time to carry out and suffers from unwanted inter- and intra-rater variability. In a study of 20 healthy volunteers, we compared three methodologies for automated delineation of the white matter fiber bundles. Using these methodologies, fiber bundle MROI for each volunteer were automatically generated. We assessed three strategies for inferring the automatic MROI utilizing nonrigid alignment of reference images and projection of template MROI. We assessed the bundle delineation error associated with alignment utilizing T1-weighted MRI, fractional anisotropy images, and full tensor images. We confirmed the smallest delineation error was achieved using the full tensor images. We then assessed three projection strategies for automatic determination of MROI in each volunteer. Quantitative comparisons were made using the root-mean-squared error observed between streamline density images constructed from fiber bundles identified automatically and by manually drawn MROI in the same subjects. We demonstrate that a multiple template consensus label fusion algorithm generated fiber bundles most consistent with the manual reference standard.

This work was done in collaboration with Ralph Suarez, Sanjay Prabhu and Simon K. Warfield [23].

### 6.2.4. Corticospinal tractography with morphological, functional and diffusion tensor MRI: a comparative study of four deterministic algorithms used in clinical routine

**Participants:** Sylvain Prima, Camille Maumet, Jean-Christophe Ferré.

In collaboration with Romuald Seizeur, Nicolas Wiest-Daesslé and Xavier Morandi, we aimed to compare four deterministic tractography algorithms used in clinical routine for the study of the corticospinal tract (the bundle mediating voluntary movement) in 15 right-handed volunteers. We found no difference between right and left sides of the brain for any of the algorithms [22].

### 6.2.5. A new multi-directional fiber model for low angular resolution diffusion imaging

**Participants:** Aymeric Stamm, Christian Barillot.

Diffusion MRI is a tool of choice for the analysis of the brain white matter fiber pathways. When translated to clinics, the short acquisition time leads to low angular resolution diffusion (LARD) images. Fiber pathways are then inferred assuming Gaussian diffusion (a.k.a. DTI) that provides one fiber orientation per voxel. In the past decade, recent researches highlight more intricate intra-voxel fiber configurations using higher angular resolution diffusion images. In collaboration with Patrick Perez (Technocolor), we have proposed a non-Gaussian diffusion model of the white matter fibers able to recover from crossing of fibers even from low angular resolution.This model enables crossing fibers to be theoretically estimated from only 8 diffusion MR images. In particular, this model allows for the retrospective study of DW data sets acquired in the past. [42] [43].

## 6.3. Medical Image Computing in Brain Pathologies

### *6.3.1. Detection of dysplasia and heterotopia*

**Participants:** Elise Bannier, Camille Maumet, Jean-Christophe Ferré, Christian Barillot.

Focal cortical dysplasia and heterotopias are a recognized cause of epilepsy. Indication for surgery relies on precise localization and delineation. However, visual depiction of focal cortical dysplasia and heterotopias is difficult, time-consuming and reader dependant. Several 3D T1 voxel based morphometry methods have been proposed to automatically identify and suggest potential abnormalities to the reader. Several studies have shown the ability of Double Inversion Recovery imaging to detect intracortical lesions in MS and Epilepsy. In this study we propose to evaluate the ability of Double Inversion Recovery voxel based analysis to detect cortical and juxtacortical lesions in pharmaco resistant epileptic patients. This work was performed in collaboration with Arnaud Biraben, Anca Pasnicu and Eduardo Pasqualini, Béatrice Carsin-Nicol [24].

### *6.3.2. MRI Estimation of $T_1$ Relaxation Time Using a Constrained Optimization Algorithm*

**Participants:** Fang Cao, Olivier Commowick, Elise Bannier, Jean-Christophe Ferré, Gilles Edan, Christian Barillot.

We propose a new method to improve $T_1$ mapping with respect to the popular *DESPOT1* algorithm. A distance function is defined to model the distance between the pure signal and the measurements in presence of noise. We use a constrained gradient descent optimization algorithm to iteratively find the optimal values of $T_1$ and $M_0$. The method is applied to MR images acquired with 2 gradient echo sequences and different flip angles. The performance of $T_1$ mapping is evaluated both on phantom and on in vivo experiments [25].

### *6.3.3. Characterization and Modeling of Multidimensional MRI signatures in Multiple Sclerosis in clinically isolated syndromes.*

**Participants:** Yogesh Karpate, Olivier Commowick, Gilles Edan, Christian Barillot.

Clinically Isolated Syndrome data contribute to critical factors in obtaining meaningful precursor and predictors of Multiple Sclerosis. Current methodologies don’t go beyond segmentation and which generalize poorly over multi-modal MRI data. The project objective is to research and develop a framework for characterization and modeling of multidimensional MRI signatures in clinically isolated syndrome(disease’s onset),based on earlier and concurrent research and developments in the lab. In on going work an algorithmic framework is being developed to address the MS lesions’ classification ,identification and retrieval in USPIO-6 database.

As part of a battery of pre- processing techniques ,the module for intensity normalization of MRI volumes based on Spatio-Temporal Robust Expectation Maximization (STREM) is developed. This work is primarily based on 3 MRI modalities viz T1-w,T2-w and FLAIR. Complementary to this work , an another intensity normalization algorithm is devised based on parametric robust as well as efficient estimation by minimizing a density power divergence (beta divergence).The proposed method is indexed by a single parameter alpha which controls the trade-off between robustness and efficiency. The methodology affords a robust extension of maximum likelihood estimation for which alpha tends to be zero. Choices of alpha near zero afford considerable robustness while retaining efficiency close to that of maximum likelihood.

Moving forward, to facilitate accurate lesion tracking,features must be selected which are robust to photometric and geometric distortions. Energy measures are used to capture lesion's multiscale orientation structure in space.To illustrate utility with respect to a lesion detection, we have developed descriptor like local energy based on 3D steerable wavelets. This will be followed by the rigorous empirical evaluations of the resulting algorithm yielding better lesion identification and retrieval.

### *6.3.4. Multiple Sclerosis Lesions Evolution in Patients with Clinically Isolated Syndrome.*

**Participants:** Alessandro Crimi, Olivier Commowick, Gilles Edan, Christian Barillot.

Multiple sclerosis (MS) is a disease with heterogeneous evolution among the patients. Some classifications have been carried out according to either the clinical course or the immunopathological profiles. Epidemiological data and imaging are showing that MS is a two-phase neurodegenerative in inflammatory disease. At the early stage it is dominated by focal in inflammation of the white matter (WM), and at a latter stage it is dominated by diffuse lesions of the grey matter and spinal cord. A Clinically Isolated Syndrome (CIS) is a first neurologic episode caused by in inflammation/demyelination in the central nervous system which may lead to MS. Few studies have been carried out so far about this initial stage. Better understanding of the disease at its onset will lead to a better discovery of pathogenic mechanisms, allowing suitable therapies at an early stage. We propose a new data processing framework able to provide an early characterization of CIS patients according to lesion patterns, and more specifically according to the nature of the inflammatory patterns of these lesions. Our method is based on a two layers unsupervised clustering. Initially, the spatio-temporal lesion patterns are classified using a tensor-like representation. The discovered lesion patterns are then used to identify group of patients and their correlation to one year follow-up total lesion loads, which is so far the only image-based figure that can potentially correlate to future evolution of the pathology. We expect that the proposed framework can infer new prospective figures from the earliest imaging sign of MS since it can provide a classification of different types of lesion across patients [30].

## 6.4. Vascular Imaging and Arterial Spin Labelling

### *6.4.1. Robust Cerebral Blood Flow Map Estimation in Arterial Spin Labeling*

**Participants:** Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Non-invasive measurement of Cerebral Blood Flow (CBF) is now feasible thanks to the introduction of Arterial Spin Labeling (ASL) Magnetic Resonance Imaging (MRI) techniques. To date, the low signal-to-noise ratio of ASL gives us no option but to repeat the acquisition in order to accumulate enough data to get a reliable signal. Perfusion signal is usually extracted by averaging across the repetitions. However, due to its zero breakdown point, the sample mean is very sensitive to outliers. A single outlier can thus have strong detrimental effects on the sample mean estimate. In this paper, we propose to estimate robust ASL CBF maps by means of M-estimators to overcome the deleterious effects of outliers. The behavior of this method is compared to z-score thresholding as recommended in [8]. validation on simulated and real data is provided. Quantitative validation is undertaken by measuring the correlation with the most widespread technique to measure perfusion with MRI: Dynamic Susceptibility weighted Contrast (DSC) [37].

### *6.4.2. A comprehensive framework for the detection of individual brain perfusion abnormalities using Arterial Spin Labeling*

**Participants:** Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Arterial Spin Labeling (ASL) enables measuring cerebral blood flow in MRI without injection of a contrast agent. Perfusion measured by ASL carries relevant information for patients suffering from pathologies associated with singular perfusion patterns. However, to date, individual identification of abnormal perfusion patterns in ASL usually relies on visual inspection or manual delineation of regions of interest. In this paper, we introduce a new framework to automatically outline patterns of abnormal perfusion in individual patients by means of an ASL template. We compare two models of normal perfusion and assess the quality of detections comparing an a contrario approach to the Generalized Linear Model (GLM) [33], [35], [36].

### 6.4.3. Using Negative Signal in Mono-TI Pulsed Arterial Spin Labeling to Outline Pathological Increases in Arterial Transit Times

**Participants:** Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Elise Bannier, Christian Barillot.

The presence of unexpected negative perfusion estimates has been sparsely discussed in the ASL literature. In the study of perfusion maps extracted from a single inversion time in ASL (mono-TI ASL), it is however common to deal with areas of significant negative signal. This is problematic since performing statistical analysis based on this data might therefore lead to inacurrate results. Though isolated negative values could be attributed to noise, clusters of significant negative signal should be explained by another phenomenon. Following [2], which outlined that negative values might arise due to increased transit times, we investigated this hypothesis based on real clinical datasets including healthy control and patient data [34].

### 6.4.4. An a contrario approach for the detection of activated brain areas in fMRI

**Participants:** Camille Maumet, Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

BOLD functional MRI (fMRI) is now a widespread imaging technique to study task-related activity in the brain. However, getting the areas of activation at the individual subject level is still an open issue. The standard massively univariate statistical analysis is usually performed after smoothing the data and makes use of a single p-value for final thresholding of the results. In group fMRI studies, the need for compensation of cross-subjects misregistrations clearly justifies the smoothing. However, at the individual level, where neat delineations of the activated areas are of interest, the use of gaussian smoothing as a pre-processing step is more questionable. In this paper, we propose to study the ability of an a contrario approach, recently adapted for basal perfusion abnormalities detection, to correctly detect areas of functional activity.

### 6.4.5. Compressive Matched Filter for Cerebral Blood Flow Quantification with ASL: sampling diversity or repetition?

**Participants:** Lei Yu, Pierre Maurel, Christian Barillot.

The Arterial Spin Labeling (ASL) is an MRI (Magnetic Resonance Imaging)-based perfusion technique which uses the magnetically tagged water as a freely diffusible tracer to measure perfusion non-invasively. This blood water is first labeled with a radio-frequency pulse in the neck of the patient. After a delay, called Inversion Time (TI), which allows the labeled blood to arrive in the brain, a labeled image of the brain is acquired. A control image is also acquired without labeling and the CBF (Cerebral Brain Flow) estimation is done on the difference between the control and labeled image. Classical method, Mono-TI, for CBF quantification is averaging repetitions with only one Inversion Time (TI) - the time delay between labeling and acquisition to allow the labeled blood to arrive the imaging slice. It improves the robustness to noise, however, cannot compensate the variety of Arterial Arrival Time (AAT).

In this work [45], Diverse-TI is proposed to exploit different TI sampling instants (sampling diversity) to improve the robustness to variety of AAT and simultaneously average repetitions with each TI (sampling repetitions) to improve the robustness to noise. Generally, the sampling diversity is relatively small and can be considered as compressed measurements, thus the Compressive Matched Filter (CMF) enlightened from sparsity is exploited to directly reconstruct CBF and AAT directly from compressed measurements. Meanwhile, regarding the CBF quantification performance, the compromise between the sampling repetition and sampling diversity is discussed and the empirical protocol to determine the sampling diversity is proposed.

The future works will consist in applying the parameter design protocol to guide the Diverse-TI technique in real ASL data acquisitions. Meanwhile, it is possible to extend CMF algorithm by considering additional priors to regularize the CBF estimation problem which might also improve the performance.

This work was done in collaboration with Remi Gribonval (Metiss team) [45].

### 6.4.6. Non-contrast enhanced neurovascular imaging

**Participants:** Elise Bannier, Hélène Raoult, Jean-Yves Gauvrit.

Detecting internal carotid artery (ICA) stenosis is a main challenge for the prevention of stroke, the third leading cause of death in the developed world. Novel non-contrast-enhanced MRA (NCE MRA) sequences have emerged as an alternative to traditional MRA approaches, especially for patients during pregnancy or with renal insufficiency.

Up to now, the inversion-prepared bSSFP NCE MRA approach has been applied to imaging of renal arteries or kidney transplants and only few studies focused on the ICA, using ECG-gating. The purpose of this first study was to assess the feasibility and image quality of an improved non-gated carotid NATIVE TrueFISP NCE MRA sequence providing an extended field of view as compared to Time-of-Flight (TOF) imaging. Sixteen healthy volunteers were included to evaluate different sequence parameter sets. In comparison to standard TOF, the used NCE MRA sequence offered equivalent to higher image quality along with larger coverage and shorter acquisition times. Improved image quality was achieved without ECG gating, which had been used in previous studies. A Partial Fourier scheme with an early acquisition of k-space center yielded higher image quality and signal intensity compared to a late acquisition.

A second study evaluated the non-contrast-enhanced ECG-gated 4D MRA combining arterial spin labeling (ASL) and bSSFP readout (bSSFP NCE 4D MRA) sequence to non invasively investigate morphological and hemodynamic patterns of cerebral arteriovenous malformations (AVM). Previous studies have shown high temporal resolution (50-100 ms), yet with temporal windows limited to a single cardiac cycle. This precludes the complete venous drainage analysis, which is necessary to evaluate AVM hemorrhagic risk. This study aimed at assessing the feasibility, quality and diagnosis performance of a bSSFP NCE 4D MRA sequence with a large acquisition time window over 2 cardiac cycles (2 RR) without a significant reduction of spatial resolution. Ten patients presenting AVM and referred to digital subtraction angiography (DSA) were included in the study. The 2-RR bSSFP NCE 4D MRA sequence yielded an image quality comparable to that of a corresponding 1-RR acquisition. AVM analysis, however, was improved due to a better depiction of venous drainage necessary to evaluate hemorrhagic risk. The simultaneous high-resolution morphologic and hemodynamic data also offered an especially accurate delineation of the nidus, target of the treatment.

### *6.4.7. ASLDEM : Arterial Spin Labeling At 3t In Semantic Dementia: Perfusion Abnormalities Detection And Comparison With Fdg-pet*

**Participants:** Isabelle Corouge, Jean-Christophe Ferré, Elise Bannier, Christian Barillot, Jean-Yves Gauvrit.

Arterial Spin Labeling (ASL) is a non invasive perfusion imaging technique which has shown great diagnosis potential in dementia. However, it has never been applied to semantic dementia (SD), a rare subtype of frontotemporal lobar degeneration characterized by the gradual loss of conceptual knowledge, which is actually explored by a now well established marker of SD: $^{18}$fluorodeoxyglucose-positron emission tomography (FDG-PET) imaging. Although ASL and FDG-PET respectively measure perfusion and metabolism, they have been shown to be strongly correlated. In this project, we explore the ability of ASL to detect perfusion abnormalities in SD in comparison with FDG-PET. We apply our analysis framework (implemented as part of the 'autoasl' and 'autoasltemplate' softwares) on patients and healthy subjects data from an ongoing clinical study with a focus on ASL data preprocessing and statistical analysis at the individual and group level. Preliminary results yield concordant observations between ASL and FDG-PET as well as expected hypoperfusions in SD, namely in the left temporal lobe, thus suggesting the potential of ASL to assess perfusion impairments in SD [29].

For this work, Aurore Esquevin was awarded the prize "Communication Jeune Chercheur 2012" at the "Journées Françaises de Radiologie (JFR)" conference.

## 6.5. Abnormal functional lateralization and activity of language brain areas in developmental dysphasia

### *6.5.1. Statistical analysis of white matter integrity for the clinical study of specific language impairment in children*

**Participants:** Olivier Commowick, Aymeric Stamm, Camille Maumet, Jean-Christophe Ferré, Clément De Guibert, Christian Barillot.

Children affected by Specific Language Impairment (SLI) fail to develop a normal language capability. To date, the etiology of SLI remains largely unknown. It induces difficulties with oral language which cannot be directly attributed to intellectual deficit or other developmental delay. Whereas previous studies on SLI focused on the psychological and genetic aspects of the pathology, few imaging studies investigated defaults in neuroanatomy or brain function. We propose to investigate the integrity of white matter in Specific Language Impairment thanks to diffusion Magnetic Resonance Imaging. An exploratory analysis was performed without a priori on the impaired regions. A region of interest statistical analysis was performed based, first, on regions defined from Catani's atlas and, then, on tractography-based regions. Both the mean fractional anisotropy and mean apparent diffusion coefficient were compared across groups. To the best of our knowledge, this is the first study focusing on white matter integrity in specific language impairment. 22 children with SLI and 19 typically developing children were involved in this study. Overall, the tractography-based approach to group comparison was more sensitive than the classical ROI-based approach. Group differences between controls and SLI patients included decreases in FA in both the perisylvian and ventral pathways of language, comforting findings from previous functional studies. This work was performed in collaboration with Emmanuel Vallée, Clément de Guibert, Catherine Allaire and Elisabeth Le Rumeur.

# CLIME Project-Team

# 6. New Results

## 6.1. New methods for data assimilation

Since the beginning, Clime is focused on developing new techniques for data assimilation in geophysical sciences. Clime is active on several of the most challenging theoretical aspects of data assimilation: data assimilation methods based on non-Gaussian assumptions, methods for estimating errors, ensemble filtering techniques, 4D variational assimilation approaches, ensemble-variational methods, etc. This year, we revisited several of these topics. A dual algorithm has been developed for the finite-size ensemble Kalman filter, that shows how to estimate optimal inflation that counteracts sampling errors. A variational method coupled to a subgrid scale statistical model has been introduced and validated to quantify the representativeness errors. We also started to work on ensemble variational methods that are brand new techniques emerging in the meteorological data assimilation field.

### *6.1.1. Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems*

**Participants:** Marc Bocquet, Pavel Sakov [NERSC, Norway].

The finite-size ensemble Kalman filter (EnKF-N) is an ensemble Kalman filter (EnKF) which, in perfect model condition, does not require inflation because it partially accounts for the ensemble sampling errors. For the Lorenz ’63 and ’95 toy-models, it was so far shown to perform as well or better than the EnKF with an optimally tuned inflation. The iterative ensemble Kalman filter (IEnKF) is an EnKF which was shown to perform much better than the EnKF in strongly nonlinear conditions, such as with the Lorenz ’63 and ’95 models, at the cost of iteratively updating the trajectories of the ensemble members. This study aims at further exploring the two filters, and at combining both into an EnKF that does not require inflation in perfect model condition and which is as efficient as the IEnKF in very nonlinear conditions.

In this study EnKF-N is first introduced and a new implementation is developed. It decomposes EnKF-N into a cheap two-step algorithm that amounts to computing an optimal inflation factor. This offers a justification of the use of the inflation technique in the traditional EnKF and why it can often be efficient. Secondly, the IEnKF is introduced following a new implementation based on the Levenberg-Marquardt optimization algorithm. Then, the two approaches are combined to obtain the finite-size iterative ensemble Kalman filter (IEnKF-N). Several numerical experiments are performed on IEnKF-N with the Lorenz ’95 model. These experiments demonstrate its numerical efficiency as well as its performance that offer, at least, the best of both filters.

### *6.1.2. Accounting for representativeness errors in the inversion of atmospheric constituent emissions: Application to the retrieval of regional carbon monoxide fluxes*

**Participants:** Mohammad Reza Koohkan, Marc Bocquet.

A four-dimensional variational data assimilation system (4D-Var) is developed to retrieve carbon monoxide (CO) fluxes at regional scale, using an air quality network. The air quality stations that monitor CO are proximity stations located close to industrial, urban or traffic sources. The mismatch between the coarsely discretized Eulerian transport model and the observations, inferred to be mainly due to representativeness errors in this context, leads to a bias (averaged simulated concentrations minus observed concentrations) of the same order of magnitude as the concentrations. 4D-Var leads to a mild improvement in the bias because it does not adequately handle the representativeness issue. For this reason, a simple statistical subgrid model is introduced and is coupled to 4D-Var. In addition to CO fluxes, the optimization seeks to jointly retrieve *influence coefficients*, which quantify each station’s representativeness. The method leads to a much better representation of the CO concentration variability, with a significant improvement of statistical indicators.

The resulting increase in the total inventory estimate is close to the one obtained from remote sensing data assimilation. This methodology and experiments suggest that information useful at coarse scales can be better extracted from atmospheric constituent observations strongly impacted by representativeness errors.

### *6.1.3. Real-time data assimilation*

**Participants:** Vivien Mallet, Anne Tilloy, Fabien Brocheton [Numtech], David Poulet [Numtech], Cécile Honoré [Airparif], Édouard Debry [INERIS].

Based on Verdandi, Polyphemus and the "Urban Air Quality Analysis" software, real-time data assimilation was carried out at urban scale. The Best Linear Unbiased Estimator (BLUE) was computed for every hourly concentration map that the ADMS model computed. A posteriori tests were conducted over Clermont-Ferrand and Paris. We addressed the key issue of the covariance of the state error. The form of the error covariance between two points was determined based on the road network, considering the distance between points along the road and the distance of each point to the road. A few parameters (primarily two decorrelation lengths) were determined thanks to cross validation with several months of simulations and observations. The results showed strong improvements even at locations where no data was assimilated.

At larger scale, the data assimilation library Verdandi was used to apply data assimilation (optimal interpolation) with the air quality model Chimere. This preliminary work will help INERIS to apply optimal interpolation for ozone and particulate matter in the operational platform Prev'air.

## 6.2. Inverse modeling

Many of this year's studies have focused on inverse modeling, including the reconstruction of the Fukushima radionuclide atmospheric and marine source terms. All were targeted to a particular application. However most of them include new methodological developments, in particular non-Gaussian data assimilation schemes.

### *6.2.1. Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the Fukushima Daiichi source term*

**Participants:** Victor Winiarek, Marc Bocquet, Olivier Saunier [IRSN], Anne Mathieu [IRSN].

The aim of this research activity is the implementation of data assimilation methods, particularly inverse modeling methods, in the context of an accidental radiological release from a nuclear power plant and their application in the specific case of the Fukushima Daiichi accident. The particular methodological focus is the a posteriori estimation of the prior errors statistics. In the case of the Fukushima Daiichi accident, the number of available observations is small compared to the number of source parameters to retrieve and the reconstructed source is highly sensitive to the prior errors. That is the why they need to be well established and justified. In this aim, three methods have been proposed: one method relies on a L-curve estimation technique, another one on the Desroziers' iterative scheme and the last method, assumed to be the most robust, relies on the maximum likelihood principle, generalised to a non-Gaussian context. These three methods have been applied to the reconstruction of cesium-137 and iodine-131 source terms from the Fukushima Daiichi accident. Because of the poor observability of the Fukushima Daiichi emissions, these methods provide lower-bounds for cesium-137 and iodine-131 reconstructed activities. Nevertheless, with the new method based on semi-Gaussian statistics for the background errors, the lower-bound estimates for cesium-137, $1.2 - 4.0\,10^{16}$ Bq with an estimated standard deviation range of $15 - 20\%$, and for iodine-131, $1.9 - 3.8\,10^{17}$ Bq with an estimated standard deviation range of $5 - 10\%$, are of the same order of magnitude as those provided by the Japanese Nuclear and Industrial Safety Agency, and about 5 to 10 times less than the Chernobyl atmospheric releases.

### *6.2.2. Assessment of the amount of Cesium-137 released into the Pacific Ocean after the Fukushima accident and analysis of its dispersion in Japanese coastal waters*

**Participants:** Claude Estournel [LA], Emmanuel Bosc [IAEA], Marc Bocquet, Caroline Ulses [LA], Patrick Marsailex [LA], Victor Winiarek, Iolanda Osvath [IAEA], Cyril Nguyen [LA,LEGOS], Thomas Duhaut [LA], Florent Lyard [LEGOS], Héloïse Michaud [LA], Francis Auclair [LA].

Numerical modeling was used to provide a new estimate of the amount of cesium-137 released directly into the ocean from the Fukushima Daiichi nuclear power plant (NPP) after the accident in March 2011 and to gain insights into the physical processes that led to its dispersion in the marine environment during the months following the accident. An inverse method was used to determine the time-dependent cesium-137 input responsible for the observed concentrations. The method was then validated through comparisons of the simulated concentrations with concentrations measured in seawater at different points in the neighborhood of the nuclear power plant. An underestimation was noticed for stations located 30 km offshore. The resulting bias in the release inventory was estimated. Finally, the maximum cesium-137 activity released directly to the ocean was estimated to lie between 5.1 and 5.5 PBq (Peta Becquerel = 1015 Bq) but uncertainties remain on the amount of radionuclides released during the first few days after the accident. This estimate was compared to previous ones and differences were further analysed. The temporal and spatial variations of the cesium-137 concentration present in the coastal waters were shown to be strongly related to the wind intensity and direction. During the first month after the accident, winds blowing toward the south confined the radionuclides directly released into the ocean to a narrow coastal band. Afterwards, frequent northward wind events increased the dispersion over the whole continental shelf, leading to strongly reduced concentrations.

### 6.2.3. What eddy-covariance measurements tell us about prior land flux errors in CO2-flux inversion schemes?

**Participants:** Frédéric Chevallier [LSCE], Tao Wang [LSCE], Philippe Ciais [LSCE], Marc Bocquet, Altaf Arain [McMaster University, Canada], Alessandro Cescatti [Joint Research Centre, Italy], Jiquan Chen [University of Toledo, USA], Johannes Dolman [Vrije Universiteit, the Netherlands], Beverly Law [Oregon State University, USA], Hank Margolis [Université Laval, Canada], Leonardo Montagnani [University of Bolzano, Italy].

To guide the future development of CO2-atmospheric inversion modeling systems, we analysed the errors arising from prior information about terrestrial ecosystem fluxes. We compared the surface fluxes calculated by a process-based terrestrial ecosystem model with daily averages of CO2 flux measurements at 156 sites across the world in the FLUXNET network. At the daily scale, the standard deviation of the model-data fit was 2.5 gC·m-2·d-1; temporal autocorrelations were significant at the weekly scale (> 0.3 for lags less than four weeks), while spatial correlations were confined to within the first few hundred kilometers (< 0.2 after 200 km). Separating out the plant functional types did not increase the spatial correlations, except for the deciduous broad-leaved forests. Using the statistics of the flux measurements as a proxy for the statistics of the prior flux errors was shown not to be a viable approach. A statistical model allowed us to upscale the site-level flux error statistics to the coarser spatial and temporal resolutions used in regional or global models. This approach allowed us to quantify how aggregation reduces error variances, while increasing correlations. As an example, for a typical inversion of grid point (300 km × 300 km) monthly fluxes, we found that the prior flux error follows an approximate e-folding correlation length of 500 km only, with correlations from one month to the next as large as 0.6.

## 6.3. Monitoring network design

In this section, we report studies that are related to the evaluation of monitoring networks and to new monitoring strategies. This year, network designs techniques have been applied to the inverse modeling of CO2 fluxes.

### 6.3.1. Network design for mesoscale inversions of CO2 sources and sinks

**Participants:** Thomas Lauvaux [Pennsylvania State University, USA], Andy Schuh [Colorado State University, USA], Marc Bocquet, Lin Wu, Scott Richardson [Pennsylvania State University, USA], Natasha Miles [Pennsylvania State University, USA], Ken Davis [Pennsylvania State University, USA].

Recent instrumental deployments of regional observation networks of atmospheric CO2 mixing ratios have been used to constrain carbon sources and sinks using inversion methodologies. In this study, we performed sensitivity experiments using observation sites from the Mid Continent Intensive experiment to evaluate the required spatial density and locations of CO2 concentration towers based on flux corrections and error reduction analysis. In addition, we investigated the impact of prior flux error structures with different correlation lengths and biome information. We show that, while the regional carbon balance converged to similar annual estimates using only two concentration towers over the region, additional sites were necessary to retrieve the spatial flux distribution of our reference case (using the entire network of eight towers). Local flux corrections required the presence of observation sites in their vicinity, suggesting that each tower was only able to retrieve major corrections within a hundred of kilometers around, despite the introduction of spatial correlation lengths ( 100 to 300 km) in the prior flux errors. We then quantified and evaluated the impact of the spatial correlations in the prior flux errors by estimating the improvement in the CO2 model-data mismatch of the towers not included in the inversion. The overall gain across the domain increased with the correlation length, up to 300 km, including both biome-related and non-biome-related structures. However, the spatial variability at smaller scales was not improved. We conclude that the placement of observation towers around major sources and sinks is critical for regional-scale inversions in order to obtain reliable flux distributions in space. Sparser networks seem sufficient to assess the overall regional carbon budget with the support of flux error correlations, indicating that regional signals can be recovered using hourly mixing ratios. However, the smaller spatial structures in the posterior fluxes are highly constrained by assumed prior flux error correlation lengths, with no significant improvement at only a few hundreds of kilometers away from the observation sites.

### 6.3.2. Potential of the International Monitoring System radionuclide network for inverse modeling

**Participants:** Mohammad Reza Koohkan, Marc Bocquet, Lin Wu, Monika Krysta [The Preparatory Commission for the Comprehensive Nuclear Test-Ban Treaty Organization, UNO].

The International Monitoring System (IMS) radionuclide network enforces the Comprehensive Nuclear-Test-Ban Treaty, which bans nuclear explosions. We have evaluated the potential of the IMS radionuclide network for inverse modeling of the source, whereas it is usually assessed by its detection capability. To do so, we have chosen the *degrees of freedom for the signal* (DFS), a well established criterion in remote sensing, in order to assess the performance of an inverse modeling system. Using a multiscale data assimilation technique, we have computed optimal adaptive grids of the source parameter space by maximizing the DFS. This optimization takes into account the monitoring network, the meteorology over one year (2009) and the relationships between the source parameters and the observations derived from the FLEXPART Lagrangian transport model. Areas of the domain, where the grid-cells of the optimal adaptive grid are large, emphasize zones where the retrieval is more uncertain, whereas areas, where the grid-cells are smaller and denser, stress regions where more source variables can be resolved. The observability of the globe through inverse modeling is studied in strong, realistic and small model error cases. The strong error and realistic error cases yield heterogeneous adaptive grids, indicating that information does not propagate far from the monitoring stations, whereas in the small error case, the grid is much more homogeneous.

In all cases, several specific continental regions remain poorly observed such as Africa as well as the tropics, because of the trade winds.

The northern hemisphere is better observed through inverse modeling (more than 60% of the total DFS), mostly because it contains more IMS stations. This unbalance leads to a better performance of inverse modeling in the northern hemisphere winter. The methodology is also applied to the subnetwork composed of the stations of the IMS network that measure noble gases.

## 6.4. Reduction and emulation

The use of environmental models raise a number of problems due to:

- the dimension of their inputs, which can easily be $10^5 - -10^8$ at every time step;

- the dimension of their state vector, which is usually $10^5 - -10^7$;
- their high computational cost.

In particular, the application of data assimilation methods and uncertainty quantification techniques may require dimension reduction and cost reduction. The dimension reduction consists in projecting the inputs and the state vector to low-dimensional subspaces. The cost reduction can be carried out by emulation, i.e., the replacement of costly components with fast surrogates.

### 6.4.1. Reduction and emulation of a chemistry-transport model

**Participants:** Vivien Mallet, Serge Guillas [University College London].

Both reduction and emulation were applied to the dynamic air quality model Polair3D from Polyphemus. The reduction relied on proper orthogonal decomposition on the input data and on the state vector. The dimension of the reduced subspace for the input data is about 80, while the dimension of the reduced state vector is less than 10. The projection of the state vector on its reduced subspace can be carried out before every integration time step, so that one can reproduce a full state trajectory (in time) using the reduced model.

Significant advances were made to emulate the reduced model, which requires about 90 inputs (reduced input data and reduced state vector) and computes about 10 outputs (reduced state vector). 90 inputs is however a large number to build an emulator using a classical approach like krigging. Promising results were however obtained with an interpolation method based on inverse distance weighting.

### 6.4.2. Reduction and emulation of a static air quality model

**Participants:** Vivien Mallet, Anne Tilloy, Fabien Brocheton [Numtech], David Poulet [Numtech].

The dimension reduction was applied to the outputs of the urban air quality model ADMS Urban, which is a static model with low-dimensional inputs and high-dimensional outputs. A proper orthogonal decomposition on the outputs allowed us to drastically reduce their dimension, from $10^4$ to just a few scalars. First attempts of emulation of the reduced model rely on Gaussian process emulation.

### 6.4.3. Motion estimation from images with a wavelets reduced model

**Participants:** Giuseppe Papari, Isabelle Herlin, Etienne Huot, Karim Drifi.

The dimension reduction was applied to an image model, composed of Lagrangian constancy of velocity and transport of image brightness. Wavelets basis have been computed on the image domain for subspaces of images, motion fields and divergence-free motion fields. Image assimilation with this reduced model allows to estimate smooth velocity fields with properties defined by user. This also solves the issue of complex geographical domains and the difficulty of applying boundary conditions on these domains. First results are obtained with a reduced dimension of motion to a few scalars, to be compared with the original problem that has the size of image domain.

## 6.5. Ensemble forecasting with sequential aggregation

The aggregation of an ensemble of forecasts is an approach where the members of an ensemble are given a weight before every forecast time, and where the corresponding weighted linear combination of the forecasts provides an improved forecast. A robust aggregation can be carried out so as to guarantee that the aggregated forecast performs better, in the long run, than any linear combination of the ensemble members with time-independent weights. The approaches are then based on machine learning. The aggregation algorithms can be applied to forecast analyses (generated from a data assimilation system), so that the aggregated forecasts are naturally multivariate fields.

### 6.5.1. Application of sequential aggregation to meteorology and air quality

**Participants:** Anne Tilloy, Vivien Mallet, Fabien Brocheton [Numtech], David Poulet [Numtech].

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. Each time new observations are available, new weights for the linear combination are computed and applied for the next forecast. We applied the discounted ridge regression algorithm, which we previously introduced for sequential aggregation of air quality forecasts, to forecast wind and temperature at given observation stations. The ensemble was generated with forecasts at different range from two models. The aggregation proved to be efficient for one-day forecasts at least.

The discounted ridge regression was also applied to the simulations of the Air Quality Modeling Evaluation International Initiative (AQMEII) over Europe and North America, for different pollutants (gases and particulate matter).

### *6.5.2. Sequential aggregation with uncertainty estimation*

**Participants:** Vivien Mallet, Sergiy Zhuk [IBM research], Paul Baudin, Gilles Stoltz [CLASSIC], Karine Sartelet [CEREA].

A new issue is the estimation of the uncertainties associated with the aggregated forecasts. One investigated direction relies on the framework of machine learning, with the aggregation of an ensemble of probability density functions instead of the point forecasts of the ensemble.

Another direction, which led to finalized results in 2012, is to reformulate the aggregation problem in a filtering problem for the weights. The weights are supposed to satisfy some dynamics with unknown model error, which defines the state equation of a filter. An observation equation compares the aggregated forecast with the observations (or analyses) with known observational error variance. The filter finally computes estimates for the weights and quantifies their uncertainties. We applied a Kalman filter and a minimax filter for air quality forecasting.

## 6.6. 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, these 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 problem is to control the location of structures using the observations,
- from the image's viewpoint, a model of the dynamics and structures has to be built from the observations.

### *6.6.1. Divergence-free motion estimation*

**Participants:** Dominique Béréziat [UPMC], Isabelle Herlin, Sergiy Zhuk [IBM Research, Ireland].

This research addresses the issue of divergence-free motion estimation on an image sequence, acquired over a given temporal window. Unlike most state-of-the-art technics, which constrain the divergence to be small thanks to Tikhonov regularization terms, a method that imposes a null value of divergence of the estimated motion is defined.

Motion is characterized by its vorticity value and assumed to satisfy the Lagragian constancy hypothesis. An image model is then defined: the state vector includes the vorticity, whose evolution equation is derived from that of motion, and a pseudo-image that is transported by motion. An image assimilation method, based on the 4D-Var technics, is defined and developed that estimates motion as a compromise between the evolution equations of vorticity and pseudo-image and the observed sequence of images.

The method is applied on Sea Surface Temperature (SST) images acquired over Black Sea by NOAA-AVHRR sensors. The divergence-free assumption is roughly valid on these acquisitions, due to the small values of vertical velocity at the surface. Fig. 2 displays data and results. As no ground truth of motion is available, the method is quantified by the value of correlation between the pseudo-images and real acquisitions [28].

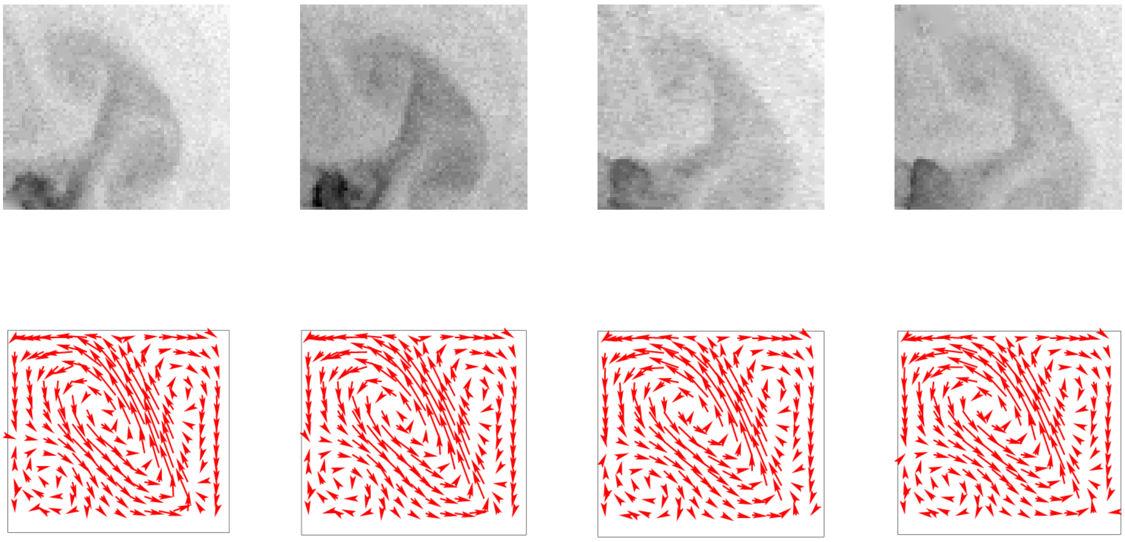


*Figure 2. SST image observations and motion results.*

### 6.6.2. Improvement of motion estimation by assessing errors on the dynamics

**Participants:** Dominique Béréziat [UPMC], Isabelle Herlin.

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, we consider adding a model error term in the evolution equation of motion and design a weak formulation of 4D-Var data assimilation. The cost function to be minimized simultaneously depends on the initial motion field, at the begining 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 [27].

This error assessment method is evaluated and quantified on twin experiments, as no ground truth would be available for real image data. Fig. 3 shows four frames of a series of observations obtained by integrating the evolution model from an initial condition on image and velocity field (the ground truth $\mathbf{w}_{\text{ref}}(0)$ displayed on the left of Fig. 4 ). An error value is added at each time step on the motion value, when integrating the simulation model. This error is a constant bias.

We performed two data assimilation experiments. The first one considers the evolution model as perfect, with no error in the evolution equation. It is denoted PM (for Perfect Model). The second one, denoted IM (for Imperfect Model) involves an error in the motion evolution equation. In Fig.4 are displayed the motion fields retrieved by PM and IM at the beginning of the temporal window.

As it can be seen, IM computes a correct velocity field while PM completely fails.

### 6.6.3. Nonlinear Observation Equation For Motion Estimation

**Participants:** Dominique Béréziat [UPMC], Isabelle Herlin.

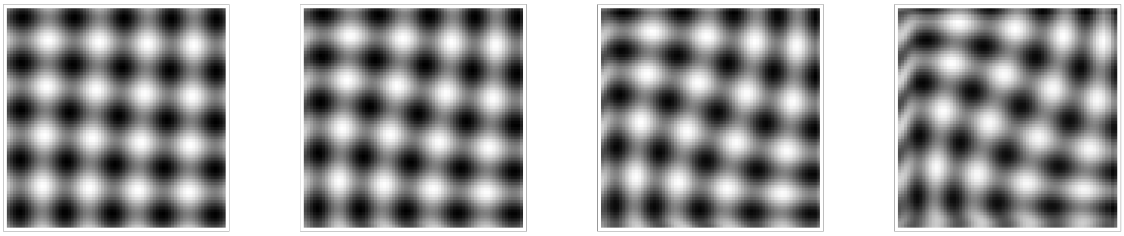


Figure 3. Observations Images.

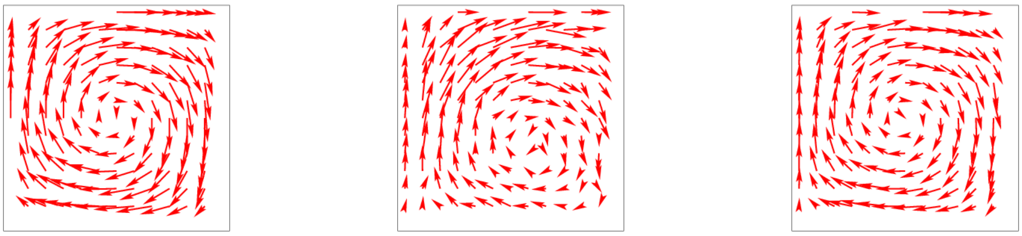


Figure 4. Left: ground-truth, middle: PM, right: IM.

In the image processing literature, the optical flow equation is usually chosen to assess motion from an image sequence. However, it corresponds to an approximation that is no more valid in case of large displacements. We evaluated improvements obtained when using the non linear transport equation of the image brightness by the velocity field [25]. A 4D-Var data assimilation method is designed that simultaneously solves the evolution equation and the observation equation, in its non linear and linearized form. The comparison of results obtained with both observation equations is quantified on synthetic data and discussed on oceanographic Sea Surface Temperature (SST) images. We show that the non linear model outperforms the linear one, which underestimates the motion norm. Fig.5 illustrates this on SST images (motion vectors are displayed by arrows).

The aim of this research is to achieve a correct estimation of motion when the object displacement is greater than its size. Howerver, in this case, coarse-to-fine incremental methods as well as the non linear data assimilation method fail to retrieve a correct value. The perspective is then to include, in the state vector, a variable describing the trajectory of pixels. The observation operator will then measure the effective displacement of pixels, according to their trajectories, and allow a better estimation of motion value.

### 6.6.4. Sliding windows method for motion estimation on long temporal image sequences

**Participants:** Karim Drifi, Isabelle Herlin.

This study concerns the estimation of motion fields from satellite images on long temporal sequences. The huge computational cost and memory required by data assimilation methods on the pixel grid makes impossible to use these techniques on long temporal intervals. For a given dynamic model (named full model), on the pixel grid, the Galerkin projection on subspaces provides a reduced model, that allows image assimilation at low cost. The definition of this reduced model however requires defining an optimal subspace of motion. A **sliding windows** method is thus designed:

- The long image sequence is split into small temporal windows that half overlap in time.

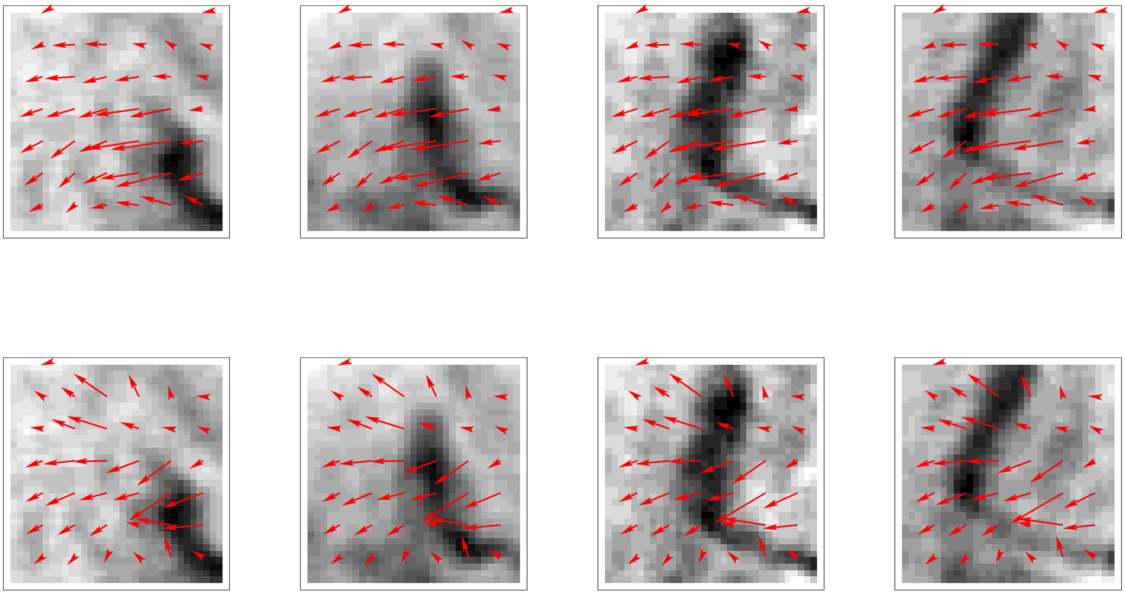


*Figure 5. Top: Non-linear observation equation. Bottom: linear.*

- Data assimilation in the full model is applied on the first window to retrieve the motion field.
- The estimate of motion field at the beginning of the second window makes it possible to define the subspace for motion and a reduced model is obtained by Galerkin projection.
- Data assimilation in the reduced model is applied for this second window.
- The process is then iterated for the next window until the end of the whole image sequence.

Experiments were designed to quantify the results of this sliding windows method with base obtained by Principal Orthogonal Decomposition or computed as bi-sine functions [29].

### *6.6.5. Tracking of structures from an image sequence*

**Participants:** Yann Lepoittevin, Isabelle Herlin, 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, and those of distance map modeling the tracked structures. The method is applied on meteorological satellite data, in order to track tropical clouds on image sequences and estimate their motion.

Part of research was concerned on the numerical schemes applied for advecting the distance map and designing its adjoint.

## 6.7. Minimax filtering

In minimax filtering for state estimation, the initial state error, the model error and the observational errors are classically supposed to belong to one joint ellipsoid. In this case, it is only assumed that the errors, stochastic or deterministic, are bounded. For each assimilation experiment, the filter computes an ellipsoid where one will find at least all states compatible with observations and errors description. The state estimate is taken as the center of the ellipsoid. No assumption on the actual distribution of the errors in needed and the state estimate minimizes the worst-case error, which makes the filter robust.

### 6.7.1. A posteriori minimax motion estimation

**Participants:** Sergiy Zhuk [IBM Research, Ireland], Isabelle Herlin, Olexander Nakonechnyi [Taras Shevchenko National University of Kyiv], Jason Frank [CWI, the Netherlands].

Data assimilation algorithms based on the 4D-Var formulation look for the so-called conditional mode estimate. The latter maximizes the conditional probability density function, provided the initial condition, model error and observation noise are realizations of independent Gaussian random variables. However this Gaussian assumption is often not satisfied for geophysical flows. Moreover, the estimation error of the conditional mode estimate is not a first-hand result of these methods. The issues above can be addressed by means of the Minimax State Estimation (MSE) approach. It allows to filter out any random (with bounded correlation operator) or deterministic (with bounded energy) noise and assess the worst-case estimation error.

The iterative MSE algorithm was developed for the problem of optical flow estimation from a sequence of 2D images. The main idea of the algorithm is to use the "bi-linear" structure of the Navier-Stokes equations and optical flow constraint in order to iteratively estimate the velocity. The algorithm consists of the following parts:

1) we construct pseudo-observations $\widehat{I}$ as the estimate of the image brightness function $I(x,y,t)$ solving the optical flow constraint such that $\widehat{I}$ fits (in the sense of least-squares) the observed sequence of images. To do so, we set the velocity field in the optical flow constraint to be the current minimax estimate of the velocity field $\mathbf{w}$, obtained at the previous iteration of the algorithm, and construct the minimax estimate $\widehat{I}$ of the solution of the resulting linear advection equation using the observed image sequence as discrete measurements of the brightness function;

2) we plug the estimate of the image gradient, obtained out of pseudo-observations $\widehat{I}$ in 1), into the optical flow constraint and the current minimax estimate $\mathbf{w}$ of the velocity field into the non linear part of Navier-Stokes equations so that we end up with a system of linear PDEs, which represents an extended state equation: it contains a linear parabolic equation for the velocity field and linear advection equation for the image brightness function; we construct the minimax estimate of the velocity field from the extended state equation using again the observed image sequence as discrete measurements of the brightness function;

3) we use the minimax estimate of the velocity field obtained in 2) in order to start 1) again.

Point 1) has been implemented and tested. As Point 2) is currently under development, it is replaced by one of our motion estimation method in order to be plugged in Point 3).

## 6.8. Fire application

### 6.8.1. Model evaluation for fire propagation

**Participants:** Vivien Mallet, Jean-Baptiste Fillipi [CNRS], Bahaa Nader [University of Corsica].

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 models performance, but currently, simulation errors are not sufficiently qualified and quantified. As the main objective is to realize a *decision support system*, it is required to establish robust forecast evaluations. In the context of the ANR project IDEA, the evaluation of model simulations has led to the definition and implementation of a series of forecast scores. The merits and shortcomings of the scores were evaluated on synthetic cases. This demonstrated the efficiency of scores that take into account the fire dynamics, where some classical scores may fail. This was also found on real fires, using field observations.

In addition, we consider that the proper evaluation of a model requires to apply it to a large number of fires – instead of carrying out a fine tuning on just one fire. We implemented a software to simulate a large number of fires (from the Prométhée database, http://www.promethee.com/) with the simulation model ForeFire (CNRS/University of Corsica) and evaluate the results with error measures. One simulation requires mainly the following data: the ignition point, the ground elevation, the vegetation cover and the wind field. See illustration on Fig. 6 .

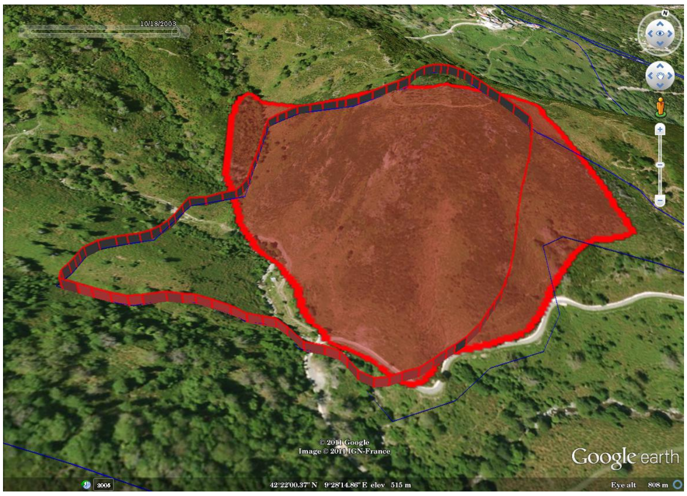


*Figure 6. Fire simulation (using ForeFire) in red elevated contour, and observation (from Prométhée) of the burned area in filled red contour, for a 2003 fire near San-Giovanni-di-Moriani (Corsica).*

# FLUMINANCE Project-Team

# 6. New Results

## 6.1. Fluid motion estimation

### *6.1.1. Stochastic uncertainty models for motion estimation*

**Participants:** Thomas Corpetti, Etienne Mémin.

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 define a natural linear scale space multiresolution 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 has been shown to provide 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 provide a useful piece of information in the context of data assimilation. This ability has been exploited to define multiscale data assimilation filtering schemes. These works have been recently published in *IEEE trans. on Image Processing* and in Numerical Mathematics: Theory, Methods and Applications [16], [18]. We intend to pursue this formalization to define dense motion estimators that allow us handling, in the same way, luminance conservation under motion uncertainty principles. An efficient GP-GPU implementation of the local estimator is also targeted.

### *6.1.2. 3D flows reconstruction from image data*

**Participants:** Ioana Barbu, Cédric Herzet, Etienne Mémin.

Our work focuses on the design of new tools for the problem of 3D reconstruction of a turbulent flow motion. 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. On the one hand, state-of-the-art methodologies such as "sparse representations" will be investigated for the characterization of the observation and fluid motion models. Sparse representations are well-suited to the representation of signals with very few coefficients and offer therefore advantages in terms of computational and storage complexity. On the other hand, the estimation problem will be placed into a probabilistic Bayesian framework. This will allow the use of state-of-the-art inference tools to effectively exploit the strong time-dependence of the fluid motion. In particular, we will investigate the use of "ensemble Kalman" filter to devise low-complexity sequential estimation algorithms.

At the beginning of Ioana Barbu's PhD, we concentrated our efforts on the problem of reconstructing the particle positions from several two-dimensional images. Our approach is based on the exploitation of a particular family of sparse representation algorithms, namely the so-called "pursuit algorithms". Indeed, the pursuit procedures generally allow a good trade-off between performance and complexity. Hence, we have performed a thorough study comparing the reconstruction performance and the complexity of different state-of-the-art algorithms to that achieved with pursuit algorithms. This work has led to the publication of two conference papers in experimental fluid mechanics.

This year, our work has focused on: i) the estimation of the 3D velocity field of the fluid flow from the reconstructed volumes of particles; ii) the design of new methodologies allowing to jointly estimate the volume of particles and the velocity field from the received image data. More particularly, we have implemented a motion estimator generalizing the local Lucas-Kanade's procedure to a 3D problem. A potential strength of the proposed approach is the possibility to consider a fully parallel (and therefore very fast) implementation. On the other hand, we have started investigating the problem of jointly estimating the volumes of particles and the velocity field. Our approach is based on the combination of sparse representation algorithms and "Lucas-Kanade"-like motion estimation methods. We are about testing the proposed approach on experimental data in order to assess its performance in practical scenarios of fluid mechanics. We also intend to collaborate with the group of Fulvio Scarano at TU Delft to assess and compare our method on experimental 3D data.

### 6.1.3. Motion estimation techniques for turbulent fluid flows

**Participants:** Patrick Héas, Dominique Heitz, Cédric Herzet, Etienne Mémin.

Based on physical laws describing the multi-scale structure of turbulent flows, this study concerns the proposition of smoothing functional for the estimation of homogeneous turbulent flow velocity fields from an image sequence. This smoothing is achieved by imposing some scale invariance property between histograms of motion increments computed at different scales. By reformulating this problem from a Bayesian perspective, an algorithm is proposed to jointly estimate motion, regularization hyper-parameters, and to select the most likely physical prior among a set of models. Hyper-parameter and model inference is conducted by likelihood maximization, obtained by marginalizing out non-Gaussian motion variables. The Bayesian estimator is assessed on several image sequences depicting synthetic and real turbulent fluid flows. Results obtained with the proposed approach in the context of fully developped turbulence improve significantly the results of state of the art fluid flow dedicated motion estimators. This series of works, which have been done in close collaboration with P. Minnini (University of Buenos Aires), have been published in several journals [21], [22], [23].

### 6.1.4. Wavelet basis for multi-scale motion estimation

**Participants:** Pierre Dérian, Patrick Héas, Cédric Herzet, Souleymane Kadri Harouna, Etienne Mémin.

This work describes the implementation of a simple wavelet-based optical-flow motion estimator dedicated to the recovery of fluid motion. The wavelet representation of the unknown velocity

field is considered. This scale-space representation, associated to a simple gradient-based optimization algorithm, sets up a natural multiscale/multigrid optimization framework for the optical flow estimation that can be combined to more traditional incremental multiresolution approaches. Moreover, a very simple closure mechanism, approximating locally the solution by high-order polynomials, is provided by truncating the wavelet basis at intermediate scales. This offers a very interesting alternative to traditional Particle Image Velocimetry techniques. As another alternative to this medium-scale estimator, we explored strategies to define estimation at finer scales. These strategies rely on the encoding of high-order smoothing functional on appropriate wavelet basis. Divergence-free bi-othogonal wavelet bases enable to further nicely enforce volume preserving motion field. Numerical results on several examples have demonstrated the relevance of the method for divergence free-2D flows. These studies have been published in the journal of Numerical Mathematics: Theory, Methods and Applications [19] and in the journal of Computer Vision [24]. The extension to 3D flows would be an interesting perspective.

### 6.1.5. Wavelet-based divergence-free fBm prior: application to turbulent flow estimation

**Participant:** Patrick Héas.

This work is concerned with the estimation of turbulent flows from the observation of an image sequence. From a Bayesian perspective, we propose to study divergence-free isotropic fractional Brownian motion (fBm) as a prior model for instantaneous turbulent velocity fields. These priors are self-similar stochastic processes, which characterize accurately second-order statistics of velocity fields in incompressible isotropic turbulence. Although, these models belong to a well-identified family of rotation invariant regularizers, there is a lack of effective algorithms in the literature to deal in practice with their fractional nature. To respond to this problem, we propose to decompose fBms on well-chosen wavelet bases. As a first alternative, we propose to design wavelets as whitening filters for divergence-free isotropic fBms, which are correlated both in space and scale. The second alternative is to use a divergence-free wavelet basis, which will take implicitly into account the divergence-free constraint arising form the physics.

### 6.1.6. Sparse-representation algorithms

**Participant:** Cédric Herzet.

The paradigm of sparse representations is a rather new concept which appears to be central in many field 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 turns out to be central in the recovering of volumes of particles in the 3D Tomo-PIV problem. In these contexts, the dimensionality of the problem can be very large and the use of sparse-representation algorithms allowing for a good trade-off between complexity and effectiveness is needed.

This year, we have therefore pursued our study of efficient sparse decomposition algorithms. In particular, we have extended our work addressing the problem of finding good sparse representations into a probabilistic framework. First, we have proposed a new family of pursuit algorithms able to take into account any type of dependence (e.g. spatial or temporal) between the atoms of the sparse decomposition. This work has led to the publication of a paper in the proceedings of the international conference IEEE ICASSP 2012.

Exploiting further this probabilistic framework, we have then considered the design of structured soft pursuit algorithms. In particular, instead of making hard decisions on the support of the sparse representation and the amplitude of the non-zero coefficients, our soft procedures iteratively update probability on the latter values. The proposed algorithms are designed within the framework of the mean-field approximations and resort to the so-called variational Bayes EM algorithm to implement an efficient minimization of a Kullback-Leibler criterion. On the other hand, the proposed methodologies can handle "structured" sparse representations, that is, sparse decompositions where some dependence exists between the non-zero elements of the support. The prior model on the support of the sparse decomposition is based on a Boltzmann machine which encompasses as particular cases many type of dependence (Markov chain, Ising model, tree-like structure, etc). This work has been published in the journal IEEE Trans. on Signal Processing in 2012.

## 6.2. Tracking and data assimilation

### 6.2.1. Stochastic filtering for fluid motion tracking

**Participants:** Sébastien Béyou, Anne Cuzol, Etienne Mémin.

We investigated the study of a recursive Bayesian filter for tracking velocity fields of fluid flows. The filter combines an Ito diffusion process associated to 2D vorticity-velocity formulation of Navier-Stokes equation and discrete image error reconstruction measurements. In contrast to usual filters designed for visual tracking problems, our filter combines a continuous law for the description of the vorticity evolution with discrete image measurements. We resort to a Monte-Carlo approximation based on particle filtering. The designed tracker provides a robust and consistent estimation of instantaneous motion fields along the whole image sequence.

When the likelihood of the measurements can be modeled as a Gaussian law, we have also investigated the use of the so-called ensemble Kalman filtering for fluid tracking problems. This kind of filters introduced for the analysis of geophysical fluids is based on the Kalman filter update equation. Nevertheless, unlike traditional Kalman filtering setting, the covariances of the estimation errors, required to compute the so-called Kalman gain, relies on an ensemble of forecasts. Such a process gives rise to a Monte-Carlo approximation for a family of non-linear stochastic filters enabling to handle state spaces of large dimension. We have recently proposed an extension of this technique that combines sequential importance sampling and the propagation law of an ensemble Kalman filter. This technique leads to an ensemble Kalman filter with an improved efficiency. We have in particular investigated the introduction of a nonlinear direct image measurement operator within this ensemble Kalman scheme. This modification of the filter provides very good results on 2D numerical and experimental flows even in the presence of strong noises. We are currently assessing its application to oceanic satellite images for the recovering of ocean streams. We are also studying the impact on the stochastic dynamics of turbulent noise defined as auto-similar Gaussian random fields and the introduction within an incremental ensemble analysis scheme of multiscale motion measurements. This work has been recently accepted for publication in the Tellus A journal [17].

### 6.2.2. Reduced-order models for flows representation from image data

**Participants:** Cédric Herzet, Etienne Mémin, Véronique Souchaud.

One of the possibilities to neglect the influence of some degrees of freedom over the main characteristics of a flow consists in representing it as a sum of $K$ orthonormal spatial basis functions weighted with temporal coefficients. To determine the basis function of this expansion, one of the usual approaches relies on the Karhunen-Loeve decomposition (refered to as proper orthogonal decomposition – POD – in the fluid mechanics domain). In practice, the spatial basis functions, also called modes, are the eigenvectors of an empirical auto-correlation matrix which is built from “snapshots" of the considered physical process.

In this axis of work we focus on the case where one does not have a direct access to snapshots of the considered physical process. Instead, the POD has to be built from the partial and noisy observation of the physical phenomenon of interest. Instances of such scenarios include situations where real instantaneous vector-field snapshots are estimated from a sequence of *images*. We have been working on several approaches dealing with such a new paradigm. A first approach consists in extending standard penalized motion-estimation algorithms to the case where the sought velocity field is constrained to span a low-dimensional subspace. In particular, we have considered scenarios where the standard optical flow constraint (OFC) is no longer statisfied and one has therefore to resort to a Discrete Finite Difference (DFD) model. The non-linearity of the latter leads to several practical issues that we have addressed this year. We are currently assessing the performance of the proposed method on experimental data in order to validate its relevance in practical scenarios. In a second approach we have studied two variational data assimilation techniques for the estimation of low order dynamical models for fluid flows. Both methods are built from optimal control recipes and rely on POD representation associated to Galerkin projection of the Navier Stokes equations. The proposed techniques differ in the control variables they involve. The first one introduces a weak dynamical model defined only up to an additional uncertainty time dependent function whereas the second one, handles a strong dynamical constraint in which the coefficients of the dynamical system constitute the control variables. Both choices correspond to different approximations of the relation between the reduced basis on which is expressed the motion feld and the basis components that have been neglected in the reduced order model construction. The techniques have been assessed on numerical data and for real experimental conditions with noisy Image Velocimetry data. This work has been published in the Journal of Computational Physics [15]. In collaboration with the University of Buenos Aires, we have also explored, a method that combines Proper Orthogonal Decomposition with a spectral technique to analyze and extract reduced order models of flows from time resolved data of velocity fields. This methodology, relying on the eigenfunctions of the Koopman operator, is specifically adapted to flows with quasi periodic orbits in the phase space. The technique is particularly suited to cases requiring a discretization with a high spatial and temporal resolution. The proposed analysis enables to decompose the flow dynamics into modes that oscillate at a single frequency. For each modes an energy content and a spatial structure can be put in correspondence. This approach has been assessed for a wake flow behind a cylinder at Reynolds number 3900 and has been recently accepted under minor revisions condition to the journal of Theoretical and Computational Fluid Dynamics.

### 6.2.3. Optimal control techniques for the coupling of large scale dynamical systems and image data

**Participants:** Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

This work aims at investigating 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 conditionally accepted for publication in Journal of Computational Physics. Along the same line of studies we have started to investigate the 3D case. 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 is 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).

### *6.2.4. Free surface flows reconstruction and tracking*

**Participants:** Benoît Combes, Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

Characterizing a free-surface flow (space and time-dependent velocity and geometry) given observations/measures at successive times is an ubiquitous problem in fluid mechanic and in hydrology. Observations can consist of e.g. measurements of velocity, or like in this work of measurements of the geometry of the free-surface. Indeed, recently developed depth/range sensors allow to capture directly a rough 3D geometry of surfaces with high space and time resolution. We have investigated the performance of the Kinect and have shown that it is likely to capture temporal sequences of depth observations of wave-like surfaces with wavelengths and amplitudes sufficiently small to characterize medium/large scale flows. Several data assimilation methods have been experimented and compared to estimate both time dependent geometry and displacement field associated to a free-surface flow from a temporal sequence of Kinect data. This study have been conducted on synthetic and real-world data. It has been presented to a data assimilation conference [35]. Finally, we explored the application of such techniques to hydrological applications. These results are currently considered for submission to Journal of Hydrology.

### *6.2.5. Stochastic filtering technique for the tracking of closed curves*

**Participants:** Christophe Avenel, Etienne Mémin.

We have proposed a filtering methodology for the visual tracking of closed curves. Opposite to works of the literature related to this issue, we consider here a curve dynamical model based on a continuous time evolution law with different noise models. This led us to define three different stochastic differential equations that capture the uncertainty relative to curve motions. This new approach provides a natural understanding of classical level-set dynamics in terms of such uncertainties. These evolution laws have been combined with various color and motion measurements to define probabilistic state-space models whose associated Bayesian filters can be handled with particle filters. This ongoing work will be continued within extensive curve tracking experiments and extended to the tracking of other very high dimensional entities such as vector fields and surfaces. This work, which corresponds to the PhD thesis of Christoph Avenel has been presented in several conferences and has been submitted to two different journals. Let us note that it has also led to a fruitful collaboration with MeteoFrance [30]

### *6.2.6. 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 (LORIA) in the context of vorticity tracking from image data.

### *6.2.7. Stochastic fluid flow dynamics under uncertainty*

**Participant:** Etienne Mémin.

In this research axis we aim at devising stochastic Eulerian expressions for the description of fluid flow evolution laws incorporating uncertainty on the particles location. Such an uncertainty modeled through the introduction of a random term, allows taking into account 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 mean 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.

### *6.2.8. Variational assimilation of images for large scale fluid flow dynamics with uncertainty*

**Participants:** Souleymane Kadri Harouna, Etienne Mémin.

In this work we explore the assimilation of a large scale representation of the flow dynamics with image data provided at a finer resolution. The velocity field at large scales is described as a regular smooth components 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 started to assess the performance of a variational assimilation technique with direct image data observation. Preliminary encouraging results obtained for a wavelet-based 2D Navier Stokes implementation and images of a passive scalar transported by the flow have been obtained. Large-scale simulation under uncertainty for the 3D viscous Taylor-Green vortex flow have been carried out and show promising results of the approach.

## 6.3. Analysis and modeling of turbulent flows

### *6.3.1. 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.

### *6.3.2. Numerical and experimental image and flow database*

**Participant:** Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The main challenge was to enlarge a database mainly based on two-dimensional flows, with three-dimensional turbulent flows. New synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with real image sequences based on wake and mixing layers flows. This new database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images.

## 6.4. Visual servoing approach for fluid flow control

### *6.4.1. Fully exploitation of the controlled degrees of freedom of the 2D plane Poiseuille flow*

**Participants:** Christophe Collewet, Xuan Quy Dao.

This works concerns the Phd of Xuan-Quy Dao and can be seen as an extension of the works carried out by Romeo Tatsambon during its post-doc position. Indeed, during this post-doc we proved that our vision-based approach overcomes the traditional approaches. Nevertheless, to compare our method with the literature, we used a traditional control law, the LQR control law. However, we can fully exploit the capabilities of visual servoing techniques by designing a more efficient control law than the LQR one. This has been done this year. We have validated our approach to the problem of minimizing the drag of the 2D plane Poiseuille flow. An important issue was also to ensure that, during the process of drag reduction, the kinetic energy density will not grow. This is of great importance since it is well known that the controlled flow may become turbulent when this kinetic energy density is growing. To cope with this problem we have proposed to design a control law based on partitioned visual servo. Indeed, following this way, we are able to simultaneously minimize the drag AND the kinetic energy density in contrast to the existing approaches. This work has been accepted to the "American control conference (ACC'12)", to the "Conférence internationale francophone d'automatique (CIFA'12)" and to the "6th AIAA Flow Control Conference". We have also explored an approach based on eigenstructure placement to ensure a strict decrease of the kinetic energy density. Another approach has also been explored, it tends to decouple all the controlled degrees of freedom of the system so that an exponential decoupled decrease of each component of the state vector is obtained. The great advantage of this approach is that all quantities depending on the state vector (like the drag) are also exponential decreasing functions.

### *6.4.2. Control behind a backward-facing step*

**Participant:** Christophe Collewet.

Instead of setting up an experimental closed loop control problem for the plane Poiseuille flow with temporal perturbations, which is theoretically based on an unrealistic infinite channel, we explore in this axis of work a closed-loop control of a flow behind a backward step. The control is expressed through the visual servoing formalism and fast velocity measurements in the recirculation zone. This work is performed in the context of the PhD thesis of Nicolas Gautier from PMMH-ESCPI. This thesis is co-supervised with Jean-Luc Aider (CNRS/PMMH-ESPCI).

# MAGIQUE-3D Project-Team

# 6. New Results

## 6.1. Inverse Problems

### 6.1.1. Reconstruction of an elastic scatterer immersed in a homogeneous fluid

**Participants:** Hélène Barucq, Rabia Djellouli, Élodie Estecahandy.

The determination of the shape of an obstacle from its effects on known acoustic or electromagnetic waves is an important problem in many technologies such as sonar, radar, geophysical exploration, medical imaging and nondestructive testing. This inverse obstacle problem (IOP) is difficult to solve, especially from a numerical viewpoint, because it is ill-posed and nonlinear. Its investigation requires as a prerequisite the fundamental understanding of the theory for the associated direct scattering problem, and the mastery of the corresponding numerical solution methods.

In this work, we are interested in retrieving the shape of an elastic obstacle from the knowledge of some scattered far-field patterns, and assuming certain characteristics of the surface of the obstacle. The corresponding direct elasto-acoustic scattering problem consists in the scattering of time-harmonic acoustic waves by an elastic obstacle $\Omega^s$ embedded in a homogeneous medium $\Omega^f$, that can be formulated as follows:

$$\begin{aligned} \Delta p + (\omega^2/c_f^2)\,p &= 0 && \text{in } \Omega^f \\ \nabla\cdot\sigma(u) + \omega^2\rho_s\,u &= 0 && \text{in } \Omega^s \\ \omega^2\rho_f u\cdot n &= \partial p/\partial n + \partial e^{i\,(\omega/c_f)\,x\cdot d}/\partial n && \text{on } \Gamma \\ \sigma(u)n &= -pn - e^{i\,(\omega/c_f)\,x\cdot d}n && \text{on } \Gamma \\ \lim_{r\to+\infty} r\,(\partial p/\partial r - i\,(\omega/c_f)\,p) &= 0 && \end{aligned} \tag{2}$$

where $p$ is the fluid pressure in $\Omega^f$ whereas $u$ is the displacement field in $\Omega^s$, and $\sigma(u)$ represents the stress tensor of the elastic material.

This boundary value problem has been investigated mathematically and results pertaining to the existence, uniqueness and regularity can be found in [86] and the references therein, among others. We propose a solution methodology based on a regularized Newton-type method for solving the IOP. The proposed method is an extension of the regularized Newton algorithm developed for solving the case where only Helmholtz equation is involved, that is the acoustic case by impenetrable scatterers [79]. The direct elasto-acoustic scattering problem defines an operator $F : \Gamma \to p_\infty$ which maps the boundary $\Gamma$ of the scatterer $\Omega^s$ onto the far-field pattern $p_\infty$. Hence, given one or several measured far-field patterns $\widetilde{p}_\infty(\widehat{x})$, corresponding to one or several given directions $d$ and wavenumbers $k$, one can formulate IOPs as follows:

$$\text{Find a shape } \Gamma \text{ such that } \quad F(\Gamma)(\widehat{x}) = \widetilde{p}_\infty(\widehat{x}); \quad \widehat{x}\in S^1.$$

We propose a solution methodology based on a regularized Newton-type method to solve this inverse obstacle problem. At each Newton iteration, we solve the forward problem using a finite element solver based on discontinuous Galerkin approximations, and equipped with high-order absorbing boundary conditions. We have first characterized the Fréchet derivatives of the scattered field. They are solution to the same boundary value problem as the direct problem with other transmission conditions. This work has been presented both in FACM11 and in WAVES 2011. A paper has been submitted.

### 6.1.2. $hp$-adaptive inversion of magnetotelluric measurements

**Participants:** Hélène Barucq, Julen Alvarez Aramberri, David Pardo.

The magnetotelluric (MT) method is a passive electromagnetic (EM) exploration technique that allows to determine the resistivity distribution in the subsurface of the area of interest on scales varying from few meters to hundreds of kilometers. Commercial uses include hydrocarbon (oil and gas) exploration, geothermal exploration, and mining exploration, as well as hydrocarbon and groundwater monitoring. MT measurements are governed by the electromagnetic phenomena, which can be described by Maxwell's equations. We solve those equations by a goal-oriented hp-adaptivity Finite Element Method (FEM).

In order to estimate the resistivity distribution in the Earth's subsurface, we solve an Inverse Problem. We define a Misfit Function that represents the difference between the measured and computed data for a particular resistivity distribution. By minimizing this misfit function using a gradient based approach with model reduction techniques, and hence solving the inverse problem, we are able to determine the properties of the subsurface materials.

## 6.2. Modeling

### 6.2.1. Implementation of a non-reflecting boundary condition on ellipsoidal boundary

**Participants:** Hélène Barucq, Anne-Gaëlle Saint-Guirons, Sébastien Tordeux.

The modeling of wave propagation problems using finite element methods usually requires the truncation of the computational domain around the scatterer of interest. Absorbing boundary condition are classically considered in order to avoid spurious reflections. This year we have implemented and tested an exact condition based on a non local Dirichlet to Neumann operator in the context of the Helmholtz equation posed on an elongated domain.

### 6.2.2. Explicit computation of the electrostatic energy for an elliptical charged disc

**Participants:** Sophie Laurens, Sébastien Tordeux.

In [32], We have described a method to obtain an explicit expression for the electro- static energy of a charged elliptical infinitely thin disc. The charge distribution is assumed to be polynomial. Such explicit values for this energy are fundamen- tal for assessing the accuracy of boundary element method codes. The main tools used are an extension of Copson's method and a diagonalization, given by Leppington and Levine, of the single-layer potential operator associated with the electrostatic potential created by a distribution of charges on an elliptical disc.

### 6.2.3. A new modified equation approach for solving the wave equation

**Participants:** Cyril Agut, Hélène Barucq, Henri Calandra, Julien Diaz, Florent Ventimiglia.

The new method involving $p$-harmonic operator described in section 3.2 has been presented in [17]. We have proved the convergence of the scheme and its stability under a CFL condition. Numerical results in one, two and three-dimensional configurations show that this CFL condition is slightly greater than the CFL condition of the second-order Leap-Frog scheme.

In the framework of the PhD thesis of Florent Ventimiglia, we are now considering the extension of this technique to the first order formulation of the acoustic and elastodynamic equations. A numerical analysis of performance in 1D indicates that, for a given accuracy, this method requires less storage than the High-Order ADER Schemes for and similar computational costs. We are now implementing this algorithm in 3D in order to confirm this analysis and to assess its performance in an RTM framework on realistic configurations.

### 6.2.4. Stability Analysis of an Interior Penalty Discontinuous Galerkin Method for the Wave equation

**Participants:** Cyril Agut, Hélène Barucq, Julien Diaz.

The Interior Penalty Discontinuous Galerkin Method [72], [69], [83] we use in the IPDGFEM code requires the introduction of a penalty parameter. Except for regular quadrilateral or cubic meshes, the optimal value of this parameter is not explicitely known. Moreover, the condition number of the resulting stiffness matrix is an increasing function of this parameter, but the precise behaviour has not been explicited neither. We have carried out a theoretical and numerical study of the pnealization parameter and of the CFL condition for quadrilateral and cubic meshes, this results have been presented in a paper accepted in M2AN [16]

### *6.2.5. Higher Order Absorbing Boundary Conditions for the Wave Equation*

**Participants:** Hélène Barucq, Juliette Chabassier, Julien Diaz.

The numerical simulation of wave propagation is generally performed by truncating the propagation medium and the team works on new ABCs, trying to improve the performance of existing conditions. Following the analysis performed in [23], we have considered the issue of constructing high-order ABCs for the Helmholtz equation. Now, to derive conditions of order greater than two is really technical. In addition, when the coefficients representing the geological properties of the medium are not regular, the method of construction of ABCs is not completely justified. That is why we turned to the construction of conditions that take into account all the characteristics of the diffraction phenomenon and not only waves that propagate like in the case of standrad ABCs. This is what we call enriched ABCs. A research report is being written, an article should be submitted in 2013. During 2012, a publication for the acoustic wave equation has been accepted in M3AS [23] and a second one has been submitted.

### *6.2.6. Multiperforated plates in linear acoustics*

**Participants:** Abderrahmane Bendali, M'Barek Fares, Sophie Laurens, Estelle Piot, Sébastien Tordeux.

Acoustic engineers use approximate heuristic models to deal with multiperforated plates in liners and in combustion chambers of turbo-engines. These models were suffering from a lack of mathematical justifications and were consequently difficult to improve. Performing an asymptotic analysis (the small parameter is the radius of the perforations), we have justified these models and proposed some improvement. Our theoretical results have been compared to numerical simulations performed at CERFACS (M'Barek Fares) and to acoustical experiments realized at ONERA (Estelle Piot). Two papers have been published in 2012 [27], [30].

### *6.2.7. Performance Assessment of IPDG for the solution of an elasto-acoustic scattering problem*

**Participants:** Hélène Barucq, Rabia Djellouli, Élodie Estecahandy.

We present a solution methodology for the direct elasto-acoustic scattering problem that falls in the category of Discontinuous Galerkin methods. The method distinguishes itself from the existing methods by combining high-order Discontinuous Galerkin approximations, local stabilizations for the coupled problem and the use of curved element edges on the boundaries. We present some numerical results that illustrate the salient features and highlight the performance of the proposed solution methodology on the resonance phenomenon existing in the elastic scatterer for simple geometries such as circles. Moreover, the designed method ensures a convergence order with a gain of two order of magnitude compared to polygonal boundaries, and a potential to address both mid- and high-frequency regimes. These results have been presented to ECCOMAS 2012 [44] and to two workshops [42] [43].

### *6.2.8. Operator Based Upscaling for Discontinuous Galerkin Methods*

**Participants:** Hélène Barucq, Théophile Chaumont, Julien Diaz, Christian Gout, Victor Péron.

Scientists and engineers generally tackle problems that include multiscale effects and that are thus difficult to solve numerically. The main difficulty is to capture both the fine and the coarse scales to get an accurate numerical solution. Indeed, the computations are generally performed by using numerical schemes based on grids. But the stability and thus the accuracy of the numerical method depends on the size of the grid which must be refined drastically in the case of very fine scales. That implies huge computational costs and in particular the limitations of the memory capacity are often reached. It is thus necessary to use numerical methods that are able to capture the fine scale effects with computations on coarse meshes. Operator-based upscaling is one of them and we present in [22] a first attempt to adapt that technique to a Discontinuous Galerkin Method (DGM). We consider the Laplace problem as a benchmark and we compare the performance of the resulting numerical scheme with the classical one using Lagrange finite elements. The comparison involves both an accuracy analysis and a complexity calculus. This work shows that there is an interest of combining DGM with upscaling.

### *6.2.9. Asymptotic Modeling for Elasto-Acoustics*

**Participants:** Julien Diaz, Victor Péron.

We present in [65] equivalent conditions and asymptotic models for the diffraction problem of elasto-acoustic waves in a solid medium surrounded by a thin layer of fluid medium. This problem is well suited for the notion of equivalent conditions : since the thickness of the layer is small with respect to the wavelength, the effect of the fluid medium on the solid is as a first approximation local. We derive and validate equivalent conditions up to the third order for the elastic displacement. These conditions approximate the acoustic waves which propagate in the fluid region. This approach leads us to solve only elastic equations. The construction of equivalent conditions is based on a multiscale expansion in power series of the thickness of the layer for the solution of the transmission problem.

Questions regarding the implementation of the conditions have been addressed carefully. Indeed, the boundary conditions have been integrated without changing the structure of the code Hou10ni.

This work has been presented in four international conferences and Workshops : Aquitaine-Euskadi Workshop on Applied Mathematics; First Russian-French Conference on Mathematical Geophysics, Mathematical Modeling in Continuum Mechanics and Inverse Problems; Workshop HPC-GA; Twelfth International Conference Zaragoza-Pau on Mathematics.

A paper with numerical results for the elasto-acoustic problem with a thin layer and a variable thickness is in preparation.

### *6.2.10. Asymptotic modeling in electromagnetism*

**Participants:** François Buret, Monique Dauge, Patrick Dular, Laurent Krähenbühl, Victor Péron, Ronan Perrussel, Clair Poignard, Damien Voyer.

The following results rely on a problematic developed in section 3.2 , item **Asymptotic modeling**.

In the paper [28], eddy current problems are addressed in a bidimensional setting where the conducting medium is non-magnetic and has a corner singularity. For any fixed skin depth we show that the flux density is bounded near the corner, unlike the perfect conducting case. Then as the skin depth goes to zero, the first two terms of a multiscale expansion of the magnetic potential are introduced to tackle the magneto-harmonic problem. The heuristics of the method are given and numerical computations illustrate the obtained accuracy.

In a forthcoming paper, we describe the magnetic potential in the vicinity of a corner of a conducting body embedded in a dielectric medium in a bidimensional setting. We make explicit the corner asymptotic expansion for this potential as the distance to the corner goes to zero. This expansion involves singular functions and singular coefficients. We introduce a method for the calculation of the singular functions near the corner and we provide two methods to compute the singular coefficients: the method of moments and the method of quasi-dual singular functions. Estimates for the convergence of both approximate methods are proven. We eventually illustrate the theoretical results with finite element computations. The specific non-standard feature of this problem lies in the structure of its singular functions: They have the form of series whose first terms are harmonic polynomials and further terms are genuine non-smooth functions generated by the piecewise constant zeroth order term of the operator. This work has been presented in the international conference WCCM 2012.

### 6.2.11. Asymptotic models for penalization methods in porous media

**Participants:** Gilles Carbou, Victor Péron.

We investigate a Stokes-Brinkman problem with Beavers and Joseph transmission conditions, adapted to a penalization method in porous media. We exhibit a WKB expansion for the solution of the fluid-porous interface problem. The main interest is to derive equivalent models for the penalization method. We explicit the first terms of the WKB expansion for the flow and the pressure in the subdomains. Each asymptotics of the flow writes as a sum of a tangential boundary layer term plus a standard term in the porous region. From the benefits or these boundary layers, we infer a collection of elementary transmission problems satisfied by the standard parts of the asymptotics for the flow and the pressure. As a consequence of the penalization of the Laplacian operator which applies to the flow in the porous media, a degenerate operator of order zero applies to the elementary velocities appears in the porous region. The main difficulty concern the proof of elliptic regularity up to the interface for the solution of each elementary problem, since exotic conditions for the flow and the pressure appears along the interface. Our strategy consists to adapt a proof of elliptic regularity for the solution of a Darcy problem set in homogeneous media and developed by Boyer-Fabrie.

### 6.2.12. Asymptotic modeling in electromagnetism

**Participants:** Marc Duruflé, Victor Péron, Clair Poignard.

We investigate asymptotic models for 3D transmission problems in electromagnetism with homogeneous thin layers (uniform thickness). We exhibit Generalized Impedance Boundary Conditions of order 1 when the thin layer is symmetric and non-symmetric with respect to its mean surface. We present also a limit model for a resistive thin layer, and an equivalent model of order 1 for large contrast in conductivities through the thin layer. We write all these models in a general form. Questions regarding the implementation of the conditions have been addressed carefully. Numerical results with the high-order finite element library Montjoie illustrate the accuracy of the asymptotic models. A paper is in preparation.

### 6.2.13. Absorbing Boundary Conditions for Tilted Transverse Isotropic Elastic Media

**Participants:** Hélène Barucq, Lionel Boillot, Henri Calandra, Julien Diaz.

The simulation of wave propagation in geophysical media is often performed in domains which are huge compared to the wavelenghts of the problem. It is then necessary to reduce the computational domain to a box. When considering acoustic or elastic isotropic media, this can be done by applying an Absorbing Boundary Condition (ABC) or by adding a Perfectly Matched Layer (PML). However, a realistic representation of the Earth subsurface must include anisotropy and, in particular, the so-called Tilted Transverse Isotropy. Perfectly Matched Layers are known to be unstable for this kind of media and, to the best of our knowledge, no ABC have been proposed yet. We have thus proposed a low-order ABC for TTI media.

This ABC has been constructed for elliptic TTI media, where the slowness curve of the P-Wave is a rotated ellipse. Then, an appropriate change of variable can be applied in order to transform this ellipse into a circle. The main idea consists in imposing the isotropic ABC in the new system of coordinates and to apply the inverse change of variable in order to obtain the elliptic TTI ABC. We have compared numerically the reflections generated by this new ABC in TTI domain to the ones generated by the classical first order ABC in isotropic domains. The results show that the new ABC performs as well as the classical first order one. Moreover, this ABC seems to be also well-suited to non elliptic TTI media. These results have been presented at the Congrès Français d'Acoustique [35], at two workshops [39], [45].

### 6.2.14. Efficient solution methodology based on a local wave tracking strategy for high-frequency Helmholtz problems.

**Participants:** Mohamed Amara, Sharang Chaudhry, Julien Diaz, Rabia Djellouli, Steven Fiedler.

We have designed a new and efficient solution methodology for solving high-frequency Helmholtz problems. The proposed method is a least-squares based technique that employs variable bases of plane waves at the element level of the domain partition. A local wave tracking strategy is adopted for the selection of the basis at the regional/element level. More specifically, for each element of the mesh partition, a basis of plane waves is chosen so that one of the plane waves in the basis is oriented in the direction of the propagation of the field inside the considered element. The determination of the direction of the field inside the mesh partition is formulated as a minimization problem. Since the problem is nonlinear, we apply Newton's method to determine the minimum. The computation of Jacobians and Hessians that arise in the iterations of the Newton's method is based on the exact characterization of the Fréchet derivatives of the field with respect to the propagation directions. Such a characterization is crucial for the stability, fast convergence, and computational efficiency of the Newton algorithm. These results are part of the Master thesis of Sharang Chaudhry (student à CSUN) and have been presented to the 6th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS, Vienna, 2012).

## 6.3. High Performance methods for solving wave equations

**Participants:** Lionel Boillot, Hélène Barucq, Henri Calandra, Julien Diaz, Emiljana Jorgji, Didier Rémy, Florent Ventimiglia.

We have recently optimized the DG code implemented in the DIVA plateform of Total by reducing the number of communications between each processors. Since this code is based on the first order formulation of the elastodynamic wave equation, we have to compute three velocities and six stresses at each degree of freedom of the mesh. One naive idea consists in communicating these nine values at each time step. On the other hand, the computation of the three velocities does not actually require the knowledge of the six stresses but of three linear combination of these stresses. Similarly, the computation of the stresses requires the knowledge of six linear combinations of the three velocities. The main idea of the optimization consists in computing the three linear combinations of the stresses and to communicate them to the other processors, while the three velocities are communicated before computing the linear computations. Hence the number of communications can be reduced to six at each time step.

This optimization, coupled with the use of Hybrid MPI and OpenMP parallel programming has allowed to prove the scalability of the code up to 512 cores. We are now planning to extend these tests up to 4000 cores.

## MOISE Project-Team

# 6. New Results

## 6.1. Mathematical Modelling of the Ocean Dynamics

### *6.1.1. Beyond the traditional approximation on the Coriolis force*

**Participant:** Antoine Rousseau.

Formerly, A. Rousseau has performed some theoretical and numerical studies around the derivation of quasi-hydrostatic models. With C. Lucas, he proved that it is sometimes necessary to take into account the cosine part of the Coriolis force (which is usually neglected, leading to the so-called Traditional Approximation). They have also shown that the non-traditional terms do not raise any additional mathematical difficulty in the primitive equations: well-posedness for both weak and strong solutions.

A. Rousseau and J. McWilliams (UCLA) proposed a mathematical justification of the tilt of convective plumes in the quasi-geostrophic regime, thanks to the account of the complete Coriolis force in the so-called quasi-hydrostatic quasi-geostrophic (QHQG) model. The new model has been presented in international conferences [59] and [60].

### *6.1.2. Coupling Methods for Oceanic and Atmospheric Models*

**Participants:** Eric Blayo, David Cherel, Laurent Debreu, Antoine Rousseau, Manel Tayachi.

#### *6.1.2.1. Interface conditions for coupling ocean models*

Many physical situations require coupling two models with not only different resolutions, but also different physics. Such a coupling can be studied within the framework of global-in-time Schwarz methods. However, the efficiency of these iterative algorithms is strongly dependent on interface conditions. As a first step towards coupling a regional scale primitive equations ocean model with a local Navier-Stokes model, a study on the derivation of interface conditions for 2-D $x-z$ Navier-Stokes equations has been performed in D. Cherel PhD thesis. It has been shown theoretically that several usual conditions lead to divergent algorithms, and that a convergent algorithm is obtained when using transmission conditions given by a variational calculation.

D. Cherel has implemented a Schwarz-based domain decomposition method, for which he developed optimized absorbing boundary conditions that mix the velocity and pressure variables on an Arakawa-C grid. The numerical results confirm the rate of convergence that has been obtained theoretically, thanks to a Fourier analysis of the semi-discretized problem.

A first step towards the coupling between Navier-Stokes and primitive equations has been made in 2012. Starting from the optimized boundary conditions obtained for the Navier-Stokes equations, we performed an asymptotic analysis in order to obtain boundary conditions that should supplement the hydrostatic Navier-Stokes equations. These results have been presented in national and international conferences [47], [46], a paper is in preparation. David Cherel defended his PhD on Dec. 12th, 2012.

#### *6.1.2.2. Coupling dimensionally heterogeneous models*

The coupling of different types of models is gaining more and more attention recently. This is due, in particular, to the needs of more global models encompassing different disciplines (*e.g.* multi-physics) and different approaches (*e.g.* multi-scale, nesting). Also, the possibility to assemble different modeling units inside a friendly modelling software platform is an attractive solution compared to developing more and more global complex models. More specifically one may want to couple 1D to 2D or 3D models, such as Shallow Water and Navier Stokes models: this is the framework of our partnership with EDF in the project MECSICO. In her PhD, M. Tayachi is aimed to build a theoretical and numerical framework to couple 1D, 2D and 3D models for river flows.

In [65], we propose and analyze an efficient iterative coupling method for a dimensionally heterogeneous problem. We consider the case of a 2-D Laplace equation with non symmetric boundary conditions with a corresponding 1-D Laplace equation. We first show how to obtain the 1-D model from the 2-D one by integration along one direction, by analogy with the link between shallow water equations and the Navier-Stokes system. Then we focus on the design of a Schwarz-like iterative coupling method. We discuss the choice of boundary conditions at coupling interfaces. We prove the convergence of such algorithms and give some theoretical results related to the choice of the location of the coupling interface, and to the control of the difference between a global 2-D reference solution and the 2-D coupled one. These theoretical results are illustrated numerically. The extension of this work to shallow0water equations and primitive equations has been started recently.

### 6.1.3. *Numerical schemes for ocean modelling*

**Participants:** Laurent Debreu, Jérémie Demange.

Reducing the traditional errors in terrain-following vertical coordinate ocean models (or sigma models) has been a focus of interest for the last two decades. The objective is to use this class of model in regional domains which include not only the continental shelf, but the slope and deep ocean as well. Two general types of error have been identified: 1) the pressure-gradient error and 2) spurious diapycnal diffusion associated with steepness of the vertical coordinate. In a recent paper [87], we have studied the problem of diapycnal mixing. The solution to this problem requires a specifically designed advection scheme. We propose and validate a new scheme, where diffusion is split from advection and is represented by a rotated biharmonic diffusion scheme with flow-dependent hyperdiffusivity satisfying the Peclet constraint.
In 2012, in collaboration with F. Lemarié at UCLA, this work has been extended in order to render the biharmonic diffusion operator scheme unconditionally stable [17]. This is particularly needed when the slopes between coordinates lines and isopycnal surfaces are important so that the rotation of the biharmonic leads to strong stability condition along the vertical coordinate where the grid size is relatively small. This work also extends more classical results on the stability of laplacian diffusion with mixed derivatives.

In his PhD, Jérémie Demange begins a work on advection-diffusion schemes for ocean models (Supervisors : L. Debreu, P. Marchesiello (IRD)). His work will focus on the link between tracers (temperature and salinity) and momentum advection and diffusion in the non hyperbolic system of equations typically used in ocean models (the so called primitive equations with hydrostatic and Boussinesq assumptions). We also investigated the use of a depth dependent barotropic mode in free surface ocean models. When most ocean models assume that this mode is vertically constant, we have shown that the use of the true barotropic mode, derived from a normal mode decomposition, allows more stability and accuracy in the representation of external gravity waves [49], [48].

Salinity at 1000 m in the Southwest Pacific ocean is shown in figure 1 . The use of traditional upwind biased schemes (middle) exhibits a strong drift in the salinity field in comparison with climatology (left). The introduction of high order diffusion rotated along geopotential surfaces prevents this drift while maintaining high resolution features (right).

## 6.2. Data Assimilation for Geophysical Models

### 6.2.1. *Development of a Variational Data Assimilation System for OPA9/NEMO*

**Participants:** Arthur Vidard, Bénédicte Lemieux-Dudon, Pierre-Antoine Bouttier.

We are heavily involved in the development of NEMOVAR (Variational assimilation for NEMO). For several years now, we built a working group (coordinated by A. Vidard) in order to bring together various NEMOVAR user-groups with diverse scientific interests (ranging from singular vector and sensitivity studies to specific issues in variational assimilation) It has led to the creation of the VODA (Variational Ocean Data Assimilation for multi scales applications) ANR project (ended in 2012). A new project, part of a larger EU-FP7 project has been submitted late 2012.

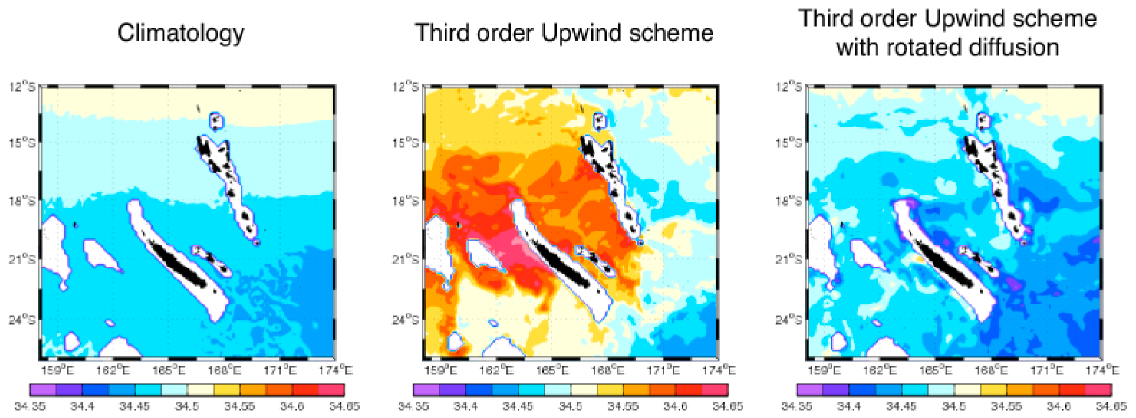


*Figure 1. Salinity at 1000m in the Southwest Pacific ocean.*

The project aims at delivering a common NEMOVAR platform based on NEMO platform for 3D and 4D variational assimilation. Following 2009-11 VODA activities, a fully parallel version of NEMOTAM (Tangent and Adjoint Model for NEMO) is now available for the community in the standard NEMO version. This version is based on the released 3.4.1 version of NEMO.

We are also investigating variational data assimilation methods applied to high resolution ocean numerical models. This part of the project is now well advanced and encouraging preliminary results are available on an idealized numerical configuration of an oceanic basin (see Figure 2 ). Several novative diagnostics have been also developed in this framework

### 6.2.2. Identification of pollution.

**Participant:** François-Xavier Le Dimet.

The problem is the next : potential sources of pollution are known but the contribution of each source to a local site is unknown. The problem is to identify the contribution of each source. This is a very common situation both at the local scale and at the synoptic scale. Thanks to second order methods we have been able to reach this goal, the theoretical part is done at FSU and application at the Institute of Mechanics of the Vietnamese Academy of Sciences. One paper has been submitted for publication [75]. At FSU M.Y. Hussaini and I. Souopgui are involved in this project

The quality of water ressources is an important problem for Vietnam. With scientists of the Institute of Mechanics (Ha Tran Thua, Hoang Van Lai, Nguyen Ba Hung) in [31] and [53] we have used the methods described in [75] for water pollution studies, in parallel Tran Thu Ha and Pham Dinh Tuan (LJK) have been working on the application of Kalman filter for this problem. Several talks have been given and papers published.

### 6.2.3. Variational data assimilation for large scale ice-sheet models

**Participants:** Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

In collaboration with C. Ritz (CNRS, Laboratoire de Glaciologie et Geophysique de l'Environnement (LGGE), Grenoble), we aim to develop adjoint methods for ice cap models.

In the framework of global warming, the evolution of sea level is a major but ill-known phenomenon. It is difficult to validate the models which are used to predict the sea level elevation, because observations are heterogeneous and sparse.

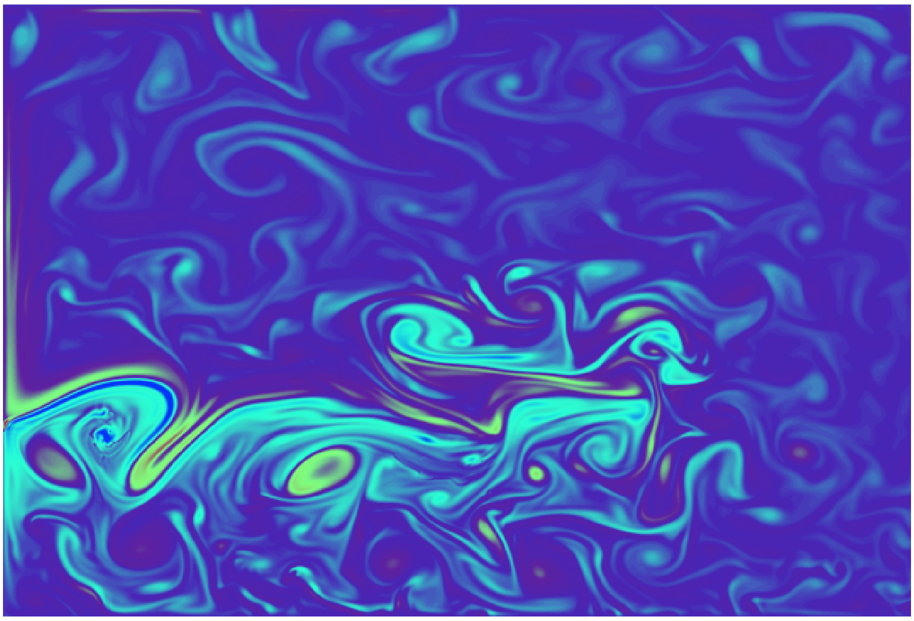


*Figure 2. Surface relative vorticity of a 1/24th of a degree NEMO configuration*

Data acquisition in polar glaciology is difficult and expensive. Satellite data have a good spatial coverage, but they allow only indirect observation of the interesting data. We wish to make the most of all available data and evaluate what new observations to add, where and when. Sensitivity analysis, and in particular the adjoint method, allows to identify the most influential parameters and variables and can help to design the observation network.

B. Bonan started his PhD in September 2010 on this subject. We implemented the 4D-Var algorithm for a flowline Shallow-Ice model, called Winnie, developed by C. Ritz at LGGE. In a simple configuration, we were able to generate the adjoint code by automatic differentiation. First results were encouraging and were presented at EGU [58] and Les Houches Summer School [30].

### 6.2.4. Ensemble Kalman filtering for large scale ice-sheet models

**Participants:** Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

We are also interested in comparing variational methods to stochastic filtering. In the framework of B. Bonan PhD, we then implemented Ensemble Transform Kalman Filter (ETKF) on Winnie, which we would like to compare to variational assimilation methods. First results are promising and were presented at three conferences [39], [37], [38].

### 6.2.5. Inverse methods for full-Stokes glaciology models

**Participants:** Olivier Gagliardini, Maëlle Nodet, Catherine Ritz.

We are investigating the means to apply inverse modeling to another class of glaciology models, called full-Stokes model. Such a model is developed by LGGE and CSC in Finland, called Elmer/Ice. Contrary to large scale models, Elmer/Ice is based on the full Stokes equations, and no assumptions regarding aspect ratio are made, so that this model is well adapted to high resolution small scale modelling, such as glaciers (and more recently the whole Greenland ice-sheet).

In collaboration with O. Gagliardini, F. Gillet-Chaulet and C. Ritz (Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), Grenoble), we investigated a new method to solve inverse problems for a Full-Stokes model of Groenland, which consisted in solving iteratively a sequence of Neumann and Dirichlet problems within a gradient descent algorithm. We also compared this method to an approximate variational algorithm, using the fact that the full Stokes equations are almost self-adjoint. These results have been published in The Cryosphere Discussion [11] and presented at three conferences [52], [28], [57]. Figure 3 presents the reconstructed surface velocities compared to the observations, where we can see a good agreement of the main features, thus success of the assimilation process.

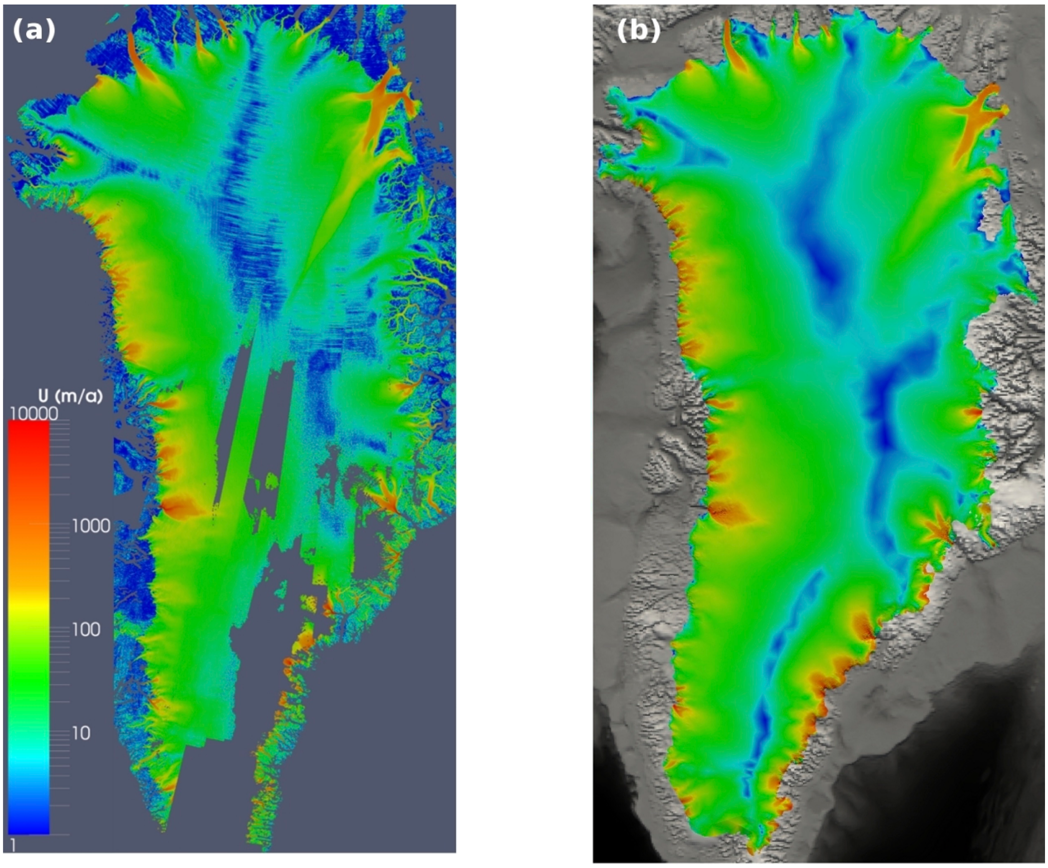


*Figure 3. Surface velocities of Greenland Ice Sheet, in meters per year. On the left (a), the velocities which are observed by satellites. On the right (b), velocities obtained after assimilation.*

### 6.2.6. Dating ice matrix and gas bubbles with DatIce

**Participants:** Eric Blayo, Bénédicte Lemieux-Dudon, Habib Toye Mahamadou Kele.

H. Toye Mahamadou Kele joined the MOISE team for 2 years as an Inria young engineer. A shared memory parallelization of the code and a more friendly user interface have been developed. Efforts have been made to calibrate the error covariance matrices by the mean of a posteriori diagnostics.

The MOISE team was involved in the Antarctic Ice Core Chronology 2012 (AICC2012) through a tight collaboration with the Laboratoire de Glaciologie et de Géophysique de l'Environnement (LGGE), the

Laboratoire des Sciences du Climat et de l'Environnement (LSCE), and other European laboratories. The AICC2012 project aimed at constructing an unified chronology for several Antarctic ice cores. A special issue is dedicated to AICC2012 in Climate of the Past http://www.clim-past-discuss.net/special_issue53.html. MOISE efforts on the DatIce code lead to two important articles currently reviewed in the Open Discussion process http://www.clim-past-discuss.net/papers_in_open_discussion.html: [2] and [22].

## 6.3. Development of New Methods for Data Assimilation

### 6.3.1. Variational Data Assimilation with Control of Model Error

**Participants:** Bénédicte Lemieux-Dudon, Arthur Vidard.

One of the main limitation of the current operational variational data assimilation techniques is that they assume the model to be perfect mainly because of computing cost issues. Numerous researches have been carried out to reduce the cost of controlling model errors by controlling the correction term only in certain privileged directions or by controlling only the systematic and time correlated part of the error.

Both the above methods consider the model errors as a forcing term in the model equations. Trémolet (2006) describes another approach where the full state vector (4D field: 3D spatial + time) is controlled. Because of computing cost one cannot obviously control the model state at each time step. Therefore, the assimilation window is split into sub-windows, and only the initial conditions of each sub-window are controlled, the junctions between each sub-window being penalized. One interesting property is that, in this case, the computation of the gradients, for the different sub-windows, are independent and therefore can be done in parallel.

This method is now implemented in a realistic Oceanic framework using OPAVAR/ NEMOVAR. An extensive documentation has been produced and we are now assessing the improvement over the previous scheme

### 6.3.2. Variational Data Assimilation and Control of Boundary Conditions

**Participant:** Eugène Kazantsev.

A variational data assimilation technique applied to the identification of the optimal discretization of interpolation operators and derivatives in nodes that are adjacent to the boundary of the domain is discussed in two contexts: a simplified case of a shallow water model and the ORCA-2 configuration of the NEMO model.

Experiments with a non-linear shallow water model in [14] show that controlling the discretization of operators near a rigid boundary can bring the model solution closer to observations both within and beyond the assimilation window. This type of control allows also to improve climatic variability of the model. These properties have been studied in two different configurations: an academic case of assimilation of artificially generated observational data in a square box configuration and assimilation of real observations in a model of the Black sea.

The sensitivity of the shallow water model in the previously described configurations has been studied in detail in [15]. It is shown in both experiments that boundary conditions near a rigid boundary influence the solution higher than the initial conditions. This fact points out the necessity to identify optimal boundary approximation during a model development.

Considering a full-physics global ocean model, we apply the 4D-Var data assimilation technique to ORCA-2 configuration of the NEMO in order to identify the optimal parametrization of boundary conditions on the lateral boundaries as well as on the bottom and on the surface of the ocean [71]. The influence of boundary conditions on the solution is analyzed as in the assimilation window and beyond the window. It is shown that optimal surface and bottom boundary conditions allow us to better represent the jet streams, such as Gulf Stream and Kuroshio. Sea Surface Height in the North Atlantic before and after control is shown in fig.4 ). Analyzing the reasons of the jets reinforcement, we notice that data assimilation has a major impact on parametrization of the bottom boundary conditions for $u$ and $v$ [55].

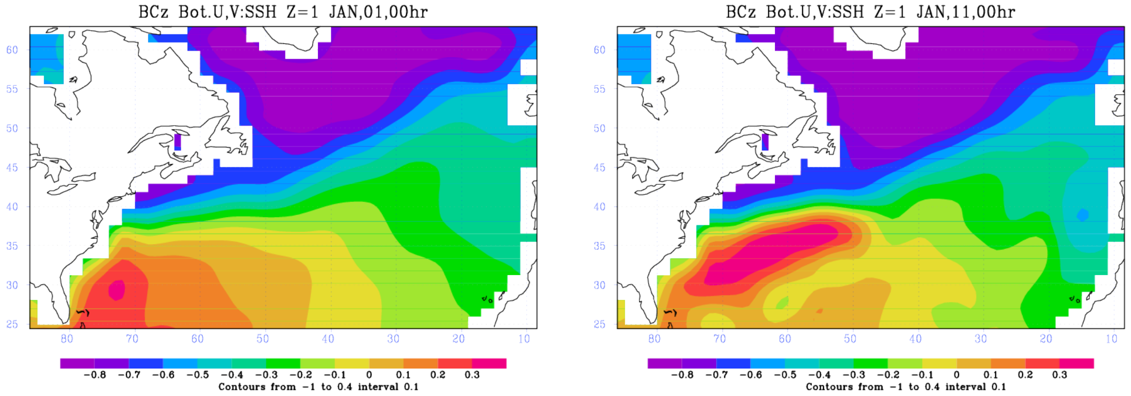


*Figure 4. Sea surface elevation in the North Atlantic on the 1st (classical boundary) and on the 11th (optimal boundary) of January, 2006.*

Adjoint models, necessary to variational data assimilation have been produced by the TAPENADE software, developed by the TROPICS 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.

### 6.3.3. Direct assimilation of sequences of images

**Participants:** François-Xavier Le Dimet, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Vincent Chabot.

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 current follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France)

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [95]. 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 [78]. 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 mains structures within each image. This can be done using, for example, a wavelet representation of images. We are also part of TOMMI, another ANR project started mid 2011, where we are investigating the possibility to use optimal transportation based distances for images assimilation.

### 6.3.4. Assimilation of ocean images

**Participants:** Vincent Chabot, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard, Alexandros Makris.

In addition with the direct assimilation approach previously described, a particular attention has been given to the study of data noise for ocean image assimilation. A journal paper is about to be submitted on this subject in Tellus A. In the context of assimilation of structures contained in satellite images, a two step registration approach using computer vision tools has been proposed within the post-doctorate of Alexandros Makris [33], [56].

## 6.4. Quantifying Uncertainty

### *6.4.1. Sensitivity analysis for West African monsoon*

**Participants:** Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry.

#### *6.4.1.1. Geophysical context*

The West African monsoon is the major atmospheric phenomenon which drives the rainfall regime in Western Africa. Therefore, this is the main phenomenon in water resources over the African continent from the equatorial zone to the sub-Saharian one. Obviously, it has a major impact on agricultural activities and thus on the population itself. The causes of inter-annual spatio-temporal variability of monsoon rainfall have not yet been univocally determined. Spatio-temporal changes on the see surface temperature (SST) within the Guinea Gulf and Saharian and Sub-Saharian Albedo are identified by a considerable body of evidences as major factors to explain it.

The aim of this study is to simulate the rainfall by a regional atmospheric model (RAM) and to analyze its sensitivity to the variability of these inputs parameters. Once precipitations from RAM are compared to several precipitation data sets we can observe that the RAM simulates the West African monsoon reasonably.

#### *6.4.1.2. Statistical methodology*

As mentioned in the previous paragraph, our main goal is to perform a sensitivity analysis for the West African monsoon. Each simulation of the regional atmospheric model (RAM) is time consuming, and we first have to think about a simplified model. We deal here with spatio-temporal dynamics, for which we have to develop functional efficient statistical tools. In our context indeed, both inputs (albedo, SST) and outputs (precipitations) are considered as time and space indexed stochastic processes. A first step consists in proposing a functional modeling for both precipitation and sea surface temperatures, based on a new filtering method. For each spatial grid point in the Gulf of Guinea and each year of observation, the sea surface temperature is measured during the active period on a temporal grid. A Karhunen-Loève decomposition is then performed at each location on the spatial grid [97]. The estimation of the time dependent eigenvalues at different spatial locations generates great amounts of high-dimensional data. Clustering algorithms become then crucial in reducing the dimensionality of such data.

Thanks to the functional clustering performed on the first principal component at each point, we have defined specific subregions in the Gulf of Guinea. On each subregion, we then choose a referent point for which we keep a prescribed number of principal components which define the basis functions. The sea surface temperature at any point in this subregion is modeled by the projection on this truncated basis. The spatial dependence is described by the coefficients of the projection. The same approach is used for precipitation. Hence for both precipitation and sea surface temperatures, we obtain a decomposition where the basis functions are functions depending on time and whose coefficients are spatially indexed and time independent. Then, the most straightforward way to model the dependence of precipitation on sea surface temperatures is through a multivariate response linear regression model with the output (precipitation) spatially indexed coefficients in the above decomposition and the input (SST) spatially indexed coefficients being predictors. A naive approach consists in regressing each response onto the predictors separately; however it is unlikely to produce satisfactory results, as such methods often lead to high variability and over-fitting. Indeed the dimensions of both predictors and responses are large (compared to the sample size).

We apply a novel method recently developed by [91] in integrated genomic studies which takes into account both aspects. The method uses an $\ell_1$-norm penalty to control the overall sparsity of the coefficient matrix of the multivariate linear regression model. In addition, it also imposes a *group* sparse penalty. This penalty puts a constraint on the $\ell_2$ norm of regression coefficients for each predictor, which thus controls the total number of predictors entering the model, and consequently facilitates the detection of important predictors. The dimensions of both predictors and responses are large (compared to the sample size). Thus in addition to assuming that only a subset of predictors enter the model, it is also reasonable to assume that a predictor may affect only some but not all responses. By the way we take into account the complex and spatio-temporal dynamics. This work has been published in [1].

#### *6.4.1.3. Distributed Interactive Engineering Toolbox*

An important point in the study described above is that the numerical storage and processing of model inputs/outputs requires considerable computation resources. They were performed in a grid computing environment with a middleware (DIET) which takes into account the scheduling of a huge number of computation requests, the data-management and gives a transparent access to a distributed and heterogeneous platform on the regional Grid CIMENT (http://ciment.ujf-grenoble.fr/).

Thus, a different DIET module was improved through this application. An automatic support of a data grid software (http://www.irods.org) through DIET and a new web interface designed for MAR was provided to physicians.
These works involve also partners from the Inria project/team GRAAL for the computational approach, and from the Laboratory of Glaciology and Geophysical Environment (LGGE) for the use and interpretation of the regional atmospheric model (RAM).

### *6.4.2. Tracking for mesoscale convective systems*

**Participants:** Anestis Antoniadis, Céline Helbert, Clémentine Prieur, Laurence Viry, Roukaya Keinj.

#### *6.4.2.1. Scientific context*

In this section, we are still concerned with the monsoon phenomenon in western Africa and more generally with the impact of climate change. What we propose in this study is to focus on the analysis of rainfall system monitoring provided by satellite remote sensing. The available data are micro-wave and IR satellite data. Such data allow characterizing the behavior of the mesoscale convective systems. We wish to develop stochastic tracking models, allowing for simulating rainfall scenari with uncertainties assessment.

#### *6.4.2.2. Stochastical approach*

The chosen approach for tracking these convective systems and estimating the rainfall intensities is a stochastic one. The stochastic modeling approach is promising as it allows developping models for which confidence in the estimates and predictions can be evaluated. The stochastic model will be used for hydro-climatic applications in West Africa. The first part of the work will consist in implementing a model developed in [96] on a test set to evaluate its performances, our ability to infer the parameters, and the meaning of these parameters. Once the model well fitted on toy cases, this algorithm should be run on our data set, and compared with previous results by [89] or by [88]. The model developed by [96] is a continuous time stochastic model to multiple target tracking, which allows in addition to birth and death, splitting and merging of the targets. The location of a target is assumed to behave like a Gaussian Process when it is observable. Targets are allowed to go undetected. Then, a Markov Chain State Model decides when the births, death, splitting or merging of targets arise. The tracking estimate maximizes the conditional density of the unknown variables given the data. The problem of quantifying the confidence in the estimate is also addressed. Roukaya Keinj started working on this topic with a two years postdoctoral position in November 2011. She left the team in October 2012, and is now replaced by Alexandros Makris.

### *6.4.3. Sensitivity analysis for forecasting ocean models*

**Participants:** Anestis Antoniadis, Eric Blayo, Gaëlle Chastaing, Céline Helbert, Alexandre Janon, François-Xavier Le Dimet, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Jean-Yves Tissot, Federico Zertuche.

#### 6.4.3.1. Scientific context

Forecasting ocean systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

#### 6.4.3.2. Estimating sensitivity indices

A first task is to develop tools for estimated sensitivity indices. Among various tools a particular attention was first paid to FAST and its derivatives. In [21], the authors present a general way to correct a positive bias which occurs in all the estimators in random balance design method (RBD) and in its hybrid version, RBD-FAST. Both these techniques derive from Fourier amplitude sensitivity test (FAST) and, as a consequence, are faced with most of its inherent issues. And up to now, one of these, the well-known problem of interferences, has always been ignored in RBD. After presenting in which way interferences lead to a positive bias in the estimator of first-order sensitivity indices in RBD, the authors explain how to overcome this issue. They then extend the bias correction method to the estimation of sensitivity indices of any order in RBD-FAST. They also give an economical strategy to estimate all the first-order and second-order sensitivity indices using RBD-FAST. A more theoretical work [77] revisit FAST and RBD in light of the discrete Fourier transform (DFT) on finite subgroups of the torus and randomized orthogonal array sampling. In [77] the authors study the estimation error of both these methods. This allows to improve FAST and to derive explicit rates of convergence of its estimators by using the framework of lattice rules. A natural generalization of the classic RBD is also provided, by using randomized orthogonal arrays having any parameters, and a bias correction method for its estimators is proposed. In variance-based sensitivity analysis, another classical tool is the method of Sobol' [94] 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 [93] and more recently by Owen [90] but the quantities they estimate still require $O(d)$ samples. In a recent work [76] the authors introduce a new approach to estimate for any $k$ all the $k$-th order Sobol' indices by using only two samples based on replicated latin hypercubes. They establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, they 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.

#### 6.4.3.3. Intrusive sensitivity analysis, reduced models

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 [12], 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. The present preprint is under review. 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 [13]. In a more recent work [69], the authors deal with asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. Implementations have to be conducted on more general models such as Shallow-Water models. 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 senses has been submitted, dealing with goal oriented uncertainties assessment [70].

#### *6.4.3.4. 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 [82] in the case where input parameters are correlated. A PhD started in October 2010 on this topic (Gaëlle Chastaing). We obtained first results [4], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs.

#### *6.4.3.5. Multy-fidelity modeling for risk analysis*

Federico Zertuche 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 ( [83], [92] ) : a new estimating 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 year 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.

### *6.4.4. Multivariate risk indicators*

In collaboration with Véronique Maume-Deschamps (ISFA Lyon 1), Elena Di Bernardino (CNAM), Anne-Catherine Favre (LTHE Grenoble) and Peggy Cenac (Université de Bourgogne), we are interested in defining and estimating new multivariate risk indicators. This is a major issue with many applications (environmental, insurance, ...). Two papers were accepted for publication and two other ones are submitted. The first submitted one deals with the estimation of bivariate tails [79]. In [81] and [68] we propose estimation procedures for multivariate risk indicators. In [5] we propose to minimize multivariate risk indicators by using a Kiefer-Wolfowitz approach to the mirror stochastic algorithm.

### *6.4.5. Quasi-second order analysis for the propagation and characterization of uncertainties in geophysical prediction*

We have developed a new approach for the propagation and characterization of uncertainties in geophysical prediction. Most of the method presently used are based on Monte-Carlo type (ensemble) methods, they are expensive from the computational point of view and have received a poor theoretical justification especially in the case of strongly non linear models. We have proposed a new method based on quasi-second order analysis, with a theoretical background and robust for strongly non linear models. Several papers have been published [20], [10], [19], [51] and the application to complex models are presently under development. Igor Gejadze and Victor Shutyaev have been staying both for a total of four weeks in MOISE.

## 6.5. Image processing

### *6.5.1. Image processing*

**Participant:** Nicolas Papadakis.

In collaboration with the Inria team MC2 of the Bordeaux-Sud-Ouest center, we investigate the application of image assimilation to medical issues. The objective is here to use MRI images in order to monitor EDP models dealing with tumor growth in lungs or brains. Using such images, we would like to define a patient specific process allowing to calibrate the numerical model with respect to the observed tumor. First works based on convex relaxation of the binary segmentation problem [34] have been realized in this direction by proposing a 3D segmentation method dedicated to glioblastomas from a set of MRI brain images. The obtained automatic segmentation results are very close to specialist manual segmentations (errors of 5%) and will be used as pseudo-observations for an assimilation system based on the numerical model describing the tumor growth. The final issue will be to define an observation operator linking images with the model in order to realize a direct assimilation.

Next, in collaboration with Vicent Caselles (Pompeu Fabra University, Barcelona, Spain) we tackled the problem of histogram equalization of different images. Our aim has been to include spatial information on color repartition during the histogram transfer for inpainting and shadow removal purposes [18]. We also focused with Jean-François Aujol (Institut de Mathématiques de Bordeaux), on the convexification of non linear problems such as optical flow estimation and submitted a jounal paper on this subject in SIAM Journal on Imaging Sciences.

### 6.5.2. *Optimal transport*

**Participants:** Maëlle Nodet, Nicolas Papadakis, Arthur Vidard.

Within the optimal transport project TOMMI funded by the ANR white program, some new algorithms had been proposed to take into account the physics (rigidity, elesticity) of the density to transport [40]. A journal paper has been submitted on this topic in M2AN.

## 6.6. Mathematical modelling for CFD-environment coupled systems

**Participant:** Antoine Rousseau.

### 6.6.1. *Minimal-time bioremediation of natural water resources*

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. The originality of the approach is to couple minimal time strategies that are determined on a simplified model with a faithful numerical model for the hydrodynamics. Based on a previous paper that deals with an an implicit representation of the spatial inhomogenity of the resource with a small number of homogeneous compartments (with a system of ODEs), we implement a coupled ODE-PDE system that accounts for the spatial non-homogeneity of pollution in natural resources. The main idea is to implement a Navier-Stokes model in the resource (such as a lake), with boundary conditions that correspond to the output feedback that has been determined to be optimal for the simple ODEs model of a (small) bioreactor. A first mathematical model has been introduced and numerical simulations have been performed in academic situations. During the internship of S. Barbier (co-advised by A. Rousseau and A. Rapaport (INRA-MODEMIC)) we built a reduced model that approximates the reference PDE model thanks to a set of ODEs with parameters. Numerical optimization is performed on these parameters in order to better fit the reference model. This will lead to a publication. In addition, bioremediation algorithms proposed by the authors have been sent to Inria Technology Transfert Services for a patent registration.

### 6.6.2. *Mathematical modelling for the confinement of lagoons*

This work deals with the concept of confinement of paralic ecosystems. It is based on a recent paper by E. Frénod that presents a modelling procedure in order to compute the confinement field of a lagoon.
In [9], A. Rousseau and E. Frénod improve the existing model in order to account for tide oscillations in any kind of geometry such as a non-rectangular lagoons with a non-flat bottom. The new model, that relies on PDEs rather than ODEs, is then implemented thanks to the finite element method. Numerical results confirm the feasibility of confinement studies thanks to the introduced model. During the internship of J.-P. Bernard, we implemented the proposed method in a realistic situation, namely the Etang de Thau in Languedoc-Roussillon, France (see Figure 5 ). This was presented in an international conference [60].

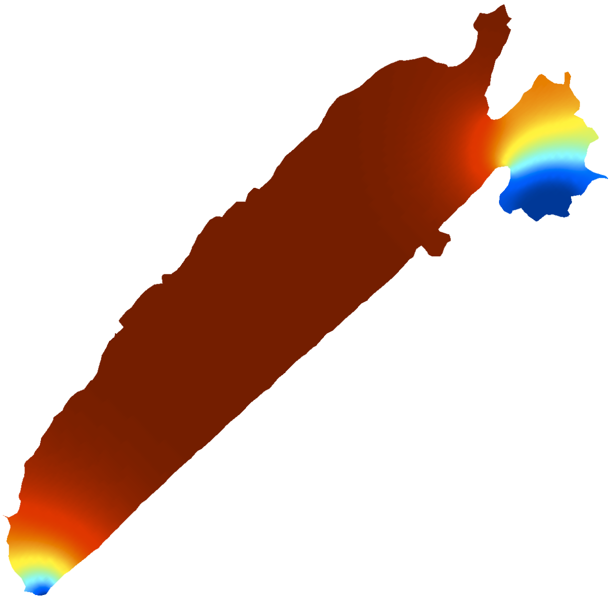


*Figure 5. Confinement map in the Thau Lagoon (France). See [60].*

## 6.7. $CO_2$ Storage

**Participant:** Céline Helbert.

In collaboration with Bernard Guy (EMSE, Saint-Etienne) and more specifically in the context the PhD of Joharivola Raveloson (EMSE, Saint-Etienne), we are interested in the study of the water-rock interactions in the case of $CO_2$ storage in geological environment. This work is following the study of Franck Diedro in the same subject [8]. In this study we focus on the scale of observation of geochemical phenomena while taking into account the heterogeneity of the reservoir. This heterogeneity at small and large scale helps to maintain a local variability of the chemical composition and influence reaction rates at the pore as well at the reservoir scale. To connect the parameters at both scale (pore and reservoir) we use deterministic and stochastic simulations of a reactive transport code developed by IFPEN.

## 6.8. Land Use and Transport models calibration

**Participants:** Clémentine Prieur, Nicolas Papadakis, Arthur Vidard.

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 lead by the STEEP EPI. This project has just been accepted and will start early 2013.

# POMDAPI Project-Team (section vide)

# SAGE Project-Team

# 6. New Results

## 6.1. Parallelism and convergence in iterative linear solvers

### 6.1.1. Generation of Krylov subspace bases

**Participant:** Bernard Philippe.

This work was done in collaboration with L. Reichel, from University of Kent, USA (see 8.3.1 ).

It is published in a journal [19].

Many problems in scientific computing involving a large sparse square matrix A are solved by Krylov subspace methods. This includes methods for the solution of large linear systems of equations with A, for the computation of a few eigenvalues and associated eigenvectors of A, and for the approximation of nonlinear matrix functions of A. When the matrix A is non-Hermitian, the Arnoldi process commonly is used to compute an orthonormal basis for a Krylov subspace associated with A. The Arnoldi process often is implemented with the aid of the modified Gram–Schmidt method. It is well known that the latter constitutes a bottleneck in parallel computing environments, and to some extent also on sequential computers. Several approaches to circumvent orthogonalization by the modified Gram–Schmidt method have been described in the literature, including the generation of Krylov subspace bases with the aid of suitably chosen Chebyshev or Newton polynomials. We review these schemes and describe new ones. Numerical examples are presented.

### 6.1.2. Parallel Adaptive Deflated GMRES

**Participants:** Jocelyne Erhel, Bernard Philippe.

This work was done in the context of the joint Inria/ NCSA laboratory on petascale computing (see 8.3.7 ), and the c2sexa project (see 8.1.3 ). Computations were done with GENCI supercomputers (see 8.1.6 ), using the software GPREMS, AGMRES, DGMRES (see 5.7 , 5.8 , 5.9 ).

It was presented at two conferences [30] [29], is published in proceedings [39] and is submitted (in revision) to a journal [46]. The algorithms are implemented in the software DGMRES and AGMRES, which are freely available in the PETSC repository.

The GMRES iterative method is widely used as Krylov subspace accelerator for solving sparse linear systems when the coefficient matrix is nonsymmetric and indefinite. The Newton basis implementation has been proposed on distributed memory computers as an alternative to the classical approach with the Arnoldi process. The aim of our work here is to introduce a modification based on deflation and augmented techniques. This approach builds an augmented subspace or a preconditioning matrix in an adaptive way to accelerate the convergence of the restarted formulation. It can be combined with preconditioning methods based for example on domain decomposition. In our numerical experiments, we show the benefits of our method to solve large linear systems.

### 6.1.3. Memory efficient hybrid algebraic solvers for linear systems arising from compressible flows

**Participants:** Jocelyne Erhel, Bernard Philippe.

This work was done in collaboration with FLUOREM company, in the context of the joint Inria/ NCSA laboratory on petascale computing (see 8.3.7 ) and the C2S@EXA project (see 8.1.3 ). Computations were done with GENCI supercomputers (see 8.1.6 ), using the software GPREMS, AGMRES, DGMRES (see 5.7 , 5.8 , 5.9 ).

It has been published in a journal [18].

This paper deals with the solution of large and sparse linear systems arising from design optimization in Computational Fluid Dynamics. From the algebraic decomposition of the input matrix, a hybrid robust direct/iterative solver is often defined with a Krylov subspace method as accelerator, a domain decomposition method as preconditioner and a direct method as subdomain solver. The goal of this paper is to reduce the memory requirements and indirectly the computational cost at different steps of this scheme. To this end, we use a grid-point induced block approach for the data storage and the partitioning part, a Krylov subspace method based on the restarted GMRES accelerated by deflation, a preconditioner formulated with the restricted additive Schwarz method and an aerodynamic/turbulent fields split at the subdomain level. Numerical results are presented with industrial test cases to show the benefits of these choices.

### *6.1.4. Efficient parallel implementation of the fully algebraic multiplicative Aitken-RAS preconditioning technique*

**Participant:** Thomas Dufaud.

This work was done in collaboration with D. Tromeur-Dervout, from ICJ, University of Lyon and has been published in a journal [14].

This paper details the software implementation of the ARAS preconditioning technique [48], in the PETSc framework. Especially, the PETSc implementation of interface operators involved in ARAS and the introduction of a two level of parallelism in PETSc for the RAS are described. The numerical and parallel implementation performances are studied on academic and industrial problems, and compared with the RAS preconditioning. For saving computational time on industrial problems, the Aitken's acceleration operator is approximated from the singular values decomposition technique of the RAS iterate solutions.

### *6.1.5. An algebraic multilevel preconditioning framework based on information of a Richardson process*

**Participant:** Thomas Dufaud.

This work was done in the context of the C2S@EXA project (see 8.1.3 ).

It has been presented at a conference [23] and submitted to the proceedings.

A fully algebraic framework for constructing coarse spaces for multilevel preconditioning techniques is proposed. Multilevel techniques are known to be robust for scalar elliptic Partial Differential Equations with standard discretization and to enhance the scalability of domain decomposition method such as RAS preconditioning techniques. An issue is their application to linear system encountered in industrial applications which can be derived from non-elliptic PDEs. Moreover, the building of coarse levels algebraically becomes an issue since the only known information is contained in the operator to inverse. Considering that a coarse space can be seen as a space to represent an approximated solution of a smaller dimension than the leading dimension of the system, it is possible to build a coarse level based on a coarse representation of the solution. Drawing our inspiration from the Aitken-SVD methodology, dedicated to Schwarz methods, we proposed to construct an approximation space by computing the Singular Value Decomposition of a set of iterated solutions of the Richardson process associated to a given preconditioner. This technique does not involve the knowledge of the underlying equations and can be applied to build coarse levels for several preconditioners. Numerical results are provided on both academic and industrial problems, using two-level additive preconditioners built with this methodology.

## 6.2. Parallel numerical algorithms

### *6.2.1. High Performance Scientific Computing*

**Participant:** Bernard Philippe.

This work was done in collaboration with several authors, from US, Greece, etc. (see 8.3.1 and 8.2.1 ).

A book appeared on this subject in 2012 [45] and a chapter of this book is devoted to a historical perspective [44].

This comprehensive text/reference, inspired by the visionary work of Prof. Ahmed H. Sameh, represents the state of the art in parallel numerical algorithms, applications, architectures, and system software. Articles in this collection address solutions to various challenges arising from concurrency, scale, energy efficiency, and programmability. These solutions are discussed in the context of diverse applications, ranging from scientific simulations to large-scale data analysis and mining.
As exascale computing is looming on the horizon while multicore and GPU's are routinely used, we survey the achievements of Ahmed H. Sameh, a pioneer in parallel matrix algorithms. Studying his contributions since the days of Illiac IV as well as the work that he directed and inspired in the building of the Cedar multiprocessor and his recent research, unfolds a useful historical perspective in the field of parallel scientific computing.

### 6.2.2. *Counting eigenvalues in domains of the complex field*

**Participant:** Bernard Philippe.

This work is done in collaboration with E. Kamgnia, from the University of Yaounde 1, Cameroon, in the context of the MOMAPLI project at LIRIMA (see 8.3.5 ).

It is accepted for publication in a journal [15], and was presented in conferences [31], [32], [38], [40].

A procedure for counting the number of eigenvalues of a matrix in a region surrounded by a closed curve is presented. It is based on the application of the residual theorem. The quadrature is performed by evaluating the principal argument of the logarithm of a function. A strategy is proposed for selecting a path length that insures that the same branch of the logarithm is followed during the integration. Numerical tests are reported for matrices obtained from conventional matrix test sets.

The procedure is now combined with the PPAT methodology (see 5.10 ). A list of triangles is built for overlapping the boundary of the pseudo-spectra. From the list of vertices, a closed polygonal line is defined and the number of enclosed eigenvalues is determined.

### 6.2.3. *Ratio-Based Parallel Time Integration*

**Participant:** Jocelyne Erhel.

This work is done in in the context of the MODNUM project (see 8.3.2 ), in collaboration with American University of Beirut (AUB), Lebanon.

It was presented at a conference [41] and is submitted to the proceedings. It was also presented at a seminar of Inria Rennes.

Because time-integration of time-dependent problems is inherently sequential, time parallelism aims mainly at reducing the computational time of some real-time evolutionary problems and may be done through predictor-corrector schemes.

We apply the rescaling method onto initial value problems having an explosive or oscillatory solution, in infinite time. We show how a relevant choice of the end-of-slice condition and the time-rescaling factor might lead to rescaled systems having a uniform convergence to a limit problem. This property provides much better predictions and enhances the relevance of RaPTI that consists mainly of (i) the little sequential computations it involves (predictions and corrections are done in parallel), (ii) the relatively low communication cost it induces and (iii) the similarity of the computation on all slices yielding similar computational times on all processors. Hence, significant speed-ups are achieved. This is illustrated on two problems: a non-linear diffusion-reaction problem having an explosive solution, and a membrane problem having an oscillatory and explosive solution.

## 6.3. Numerical models and simulations applied to physics

### 6.3.1. *Heat transfer modeling in saturated porous media*

**Participant:** Édouard Canot.

This work is done in the context of the ARPHYMAT project (see 8.3.3 ) and the MODNUM project (see 8.3.2 ), in collaboration with Archeosciences, IPR and Lebanese International University (LIU), Lebanon. It was also done in the context of Caroline Thoux's internship (L3, INSA Rennes).

This work is published in [17].

In this paper, the authors introduce a robust numerical strategy to estimate the temperature dependent heat capacity, thermal conductivity and porosity of a saturated porous medium, basing on the knowledge of heating curves at selected points in the medium. In order to solve the inverse problem, we use the least squares criterion (in which the sensitivity coefficients appear), leading to a system of ordinary differential equations (ODE). At the stage of numerical computations, we propose a new global approach, based on the method of lines and ordinary differential equations solvers, combined with a modified Newton method to deal with the nonlinearities presented in the system of coupled equations.

Concerning strong thermal transfer in saturated porous media, the LHA method (Latent Heat Accumulation) is able to take into account phase changes by considering heat accumulation at the local level. The explicit knowledge of the cells which are changing their state allows the build of the liquid-gas interface position. A 2D configuration has been considered, together with a structured mesh but without refinement. The validation of this new method has been checked by making comparison between numerical results and an analytical solution.

### *6.3.2. Granular materials*

**Participant:** Édouard Canot.

This work is done in collaboration with IPR and is published in [11].

We first studied the granular flows by the "discrete elements" method in silo geometries. By changing the micro-mechanical properties of the grains (restitution and friction) we showed that they had a significant influence on the flow discharge. Although models such as "discrete elements" provide access to all the individual properties of the grains, they have one major drawback: the computation time is very important that prohibits the modeling of geophysical and industrial situations. To overcome this problem, we used the "continuous medium" approach, which consider that the granular medium studied follows a rheology recently proposed in the literature. After discussing the numerical implementation, we have studied this rheology for steady and fully developed flows with a semi-analytical method in two configurations: a shear cell and a channel. This allowed us to highlight the differences between a granular medium and a Newtonian fluid.

## 6.4. Models and simulations for flow and transport in porous media

### *6.4.1. Flow and transport in highly heterogeneous porous medium*

**Participants:** Jocelyne Erhel, Grégoire Lecourt, Géraldine Pichot.

This work is done in the context of the H2MNO4 project (see 8.1.1 ), the H2OGUILDE project (see 8.1.4 ), the HEMERA project (see 8.1.2 ). Computations are partly done with GENCI supercomputers (see 8.1.6 ), using the platform H2OLab (see 5.1 ) and the software GWNUM, GWUTIL, PARADIS (see 5.3 , 5.2 , 5.5 ).

This work was done in collaboration with A. Beaudoin, from University of Poitiers (Pprime) and J.-R. de Dreuzy, from Geosciences Rennes (who is on leave until 2013 at UPC, Barcelona, Spain, see 8.2.1 ). It is also done in collaboration with A. Debussche, from ENS-Cachan-Rennes/Ipso Inria team. It was also done in the context of Grégoire Lecourt's internship (M2, INSA Rennes).

It has been presented at a conference (plenary talk) [26] and a paper is submitted to a journal.

Models of hydrogeology must deal with both heterogeneity and lack of data. We consider in this paper a flow and transport model for an inert solute. The conductivity is a random field following a stationary log normal distribution with an exponential or Gaussian covariance function, with a very small correlation length. The quantities of interest studied here are the expectation of the spatial mean velocity, the equivalent permeability and the macro spreading. In particular, the asymptotic behavior of the plume is characterized, leading to large simulation times and in turn to large physical domains. Uncertainty is dealt with a classical Monte Carlo method, which turns out to be very efficient, thanks to the ergodicity of the conductivity field and to the very large domain. These large scale simulations are achieved by means of high performance computing algorithms and tools.

### 6.4.2. *Solving flow equations in highly heterogeneous porous medium*

**Participant:** Thomas Dufaud.

This work was done in collaboration with L. Berenguer and D. Tromeur-Dervout, from University of Lyon (ICJ).

It is published in a journal [12].

This paper is devoted to the acceleration by Aitken's technique of the convergence of the Schwarz domain decomposition method applied to large scale 3D problems with non separable linear operators. These operators come from the discretization of groundwater flow problems modeled by the linear Darcy equation, where the permeability field is highly heterogeneous and randomly generated. To be computationally efficient, a low-rank approximation of the Aitken's formula is computed from the singular value decomposition of successive iterated solutions on subdomains interfaces. Numerical results explore the efficiency of the solver with respect to the random distribution parameters, and specific implementations of the acceleration are compared for large scale 3D problems. These results confirm the numerical behavior of the methodology obtained on 2D Darcy problems [49].

### 6.4.3. *Transport in discontinuous porous medium*

**Participants:** Lionel Lenôtre, Géraldine Pichot.

This work was done in collaboration with A. Lejay, from Inria Nancy, in the context of the H2MNO4 project (8.1.1 ).

It is published in a journal [16].

We propose new Monte Carlo techniques for moving a diffusive particle in a discontinuous media. In this framework, we characterize the stochastic process that governs the positions of the particle. The key tool is the reduction of the process to a Skew Brownian Motion (SBM). In a zone where the coefficients are locally constant on each side of the discontinuity, the new position of the particle after a constant time step is sampled from the exact distribution of the SBM process at the considered time. To do so, we propose two different but equivalent algorithms: a two-steps simulation with a stop at the discontinuity and a one-step direct simulation of the SBM dynamic. Some benchmark tests illustrate their effectiveness.

### 6.4.4. *Adaptive stochastic collocation method for an elliptic problem with random data*

**Participants:** Jocelyne Erhel, Mestapha Oumouni.

This work is done in collaboration with Z. Mghazli, from the university of Kenitra, Morocco, in the context of the joint PhD supervision and the HYDRINV project (see 8.3.8 ,8.3.4 ).

This work has been presented at two conferences [43] [42].

Stochastic collocation methods are frequently used for elliptic equations with random coefficients. However, sparse grid methods are quite expensive and adaptive approaches are designed to save computations.

### 6.4.5. *Reactive transport*

**Participants:** Édouard Canot, Jocelyne Erhel, Souhila Sabit.

This work is done in the context of the MOMAS GNR (8.1.7 ), the contract with Andra (7.1 ) and the C2S@EXA project (see 8.1.3 ). Computations use the software GRT3D (see 5.6 ).

It has been presented at a conference and a workshop [35] [36].

Modeling reactive transport of contaminants in porous media is a complex time-dependent problem, due to combining the difficulties of modeling transport and chemistry, especially the coupling between them. In this work, we are interested to solve this type of coupling. Several methods have been developed for the resolution for solving this type problem. We choose to solve this problem by a global approach, which considers all the equations as a whole system of differential algebraic equations (DAE), which come from the spatial-only discretization of the equations (method of lines). This approach uses implicit schemes, which imply solving many large linear systems with the Jacobian matrix. The differential algebraic system (DAE) is solved by the solver IDA Sundials. Our new technique is implemented in the GRT3D software; we have observed that the CPU time increases very fast with the size of the system. Our aim is thus to reduce this computation time. Profiling tools have shown that an important part of this computation is due to the linear solving related to the Jacobian matrix. We focus our effort on improving this part, by exploiting the 3x3 block-structure of the Jacobian matrix, via a Gaussian block elimination technique. Our simulations are performed on academic test cases, which involve few chemical components (4 to 5) for both 1D and 2D geometries, giving a number of unknowns up to 72000. First results have shown that our technique is very promising, because the CPU time is reduced by approximately 40 After this part, we eliminated the tracer in our test cases. In GRT3D-SL software, we calculated the concentrations directly without using the Logarithms and with this software, we have reduced the CPU time to 50 %.

## 6.5. Models and simulations for flow in porous fractured media

This work is done in collaboration with J.-R. de Dreuzy, from Geosciences Rennes (who is on leave until 2013 at UPC, Barcelona, Spain, see 8.2.1 ). It is done in the context of the GEOFRAC project (see 8.1.5 ), the H2OGUILDE project (see 8.1.4 ), the HEMERA project (see 8.1.2 ), and the Joint Laboratory for Petascale Computing (see 8.3.7 ). Computations are partly done with GENCI supercomputers (see 8.1.6 ), using the platform H2OLab (see 5.1 ) and the software GWNUM, GWUTIL, MPFRAC (see 5.3 , 5.2 , 5.4 ).

### *6.5.1. Influence of fracture scale heterogeneity on the flow properties of three-dimensional Discrete Fracture Networks*

**Participant:** Géraldine Pichot.

This work is published in a journal [21].

While permeability scaling of fractured media has been so far studied independently at the fracture- and network- scales, we propose a numerical analysis of the combined effect of fracture-scale heterogeneities and the network-scale topology. The analysis is based on $2 \times 10^6$ discrete fracture network (DFNs) simulations performed with highly robust numerical methods. Fracture local apertures are distributed according to a truncated Gaussian law, and exhibit self-affine spatial correlations up to a cutoff scale $L_c$. Network structures range widely over sparse and dense systems of short, long or widely-distributed fracture sizes and display a large variety of fracture interconnections, flow bottlenecks and dead-ends. At the fracture scale, accounting for aperture heterogeneities leads to a reduction of the equivalent fracture transmissivity of up to a factor of 6 as compared to the parallel plate of identical mean aperture. At the network scale, a significant coupling is observed in most cases between flow heterogeneities at the fracture and at the network scale. The upscaling from the fracture to the network scale modifies the impact of fracture roughness on the measured permeability. This change can be quantified by the measure $\alpha_2$, which is analogous to the more classical power-averaging exponent used with heterogeneous porous media, and whose magnitude results from the competition of two effects: (i) the permeability is enhanced by the highly transmissive zones within the fractures that can bridge fracture intersections within a fracture plane; (ii) it is reduced by the closed and low transmissive areas that break up connectivity and flow paths.

### *6.5.2. Synthetic benchmark for modeling flow in 3D fractured media*

**Participants:** Jocelyne Erhel, Géraldine Pichot.

This work is published in a journal [22].

Intensity and localization of flows in fractured media have promoted the development of a large range of different modeling approaches including Discrete Fracture Networks, pipe networks and equivalent continuous media. While benchmarked usually within site studies, we propose an alternative numerical benchmark based on highly-resolved Discrete Fracture Networks (DFNs) and on a stochastic approach. Test cases are built on fractures of different lengths, orientations, aspect ratios and hydraulic apertures, issuing the broad ranges of topological structures and hydraulic properties classically observed. We present 18 DFN cases, with 10 random simulations by case. These 180 DFN structures are provided and fully documented. They display a representative variety of the configurations that challenge the numerical methods at the different stages of discretization, mesh generation and system solving. Using a previously assessed mixed hybrid finite element method (Erhel et al., 2009a), we systematically provide reference flow and head solutions. Because CPU and memory requirements stem mainly from system solving, we study direct and iterative sparse linear solvers. We show that the most cpu-time efficient method is a direct multifrontal method for small systems, while conjugate gradient preconditioned by algebraic multrigrid is more relevant at larger sizes. Available results can be used further as references for building up alternative numerical and physical models in both directions of improving accuracy and efficiency.

### *6.5.3. Robust numerical methods for solving flow in stochastic fracture networks*

**Participants:** Jocelyne Erhel, Géraldine Pichot.

This work is published in a journal [20] and was presented at a conference (plenary talk) [33].

Working with random domains requires the development of specific and robust numerical methods to be able to solve physical phenomena whatever the generated geometries. Hydrogeology is a typical area of application where one has to face uncertainty about the geometry and the properties of the domain since the available information on the underground media is local, gathered through in-situ experiments with outcrops and wells. From measurements, statistical laws are derived that allow the generation of natural-like random media. The focus of this talk will concern flow in discrete fracture networks. The parameters governing the fractures lengths, shapes, orientations, positions as well as their hydraulic conductivity are stochastic. Our objective is to design robust numerical methods to solve Poiseuille's flow in large and heterogeneous stochastic fracture networks. The first part deals with the meshing strategies required to obtain a good quality mesh for any generated networks. The second part is devoted to numerical techniques to solve the flow equations. A Mortar-like method to deal with nonconforming meshes at the fracture intersections is presented as well as a Schur complement approach to solve the linear system of interest in parallel.

### *6.5.4. Deflation and Neumann-Neumann Preconditionner for Schur Domain Decomposition Method*

**Participants:** Jocelyne Erhel, Géraldine Pichot.

This work was presented at a conference [34]. A paper is in preparation.

We study a domain decomposition method, which takes advantages from both the direct method and the Preconditioned Conjugate Gradient (PCG). This Schur method reduces the global problem to an interface problem, with a natural domain decomposition based on fractures or fracture packs. We propose an original approach for optimizing the algorithm and a global preconditioning of deflation type. Since the Schur complement $S$ is spd, we apply PCG to solve the linear system $Sx = b$. We use the classical Neumann-Neumann (NN) preconditioner. To gain in efficiency, we use only one Cholesky factorization of the subdomain matrices for the preconditionning and the conjugate gradient steps. We also define a coarse space, based on the subdomain definition, to apply a deflation preconditioner. We do a theoretical complexity study of our algorithm. We use this study, with the numerical data, to compute experimental complexity. We compare the results between several combination for the preconditioner. Then, we confront our results with existing solvers.

### *6.5.5. Flow in complex 3D geological fractured porous media*

**Participants:** Thomas Dufaud, Jocelyne Erhel, Géraldine Pichot.

This work was presented at a conference [24].

This communication focuses on numerical techniques to compute flow in complex 3D geological fractured porous media, where water can flow both in the rock matrix and in the fractures. This study is an extension of the models designed in the teams SAGE and POMDAPI. The numerical model deals with steady-state flow for single phase and incompressible fluid. In the rock matrix, the flow is governed by Darcy's law, while the flow in the fractures is governed by Poiseuille's law. For both, the law of mass conservation is verified. In a first part, we present the model. Then we propose a test case and its discretization considering a Mixed Hybrid Finite Element Method.

# STEEP Exploratory Action

# 6. New Results

## 6.1. Calibration of TRANUS Adjustment Parameters

One of the most difficult steps in calibrating the parameters of the TRANUS land use model, concerns the estimation of its adjustment parameters (so-called shadow prices), that allow to "absorb" imperfections of the model or the data. The main difficulties are the non-linearity of the underlying equations and the fact that some of these equations give rise to loops between intermediate system variables: modifications of some of these variables entail modifications of others and vice-versa. In other words, the concerned part of TRANUS is a dynamic system. Currently, users of TRANUS perform the calibration by semi-automatic (at best) trial-and-error.

We have started investigating more systematic solutions to this. A first step has been to explicitly pose the estimation problem in the form of an optimization problem, with clearly stated cost function and constraints. Next, we have found ways of splitting the problem into separable subproblems, concerning the estimation of adjustment parameters for different economic sectors. In particular, the housing/land sectors can be calibrated independently of the others. A simple gradient descent was shown to be sufficient, both theoretically and experimentally, to achieve this calibration. We are currently investigating strategies to estimate the adjustment parameters of the remaining sectors.

## 6.2. Calibration of TRANUS Using Maximum Likelihood Estimation

Calibration of the TRANUS land use module typically involves determination of key parameters which dictate land use assignments and prices. As mentioned earlier, It is a difficult task to calibrate a LUTI model as the number of parameters involved are large and are uncertain. Traditionally, these models are calibrated manually by experts, who try to estimate the parameters using their prior experience. However, such a method is difficult as well as time consuming, especially when the parameter space is large and uncertain. Hence, an algorithmic procedure to estimate parameters from mathematical model is desired.

We have proposed an algorithm to calibrate the land use module of TRANUS using maximum likelihood estimation (MLE). The observed outputs of the land use module is modeled to follow a Gaussian process. The covariance matrix is represented as a function of inputs of the land use module and hyperparameters. A MLE optimization problem is then formulated to estimate the parameters of the land use module and the hyperparameters of the Gaussian covariance kernel. The resulting nonlinear programming (NLP) problem is then solved using NLP solvers based on sequential quadratic programming.

The proposed calibration algorithm has been successfully applied to the model of Grenoble, France ; and the performance of the proposed calibration methodology, has been compared to traditional calibration techniques. The metric to judge performance is assumed to be the $\mathcal{L}_2$ norm of the difference between observed and calculated land use assignments obtained using the calibrated model.

Before this calibration task is performed, a sensitivity analysis has been carried out. Hence, sensitivity analysis of the parameters on the output is important as is helps us identify major sources of uncertainty in terms of their contribution towards output space variability. Here, the *total effect* of the land use parameters on a *quantity of interest* or *QoI* is assessed. The *QoI* is assumed to be the $\mathcal{L}_2$ norm of the difference between observed and calculated land use assignments. For this Grenoble model, the number of uncertain parameters involved are 100, and finally is is observed that only 3 amongst them contribute towards 99.2% of *QoI* variability. [14], [13]

## 6.3. Material flows, production and consumption at sub-national geographic levels

As explained earlier, estimating the actual environmental impact of an urban area on the one hand, and the efficiency of (local or national) policy options in reducing these impacts on the other, requires an understanding of the material flows and material uses generated by the considered urban area. It is important to realize that impacts (both local and distant) can vary greatly from one region or departement to the next, depending on its agricultural and industrial characteristics. The whole point of this work is to evaluate as best as possible these variations, in order to best adapt public policies in terms of environmental impacts, for given socio-economic conditions and objectives.

The first step in this analysis is to establish a database of production, consumption and exchanges (import and export) at the various geographic levels of interest, and for the various types of material of interest. In practice, the finest scale of available data is a French *département*, and the publicly available data refer to the national, regional or "departemental" level. Only major primary materials are accounted for, through the content of end products and waste in these primary materials (toxic waste are accounted for separately). For example, for cereals such as wheat, production at the departement level is available through the national *Agreste* database, variations of stock are small once averaged over a few years period, import and export are obtained from the *Sitram* database (a database initially elaborated by the ministry of transportation and now maintained by the Ministry of Environment), which follows all national and international transport by transportation mode and by type, through annual stratified polls of transportation companies. Productions of non-agricultural products in France is very low except for construction materials (most notably cement), for which the industry maintains its own publicly available database. Following transformations requires information from various industrial sectors, e.g., the flour trade and food industry for wheat use, taking into account animal farming which consumes a non-negligible fraction of primary agricultural products.

Once this database is constructed, one also needs to estimate production, consumption and imports and exports at finer scale than the departement. In practice, this is performed by correlating the desired information at the national, regional and departemental scale with another auxiliary quantity serving as proxy, that is also known at the desired smaller scale. For example, wheat production can easily be correlated with available surfaces in wheat growing areas, that are known from the Corine Land Cover database at scales of the order of a few hundred meters. More generally, auxiliary quantities are constructed from relevant demographic and economic and geographic data, that are mostly available through the various INSEE databases. This requires some educated guess-work to find the most likely auxiliary quantities, and evaluate their correlation with the quantities of interest at scales where data on both are available. This aspect of the problem has been completed only for food staples at this stage.

An important aspect of the problem is to estimate the errors in the data. Errors can be detected when quantities of a given material are not conserved through transportation and transformation processes. It appears that the largest source of error comes from the transportation database, because the stratified polling methodology is optimized with respect to total transport from a pair of origin and destination, independently of the nature of the transported goods. It is in principle possible to compute confidence intervals per type of material and not only on total volumes of exchanges, but this requires access to some non public information. Discussions have been initiated with the Ministry to have access to this information, in order to estimate the reliability of this method of transport quantification. If the *Sitram* database turns out to be too imprecise, the method described above to estimate lacking data can be applied to transport as well with appropriate auxiliary quantities, but the results also suffer from various sources of error.

This first stage of the Material Flow analysis is nevertheless largely underway. The two next steps consist in environmental impact evaluation on the one hand, at the present date, and in developing a method of analysis of changes of such impacts under various policy scenarios and options. Bith will rely on the use of Life Cycle Analysis databases, as mentioned above.

## 6.4. Computer vision

Three of our permanent staff have previously been active in computer vision. This activity is gradually coming to an end: the last PhD student has defended his thesis in 2012 and no new projects are started. Since this topic is not central to STEEP, results are only summarized very briefly. The main scientific result has been the development of a novel approach for 3D modeling of semi-transparent objects, which couples both, geometric and photometric information [1]. This allows 3D modeling with fewer input images than previously and potentially, with a higher accuracy. Besides this, our main activity in computer vision has been related to industrial projects, the main goal being to finalize our work of the last years with an industrial transfer.

## BANG Project-Team

# 6. New Results

## 6.1. Proliferation dynamics and its control

### 6.1.1. Cell division dynamics in structured cell populations

**Participants:** José Luís Avila Alonso [DISCO project-team, Inria Saclay IdF], Annabelle Ballesta, Frédérique Billy, Frédéric Bonnans [Commands project-team, Inria Saclay IdF], Catherine Bonnet [DISCO project-team, Inria Saclay IdF], Jean Clairambault, Luna Dimitrio, Marie Doumic-Jauffret, Xavier Dupuis [Commands project-team], Olivier Fercoq [MaxPlus project-team, Inria Saclay IdF], Stéphane Gaubert [MaxPlus project-team, Inria Saclay IdF], Germain Gillet [IBCP, Université Cl. Bernard Lyon 1], Philippe Gonzalo [IBCP, Université Cl. Bernard Lyon 1], Pierre Hirsch [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Thomas Lepoutre [now in DRACULA project-team, Inria Rhône-Alpes, Lyon], Jonathan Lopez [IBCP, Université Cl. Bernard Lyon 1], Pierre Magal [University Bordeaux II], Anna Marciniak-Czochra [Institute of Applied Mathematics, Universität Heidelberg], Jean-Pierre Marie [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Roberto Natalini [IAC-CNR, Università Sapienza, Rome], Silviu Niculescu [DISCO project-team, Inria Saclay IdF], Hitay Özbay [Bilkent University, Ankara, Turkey], Benoît Perthame, Ruoping Tang [INSERM Paris (Team18 of UMR 872) Cordeliers Research Centre and St. Antoine Hospital, Paris], Vitaly Volpert [CNRS Lyon, UMR5208, Camille Jordan Institute, Lyon], Jorge Zubelli [IMPA, Rio de Janeiro].

1. *Transition kernels in a McKendrick model of the cell division cycle.* This theme has continued to be developed with identification of model parameters by FUCCI imaging in collaboration with G. van der Horst's team in Amsterdam and with F. Delaunay's team in Nice, within the C5Sys European network, coordinated by F. Lévi (Villejuif) [10], [11], [12], [39], [42], [43]. Main young researchers on this theme, F. Billy has concluded her 2-year Inria postdoc at Bang, leaving for an industrial company in November 2012, and O. Fercoq (team MaxPlus, Saclay) has defended his PhD thesis at École Polytechnique in September 2012, only to leave for a postdoc position dedicated to optimisation theory in Edinburgh.
2. *Modelling haematopoiesis with applications to AML.* This theme has been active through a collaboration with Inria teams Commands (F. Bonnans, X. Dupuis) and Disco (JL Avila, C. Bonnet), and J.-P. Marie's team at St Antoine Hospital leukaemic tumour bank, where A. Ballesta, Cancéropole IdF-Inria postdoc has been detached (ending in March 2013) to identify parameters of a model of acute myeloblastic leukaemia (AML) in patient fresh cell cultures with and without anticancer drugs. This work has led to several presentations, and publications are in preparation.
3. *Hybrid models* Systems combining PDEs and discrete representations in hybrid models, with applications to cancer growth and therapy, in particular for AML, are the object of study of the ANR program *Bimod*, coordinated by V. Volpert (Lyon), associating CNRS (V. Volpert, Lyon), Bordeaux II University (P. Magal) and the Bang project-team.
4. *Molecular model of the activity of the p53 protein.* This work, the object of Luna Dimitrio's PhD thesis [1], co-supervised by J. Clairambault and R. Natalini (Rome), has led to her PhD defence in September 2012 at UPMC, and to a first publication [18], that should be followed by others. After L. Dimitrio's leave for the pharmaceutic industry, a new PhD student, Ján Eliš, has taken over this theme in September 2012 in a new PhD thesis at UPMC, under the supervision of J. Clairambault and B. Perthame

### 6.1.2. Physiological and pharmacological control of cell proliferation

**Participants:** Annabelle Ballesta, Frédérique Billy, Jean Clairambault, Sandrine Dulong [INSERM Villejuif (U 776)], Olivier Fercoq [MaxPlus project-team], Stéphane Gaubert [MaxPlus project-team], Thomas Lepoutre [Dracula project-team], Francis Lévi [INSERM Villejuif (U 776)].

1. *Periodic (circadian) control of cell proliferation in a theoretical model of the McKendrick type.* This theme (cf. supra “transition kernels...”) has been continued [39], [11], [12], [10], [42], [43]. Whereas transition kernels between cell cycle phases without control have been experimentally identified in cell cultures by FUCCI imaging [12], their circadian control remains elusive and has been modelled on the basis of gating by plain cosines representing the influence exerted on these transition kernels by circadian clocks. To go further, it would be necessary to have access by cell imaging to the activity of the best physiological candidates to such gating, namely the cyclin-Cdk complexes, together with the activities of the clock-controlled proteins Wee1 and p21, which thus far have remained unavailable to us through biological experimentation with imaging.
2. *Intracellular pharmacokinetic-pharmacodynamic (PK-PD) models for anticancer drugs.* This theme has continued to be developed with new publications for the drugs irinotecan [40], [44], 5-fluorouracil and oxaliplatin [43].

### 6.1.3. Optimisation of cancer chemotherapy

**Participants:** Annabelle Ballesta, Frédérique Billy, Frédéric Bonnans [Commands project-team], Jean Clairambault, Sandrine Dulong [INSERM Villejuif (U 776)], Xavier Dupuis [Commands project-team], Olivier Fercoq [MaxPlus project-team], Stéphane Gaubert [MaxPlus project-team], Thomas Lepoutre [Dracula project-team], Alexander Lorz, Francis Lévi [INSERM U 776, Villejuif], Michael Hochberg [ISEM, CNRS, Montpellier], Benoît Perthame.

Optimising cancer chemotherapy, in particular chronotherapy, is the final aim of the activities mentioned above. This theoretical activity has been continued, using the McKendrick paradigm in works involving the *C5Sys* network [12], [42], [43], with numerical optimisation algorithms for the toxicity constraint, and also in more general settings taking into account another major issue of anticancer treatment, namely resistance to drugs in cancer cells. To this latter aim, we have developed another type of models based on integro-differential equations, which are inspired from those used in ecology for Darwinian evolution. These are aimed at studying another major issue in cancer therapy: appearance of resistances to treatment in tumour cell populations. Indeed, these cell populations, because of their heterogeneity and genomic instability, present an ability to adapt and evolve (in the Darwinian sense) that is much higher than in healthy cell populations [10], [27], [39]. The time scales under investigation, much shorter than in ecology, are still much longer than in microbiology, and are those of clinical treatments.

From a molecular point of view, studying drug resistance leads to the study of ABC transporters, which is one of the tracks followed by A. Ballesta, following her PhD thesis, in collaboration with F. Lévi’s INSERM team in Villejuif [40], [44].

Underway is also the use of methods of optimal control developed by the Commands project-team (F. Bonnans, X. Dupuis) to optimise therapies in the treatment of Acute Myeloblastic Leukaemia (AML, cf. supra “Modelling haematopoiesis with applications to AML”).

### 6.1.4. Protein polymerisation and application to amyloid diseases (ANR grant TOPPAZ)

**Participants:** Annabelle Ballesta, Vincent Calvez [ENS Lyon], Marie Doumic-Jauffret, Pierre Gabriel, Hadjer Wafaâ Haffaf, Benoît Perthame, Stéphanie Prigent [BPCP, INRA Jouy-en-Josas], Human Rezaei [BPCP, INRA Jouy-en-Josas], Léon Matar Tine [SIMPAF project-team, Inria Lille Nord-Europe].

Published in PLoS One in collaboration with the biologists’ team of H. Rezaei [29], a new and very complete PDE model for protein polymerisation has been designed. Following F. Charles’s work, A. Ballesta has applied this model to Huntington’s disease (PolyQ expansion) and compared it with its ODE counterpart, leading to a better understanding of the leading mechanisms responsible for PolyQ fibrillisation. New applications of this framework model are in progress with H.W. Haffaf and S. Prigent.

The eigenvalue problem playing a major role in the representation of Prion proliferation dynamics and, in a more general way, of many fragmentation-coalescence phenomena, the article [15] published in J. de Math. Pur. Appl. investigated the dependency of the principal eigenvector and eigenvalue upon its parameters. We exhibited possible nonmonotonic dependency on the parameters, conversely to what would have been conjectured on the basis of some simple cases.

### *6.1.5. Inverse problem in growth-fragmentation equations*

**Participants:** Marie Doumic-Jauffret, Marc Hoffmann [ENSAE], Nathalie Krell [Univ. Rennes I], Patricia Reynaud [CNRS, Nice Univ.], Lydia Robert [UPMC], Vincent Rivoirard [Paris IX Univ.], Léon Matar Tine [SIMPAF project-team, Inria Lille Nord-Europe].

In collaboration with statisticians (M. Hoffman, Professor at Université de Marne-la-Vallée, V. Rivoirard, MC at Université d'Orsay, and P. Reynaud, CR CNRS at Université de Nice), in the article [19] published in SIAM Num. Anal., we explored a statistical viewpoint on the cell division problem. In contrast to a deterministic inverse problem approach, we take the perspective of statistical inference. By estimating statistically each term of the eigenvalue problem and by suitably inverting a certain linear operator, we are able to construct an estimator of the division rate that achieves the same optimal error bound as in related deterministic inverse problems. Our procedure relies on kernel methods with automatic bandwidth selection. It is inspired by model selection and recent results of Goldenschluger and Lepski.

An extension of this work, which consists of the statistical estimation of a branching process modelling the same growth and fragmentation dynamics, has been submitted in [49], in collaboration with N. Krell, M. Hoffmann and L. Robert.

In [20], published in J. Math. Biol. with L. Matar Tine, we generalised the inverse techniques proposed previously in [53], [57], in order to adapt them to general fragmentation kernels and growth speeds. The potential applications of this problem are numerous, ranging from polymerisation processes to the cell division cycle. An extension of this work is in progress with M. Escobedo and T. Bourgeron.

## 6.2. Tissue growth, regeneration and cell movements

### *6.2.1. Chemotaxis, self-organisation of cell communities (KPP-Fisher and Keller-Segel)*

**Participants:** Nikolaos Bournaveas [Univ. Edinburgh], Axel Buguin [UPMC, Institut Curie], Vincent Calvez [ENS Lyon], François James [univ. Orléans], Alexander Lorz, Grégoire Nadin [UPMC], Benoît Perthame, Jonathan Saragosti [Institut Curie], Pascal Silberzan [Institut Curie], Min Tang [Shanghai Jiaotong University], Nicolas Vauchelet.

Chemotaxis denotes the ability of some cells to undergo a directed movement in response to an extracellular chemical substance. A mathematical description of chemotaxis is a major issue in order to understand collective movements of bacterial colonies. Numerous mathematical models, at various scales, have been proposed, allowing for a good description of cell aggregation under chemotaxis at the macroscopic level, the first of all being that of Keller-Segel (1971), that is now at the centre of an abundant international scientific literature.

At the cell scale, one uses kinetic equations. Numerical simulations have been performed and blow-up is also observed, which differs highly from pointwise blow-up in parabolic models. Representing them leads to various theoretical questions and amounts to define measure solutions [25], [24] or to develop an existence theory.

### *6.2.2. Single-cell-based and continuum models of avascular tumours*

**Participants:** Ibrahim Cheddadi, Dirk Drasdo, Benoît Perthame, Min Tang [Shanghai Jiaotong University], Nicolas Vauchelet, Irène Vignon-Clémentel [REO project-team].

The recent biomechanical theory of cancer growth considers solid tumours as liquid-like materials comprising elastic components. In this fluid mechanical view, the expansion ability of a solid tumour into a host tissue is mainly driven by either diffusion of cells (emerging on the mesoscopic scale by coarse graining from the cell micro-motility) or by cell division depending either on the local cell density (contact inhibition), on mechanical stress in the tumour, or both. For the two by two degenerate parabolic/elliptic reaction-diffusion system that results from this modelling, we prove there are always travelling waves above a minimal speed and we analyse their shapes. They appear to be complex with composite shapes and discontinuities. Several small parameters allow for analytical solutions; in particular the incompressible cells limit is very singular and related to the Hele-Shaw equation. These singular travelling waves are recovered numerically. See [32].

### 6.2.3. Single cell-based models of tumour growth, tissue regeneration

**Participants:** Gregory Batt [CONTRAINTES project-team], François Bertaux, Géraldine Cellière, Chadha Chettaoui, Ibrahim Cheddadi, Dirk Drasdo, Adrian Friebel, Rolf Gebhardt [Univ. of Leipzig, Germany], Adriano Henney [Director Virtual Liver Network and VLN consortium], Jan G. Hengstler [Leibniz Research Centre, Dortmund, Germany and CANCERSYS consortium], Stefan Höhme, Elmar Heinzle [University of Saarbrücken and NOTOX consortium], Isabelle Hue [INRA], Nick Jagiella, Ursula Klingmüller [German Cancer Centre, Heidelberg and LungSys Consortium], Axel Krinner, Johannes Neitsch, Benoît Perthame, Ignacio Ramis-Conde, Luc Soler [ IRCAD, Coordinator EU-project PASSPORT and PASSPORT consortium], Jens Timmer [University of Leipzig, Germany], Irène Vignon-Clémentel [REO project-team], Juhui Wang [INRA], William Weens.

#### 6.2.3.1. A Multi-scale model for clonal competition in growing tumours

In this work we set up a multi-scale model testing the impact of three experimentally found variants of a signal transduction pathway controlling cell-cell adhesion on multi-cellular growth as well as the possible consequences of inhomogeneous populations where each of the three phenotypes competed [30].

#### 6.2.3.2. Growth of cell populations in embedding granular and cell-like matter

In this work simulations of growing 2D and 3D clones embedded in granular and cell-like matter were mimicked [21]. The influence of active directed cell motion vs. passive pushing triggered by cell proliferation, as well as of various parameters of the embedding matter, such as the friction of embedding objects with its environment, adhesion strength, size of objects, elastic modulus etc. on the growth kinetics and the spatial pattern has been studied. The emerging patterns are strongly reminiscent of a fingering instability (a type of a Saffman-Taylor instability) occuring if a viscous fluid is injected into a more viscous fluid constrained between two plates (Hele Shaw cell).

#### 6.2.3.3. Quantitative modelling of multi-cellular spheroids

Nick Jagiella in his thesis has worked out how stepwise and iteratively mechanisms controlling the spatial-temporal growth dynamics can be inferred by combining information from bright field micrographs stained for proliferating, dying cells, cell nuclei and extra-cellular matrix with the macroscopic growth kinetics.

This thesis, pursued within the German network project LUNGSYS was defended in September 2012. The thesis work was mainly supervised by Dirk Drasdo, PI for this part within the LUNGSYS project. Main collaborators were Margareta Mueller (previously DKFZ, Heidelberg) and Ursula Klingmueller, (DKFZ Heidelberg).

Moreover, Géraldine Cellière has worked out a model to mimic the aggregation of cells in the hanging drop method, a standard method to generate 3D multi-cellular aggregates. The kinetics and final configuration give information on multicellular aggregates. This work is pursued within the EU NOTOX project. Main collaborators are Fozia Noor and Elmar Heinzle (Univ. of Saarbruecken).

#### 6.2.3.4. Image reconstruction of 3D liver architecture at subcellular level

In order to permit simulation liver function we started to set up an image processing pipeline resolving liver at subcellular scale. This will enable us to mimic all flows in liver, which comprises of blood flow through the micro-vessels (sinusoids), of blood plasma through the space between micro-vessel wall and hepatocytes, the main type of liver cells (called space of Disse), and of the bile through a network of bile canaliculi. Besides image analysis, also setting up the models of the flows has been started.

This work is conducted by the PhD student Adrian Friebel (IZBI, University of Leipzig) co-supervised by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network (VLN; PI from IZBI, Leipzig: Dirk Drasdo). Main collaborator is Jan G. Hengstler from the IfADo (directeur at the Leibniz Institute in Dortmund, Germany).

#### *6.2.3.5. Ammonia metabolism during liver regeneration*

Based upon the paper on liver regeneration after drug-induced damage (Hoehme et. al. PNAS 2010 [55]) we in a next step investigated the change of ammonia metabolism during the regeneration process. Ammonia is toxic for the body. We linked our spatial-temporal liver lobule model with a compartment model for the ammonia, glutamine and urea metabolism. In the latter we consider a compartment (the peri-central compartment) in which glutamine synthetase, a strongly ammonia-detoxifying enzyme, is degraded efficiently and a (peri-portal) compartment, in which this is not the case. By testing different hypotheses on the chemical reactions taking place during the degradation process and quantitatively comparing to time-space data of the regeneration process including data on the activity of glutamine synthetase we were able to propose a potentially missing chemical reaction. Validation experiments have been started and suggest that the original reaction scheme was indeed incomplete.

This work is conducted by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) partly within the Germany funded grant project Virtual Liver Network (VLN; PI from IZBI, Leipzig: Dirk Drasdo) and the EU project NOTOX. Main collaborators are Rolf Gebhardt (chair for Biochemistry, University of Leipzig), Jan G. Hengstler from the IfADo (Leibniz Institute in Dortmund, Germany) and BioControl Jena GmbH, a company in Jena, Germany.

#### *6.2.3.6. Multi-scale simulation of cell cycle progression during liver regeneration*

In previous work on liver regeneration after drug induced damage (Hoehme et. al. PNAS 2010 [55]) the experimentally observed spatial-temporal proliferation pattern has been used as an input parameter. We have now started to study the molecular control of cell cycle progression by hepatocyte growth factor (HGF). Based on model predictions with a hypothesized model linking the downstream activation of the HGF-pathway with cell cycle progression, experiments were performed which now led to a validated intracellular model of cell cycle progression by HGF. Moreover, based on model simulations predicting that two sources of HGF are necessary to explain the experimentally observed proliferation pattern, experiments detecting the potential sources of HGF have been initiated. The models are multi-scale in that the precise spatial architecture of a piece of liver tissue is modelled representing each individual hepatocyte as well as the blood micro-vessels. A system of ODE's mimicking the HGF signalling and its impact on cell cycle progression is solved inside each individual cell. The project works out a systematic strategy to stepwise identify multi-scale multi-level processes in tissue organisation extending the lines pursued in Hoehme et. al. [55] and Holzhuetter et. al. [23].

This work is conducted by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network (VLN; PI from IZBI, Leipzig: Dirk Drasdo). Main collaborators are Ursula Klingmueller and Lorenza D'Alessandro (UK is Professor at Heidelberg University and department head at German Cancer Research Centre (DKFZ), Heidelberg, Germany) as well as Jens Timmer and Andreas Raue (JT is Professor University of Freiburg, Germany).

#### *6.2.3.7. Phenotypes in early liver cancer*

The model of a liver lobule, the smallest functional unit of liver (Hoehme et. al., PNAS 2010 [55]) has been used as a starting point to explain the experimentally observed early tumour phenotypes. We made a sensitivity analysis to identify the parameters that influence the tumour phenotype. Each simulation mimicked a monoclonal tumour. We could show that the observed early phenotypes could be explained by only a few sensitive parameters which are the direction of cell division, cell-micro-vessel adhesion, and destruction of micro-vessels by the tumour cells.

This work has been taken over from the previous PhD student William Weens by the PhD student François Bertaux who is co-supervised by Dirk Drasdo and Gregory Batt. Main collaborator is Jan G. Hengstler from the IfADo (directeur at the Leibniz Institute in Dortmund, Germany).

#### *6.2.3.8. Regeneration of liver after partial hepatectomy*

We continued this earlier activity by initiating experiments on pigs to test the model prediction that the 2nd wave of proliferation during regeneration after partial hepatectomy in pig should occur only close to the Glisson capsule, that encloses the liver, while in mouse proliferation occurs homogeneously and isotropically distributed over the whole liver lobe.

This work is conducted by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network. Main collaborators are Jan G. Hengstler from the IfADo (Leibniz Institute in Dortmund, Germany) and Eric Vilbert, Centre Hépato-Biliaire (CHB)- INSERM U785, Hospital Paul Brousse, Villejuif.

#### *6.2.3.9. High resolution model for eukaryotic cells*

In order to permit simulations directly out of 3D reconstructions of confocal laser scanning micrographs at subcellular resolution we developed a model that is capable to resolve complex cell shapes. The model parameters were calibrated by comparison with experiments probing the material properties of cells. Moreover, the cell division was implemented. The model was integrated into the CellSys software.

This work is conducted by the PhD student Johannes Neitsch (IZBI, University of Leipzig) co-supervised by Dirk Drasdo and Stefan Hoehme (IZBI, University of Leipzig) within the Germany funded grant project Virtual Liver Network (VLN). Main collaborators are Jan G. Hengstler from the IfADo (directeur at the Leibniz Institute in Dortmund, Germany) and Josef Kaes (Prof. for Experimental Physics, Univ. Leipzig).

#### *6.2.3.10. Yeast cells playing the Game of Life*

Within a collaboration with a synthetic biology lab at MIT, we work on the multicellular modelling of engineered yeast cell populations. Those cells secrete a messenger molecule (IP) which diffuse in the medium, bind to other cells, and trigger a signalling cascade which finally induce expression of lethal genes. A model has been established based on our single-cell-based model framework associated with PDE's simulations, and it is currently used to explain and guide experiments obtained at MIT.

This work is conducted within the project Sine2Arti by François Bertaux co-supervised by Gregory Batt and Dirk Drasdo, and by Szymon Stoma. Main collaborator is Ron Weiss, MIT, Boston, USA.

#### *6.2.3.11. Stochastic modelling of extrinsic apoptosis*

Here we extended a well-established ODE model of TRAIL-induced apoptosis developed by Sorger's group in Harvard by the possible effect of cell-to-cell variability due to stochasticity of rare events in the cascade.

This work is conducted within the project Sine2Arti by François Bertaux co-supervised by Gregory Batt and Dirk Drasdo, and by Szymon Stoma as well as Xavier Duportet for the experimental part.

#### *6.2.3.12. Artificial Homeostasis in HeLa cells*

The aim is to genetically engineer human cancer cells (HeLa cell line) such that they perform population control in a petri dish. To do so, it is made use of extrinsic apoptosis by forcing cells to produce a messenger molecule able to trigger apoptosis above a certain threshold concentrations in the medium. We developed a mathematical model which integrates both PDEs and intracellular components into a single-cell-based model framework. Such model allows to help designing the genetic system that should be integrated into cells as well as guiding experiments.

This work is conducted within the project Sine2Arti by François Bertaux who is co-supervised by Gregory Batt and Dirk Drasdo. Moreover Szymon Stoma for the modelling part, as well as Xavier Duportet for the experimental part from the CONTRAINTES team are included.

### *6.2.4. Modelling flow in tissues*

**Participants:** Lutz Brusch [TU Dresden], Dirk Drasdo, Adrian Friebel [IZBI, University of Leipzig], Stefan Hoehme [IZBI, University of Leipzig], Nick Jagiella [Inria and IZBI, University of Leipzig], Hans-Ulrich Kauczor [University of Heidelberg, Germany], Fabian Kiessling [University Clinics, Technical University of Aachen, Germany], Ursula Klingmueller [German Cancer Research Centre (DKFZ), Heidelberg, Germany], Hendrik Laue [Fraunhofer Mevis, Bremen, Germany], Ivo Sbazarini [MPI for Molecular Cell Biology and Genetics, Dresden, Germany], Irène Vignon-Clémentel [REO project-team], Marino Zerial [MPI for Molecular Cell Biology and Genetics, Dresden, Germany].

#### *6.2.4.1. Flow and perfusion scenarios in cancer*

In this subject we simulated typical flow and perfusion scenarios in tumour and tissue including, how the spatial-temporal pattern look like on the scale of non-invasive medical image modalities currently applied, to infer parameters that are used to or may permit to evaluate the perfusion of tumors in patients. The simulations use Poiseuille flow and Kirchhoff rule in 3D blood network representing typical architectures.

The work was part of the PhD thesis of Nick Jagiella, defended in September 2012 co-supervised by Dirk Drasdo and Irene Vignon-Clementel, and conducted within the grant funded network projects LUNGSYS and LUNGSYS II. Main collaborators were Oliver Sedlaczek, DKFZ Heidelberg and University of Heidelberg, Fabian Kissling, Technical University of Aachen and Hendrik Laue, Fraunhofer Mevis, Bremen (all in Germany).

#### *6.2.4.2. Flow in liver lobules*

The aim of this project is to simulate realistically the flow of matter within liver lobules from images generated with different image modalities at histological scales. So far we have established a model of blood flow and perfusion in liver lobules based upon 3D reconstruction of confocal micrographs.

This work is conducted by collaboration of different groups within the Germany funded grant project Virtual Liver Network. From our group Nick Jagiella, Adrian Friebel, and Stefan Hoehme, Dirk Drasdo are involved, main collaborators are Irene Vignon-Clementel (REO project team Inria), Marino Zerial and Ivo Sbazarini (Max-Planck Institute for Molecular Cell Biology and Genetics, Dresden, Germany), Lutz Brusch (Technical University of Dresden) and Jan G. Hengstler from the IfADo (Leibniz Institute in Dortmund, Germany).

### *6.2.5. Contraction of acto-myosin structures in morphogenesis and tissue repair*

**Participants:** Luís Almeida, P. Bagnerini [Univ. Genova], A. Habbal [Univ. Nice], A. Jacinto [CEDOC, Lisbon], M. Novaga [Univ. Padova], A. Chambolle [École Polytechnique], J. Demongeot [Univ. Grenoble].

Contraction of actin structures (in one, two or three dimensions) plays an important role in many cellular and tissue movements, both at a multicellular tissue level and at a cellular (and even intracellular) one: from muscle contraction to neural tube closure, epiboly in zebrafish embryo, the contractile ring in cytokinesis, cell crawling,... examples are everywhere in the living world. These structures consist of meshworks of actin filaments (which are like fibers) that are cross-linked by molecular motors (Myosin II) which can make the actin filaments slide relative to each other, thus generating deformation movements.

In [4] we are particularly interested in modelling the contraction of acto-myosin cables in morphogenesis and tissue repair. The experiments done in collaboration with A. Jacinto's lab show that the local curvature (and in particular its sign) plays an important role in the contractile behaviour of the acto-myosin cables. These experimental results led us to develop some of these ideas in [6] and to do a more abstract study of flows by the positive part of the curvature in [5].

## 6.3. Neurosciences

**Participants:** M. Galtier, G. Hermann, M. Magnasco, T. Taillefumier, Jonathan Touboul.

We pursued the analysis of the dynamics of networks of neurons in the presence of noise. Limit theorems in simple cases were treated in [9], and more refined models including space, delays and heterogeneities were analysed in [34], [35], toubouldelays:12,touboulNeuralFieldsDynamics:12. In all these contributions we analysed the eminently important role of noise and heterogeneity on the qualitative dynamics of networks. Mathematical results were obtained for representation of the solutions to linear functional differential equations [22] that were motivated by plasticity phenomena in the cortex.

## 6.4. Free surface geophysical flows

**Participants:** Emmanuel Audusse [LAGA - Université Paris 13, Institut Galilée], Anne-Céline Boulanger, Marie-Odile Bristeau, Benoît Perthame, Jacques Sainte-Marie, Nicolas Seguin, Edwige Godlewski, Anne Mangeney, Yohan Penel, Raouf Hamouda, Philippe Ung.

The ANGE team has been created in november 2012. This new team (led by J. Sainte-Marie) resumes the activities of the BANG team concerning geophysical flows.

We are involved in research concerning the numerical simulation of free surface geophysical flows such as rivers, lakes, coastal areas and also overland flows. Many applications related to environmental problems are concerned : floodings, dam breaks, swell, transport and diffusion of pollutants, water quality, upwellings, sustainability of aquatic ecosystems, ...

The basic model for these problems is the 3D free surface Stokes system leading to a 3D solver [52] with a moving mesh. However for efficiency reasons, vertically averaged models such as the Saint-Venant system [54] are often used.

The Saint-Venant equations are deduced of the Navier-Stokes system with two main assumptions:

- the pressure is hydrostatic,
- the horizontal velocity is represented by its average.

We have developed extensions of the Saint-Venant system where the basic Saint-Venant solver [51] is still used and, in that way, the robustness, the efficiency and the easiness to treat the free surface are preserved while the domain of validity is larger.

In these extensions, we relax the two above assumptions. Actually, we have derived a non-hydrostatic shallow water model and a multilayer Saint-Venant system.

We have coupled the hydrodynamics of free surface flows with other phenomena such as biology (phytoplankton culture) or erosion.

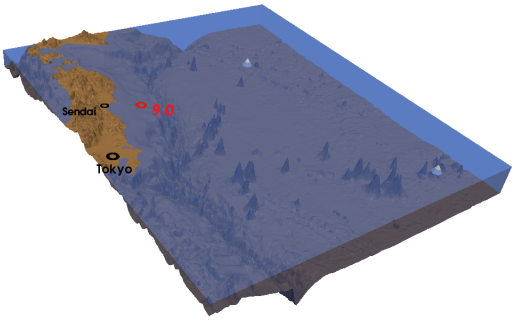


*Figure 1. Map of Japan with the seism epicentre and the DART buoys 21418 and 21413.*

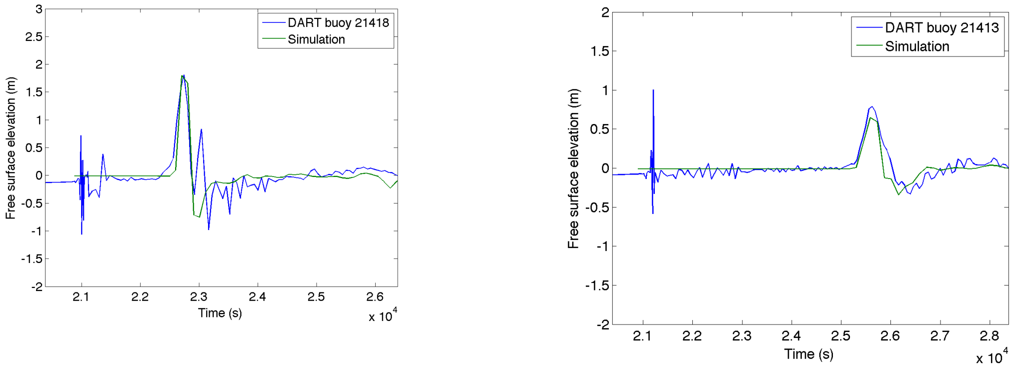


*Figure 2. Free surface elevation of the sea, comparison between the recorded data by the buoys 21418 and 21413 and the simulation obtained with our 3d Navier-Stokes code.*

## BIGS Project-Team

# 6. New Results

## 6.1. Modern methods of data analysis

Participants: R. Bar, B. Lalloué, J-M. Monnez, C. Padilla, D. Zmirou, S. Deguen.

In 2012, our contributions to data analysis in a Biological context are twofold:

- At a theoretical level, we have kept on working on the so-called online data analysis alluded to at the *Scientific Foundations* Section. Specifically we have carried on in [15] (see also [4]) the analysis of data whose characteristics such as mathematical expectation or covariance matrix may vary with time, a problem which arises very naturally in this context. Moreover, in order to save computation time and thus take into account more data, a method considering several data at each step (we talk about data blocks) is proposed. This technique can also be useful if data are sent and received blockwise. In parallel, a R package performing most of the methods of factorial analysis in an online way is under development.
- At a practical level, our efforts have focused (cf. [19]) on an interesting study concerning the construction of a socio-economic neighborhood index which might quantify health inequalities. While several socio-economic indices already exist in this application field, most of them are very simple both in term of methodological construction and of number of variables taken into account, and only a few use data mining techniques. In order to exploit the large data sets of socio-economic variables provided by censuses and create neighborhood socio-economic indices yielding a better highlight of social health inequalities, a procedure was set in order to automatically select the best indicators in a set of socio-economic variables and synthesize them in a quantitative index. Application to three French metropolitan areas allowed testing the procedure and confirming both its reproducibility on various urban areas and the quality of the neighborhood socio-economic indices we had created (according to field experts and study partners). In this context, our expertise in data analysis allows for a good prediction by means of rigorous methods. Eventually, in order to simplify the application of the creation procedure of a socio-economic index for non-statisticians, a R package called SesIndexCreatoR was created to implement it.
- Publication of the sharp results obtained in [8] on local regression techniques.

## 6.2. Tumor growth modeling

Participants: R. Keinj, T. Bastogne, P. Vallois.

Up to now, the treatment planning systems used in radiotherapy only use mathematical models to describe the delivery of physical doses of radiation within biological tissues but cannot accurately predict the biological damages caused by such treatments. One important bottleneck is to account for the cell damage heterogeneity in the treated tumor. To this aim we firstly introduced in [51] a stochastic model based on multi-state Markov chains able to describe both treatment damage and cell reparation process.

More recently, we have proposed another model describing the lifespan of heterogenous tumors treated by radiotherapy. It is a bi-scale model in which the cell and tumor lifespans are represented by random variables. First and second-order moments, as well as the cumulative distribution functions and confidence intervals are expressed for the two lifespans with respect to the model parameters. One interesting result is that the mean value of the tumor lifespan can be approached by a logarithmic function of the initial cancer cell number. Moreover, we show that TCP (Tumor Control Probability) and NTCP (Normal Tissue Complication Probability), used in radiotherapy to evaluate, optimize and compare treatment plans, can be derived from the tumor lifespan and the surrounding healthy tissue respectively. Finally, we propose a ROC curve, entitled ECT (Efficiency-Complication Trade-off), suited to the selection by clinicians of the appropriate treatment planning (see [10]).

One difference between photodynamic therapy (PDT) and radiotherapy (RT) is the irradiation signal (X ray in RT and light beam in PDT). Another one is the treatment planning: 10 to 30 daily sessions of treatment in RT against only one for PDT. To adapt the previous model to PDT, a continuous-time version was developed and proposed in [18]. The model has been implemented into Matlab and numerical simulations have emphasized the effects of the model parameters on the model output.

In the framework of a new collaboration with S. Niclou (NorLux Neuro-Oncology Laboratory, Department of Oncology, Centre de Recherche Public de la Santé, Luxembourg), we have extended our stochastic model of cell damage to describe the phenotypic heterogeneity in brain tumors. Preliminary results have recently been presented in [16]. Cancer stem cell (CSC) hypothesis suggests that tumor progression and recurrence rely on a small subpopulation of cancer cells with stem-like properties. The unresolved question is whether cancer stem cells lead to organisation of intratumoral phenotypic heterogeneity by hierarchical differentiation events or whether they represent one of the transitory phenotypic states. This is crucial not only for our understanding of tumor progression, but also for the successful design of novel therapeutic strategies targeting CSCs. Let us also highlight the fact that those studies are related to a more application oriented research synthesized in [3], [13], [21]

## 6.3. Piecewise deterministic Markov processes

Participants: A. Crudu, A. Debussche, A. Muller-Gueudin, O. Radulescu.

Piecewise deterministic Markov processes are models which feature in a prominent way in Biomedical applications. They appear in two contributions of our team this year.

**(1)***Convergence of stochastic gene networks.* In [24], [5], we propose simplified models for the stochastic dynamics of gene network models arising in molecular biology. Those gene networks are classically modeled by Markov jump processes, which are extremely time consuming. To overcome this drawback, we study the asymptotic behavior of multiscale stochastic gene networks using weak limits of Markov jump processes.

We consider a set of chemical reactions $R_r$, $r \in \mathcal{R}$; $\mathcal{R}$ is supposed to be finite. These reactions involve species indexed by a set $S = 1, \cdots, M$, the number of molecules of the species $i$ is denoted by $n_i$ and $X \in \mathbb{N}^M$ is the vector consisting of the $n_i$'s. Each reaction $R_r$ has a rate $\lambda_r(X)$ which depends on the state of the system, described by $X$ and corresponds to a change $X \to X + \gamma_r$, $\gamma_r \in \mathbb{Z}^M$.

Mathematically, this evolution can be described by the following Markov jump process. It is based on a sequence $(\tau_k)_{k\geq 1}$ of random waiting times with exponential distribution. Setting $T_0 = 0$, $T_i = \tau_1 + \cdots + \tau_i$, $X$ is constant on $[T_{i-1}, T_i)$ and has a jump at $T_i$. The parameter of $\tau_i$ is given by $\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))$:

$$\mathbf{P}(\tau_i > t) = \exp\left(-\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))t\right).$$

At time $T_i$, a reaction $r \in \mathcal{R}$ is chosen with probability $\lambda_r(X(T_{i-1}))/\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))$ and the state changes according to $X \to X + \gamma_r$: $X(T_i) = X(T_{i-1}) + \gamma_r$. This Markov process has the following generator:

$$Af(X) = \sum_{r\in\mathcal{R}} \left[f(X + \gamma_r) - f(X)\right] \lambda_r(X).$$

In the applications we have in mind, the numbers of molecules have different scales. Some of the molecules are in small numbers and some are in large numbers. Accordingly, we split the set of species into two sets $C$ and $D$ with cardinals $M_C$ and $M_D$. This induces the decomposition $X = (X_C, X_D)$, $\gamma_r = (\gamma_r^C, \gamma_r^D)$. For $i \in D$, $n_i$ is of order 1 while for $i \in C$, $n_i$ is proportional to $N$ where $N$ is a large number. For $i \in C$, setting $\widetilde{n}_i = n_i/N$, $\widetilde{n}_i$ is of order 1. We define $x_C = X_C/N$ and $x = (x_C, X_D)$.

For this kind of system, we are able to give in [5] some relevant information on the asymptotic regime $N \to \infty$ when different type of reactions are involved. Depending on the time and concentration scales of the system we distinguish four types of limits:

- Continuous piecewise deterministic processes (PDP) with switching.
- PDP with jumps in the continuous variables.
- Averaged PDP.
- PDP with singular switching.

We justify rigorously the convergence for the four types of limits.

**(2)***Variable length Markov chains.* A classical random walk $(S_n,\, n \in \mathbb{N})$ is defined by $S_n := \sum_{k=0}^{n} X_k$, where $(X_k)$ are i.i.d. When the increments $(X_k)_{k\in\mathbb{N}}$ are a one-order Markov chain, a short memory is introduced in the dynamics of $(S_n)$. This so-called "persistent" random walk is no longer Markovian and, under suitable conditions, the rescaled process converges towards the integrated telegraph noise (ITN) as the time-scale and space-scale parameters tend to zero (see [70], [71], [50]). The ITN process is effectively non-Markovian too. In [28] our aim has been to consider persistent random walks $(S_t)$ whose increments are Markov chains with variable order which can be infinite.

Associated with a process $(X_n)$ which takes its values in a finite set, we consider an integer valued process $(M_n)$ so that $(X_n, M_n)$ is Markov and $M_n$ measures the size of the memory at time $n$. This variable memory is justified by a one-to-one correspondence between $(X_n)$ and a suitable Variable Length Markov Chain (VLMC), since for a VLMC the dependency from the past can be unbounded. We prove in [28] that, under a suitable rescaling, $(S_n, X_n, M_n)$ converges in distribution towards a time continuous process $(S^0(t), X(t), M(t))$. The process $(S^0(t))$ is a semi-Markov and Piecewise Deterministic Markov Process whose paths are piecewise linear.

Observe that, though our study in [28] is made at a theoretical level, it leads to potentially interesting applications in growth models for tumors. This kind of link will be developed in the next future.

## 6.4. Inference for Gaussian systems

Participants: T. Cass, S. Cohen, M. Hairer, C. Litterer, F. Panloup, L. Quer, S. Tindel.

As mentioned at the *Scientific Foundations* Section, 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 (see [69]) is of the form:

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

where $\theta$ 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 (see the *Application Domains* Section) require the analysis of the following $\mathbb{R}^n$-valued equation:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\, du + \int_0^t \sigma(\theta; X_u)\, dB_t, \tag{3}$$

where $\theta$ enters non linearly in the coefficient, where $\sigma$ 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 2012:

- An implementable numerical scheme for equations driven by irregular processes, which is one of the ingredients one needs in order to perform an accurate statistical estimation procedure (see [6]).
- A better understanding of the law of the solution $X_t^\theta$ to equation (1 ), carried out in [25]. This step allows to obtain smoothness of density for our equation of interest in a wide range of contexts, which is an essential prerequisite for a good estimation procedure.
- Another important preliminary step for likelihood estimates for stochastic equations is a good knowledge of their invariant measure in the ergodic case. This is the object of our article [27].
- Finally we have also progressed in our knowledge of noisy differential systems by extending the range of applications of rough paths methods [11], [14].

# BIOCORE Project-Team

# 6. New Results

## 6.1. Mathematical methods and methodological approach to biology

### 6.1.1. Mathematical analysis of biological models

**Participants:** Jean-Luc Gouzé, Olivier Bernard, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Francis Mairet, Rafael Muñoz-Tamayo, Elsa Rousseau.

#### 6.1.1.1. Mathematical study of semi-discrete models

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [95]. We used such models and analysed their properties in several situations that are developed in 6.2.3 , most of them requiring such a modeling in order to take seasonality into account. Such is the case when 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 [59] or in the co-existence analysis of epidemiological strains [19], [50]. Seasonality also plays a big role in the semi-discrete modeling required for the analysis of consumers’ adaptive behavior in seasonal consumer-resource dynamics, where only dormant offspring survives the ’winter’ [52].

#### 6.1.1.2. Mathematical study of models of competing species

When several species are in competition for a single substrate in a chemostat, and when the growth rates of the different species only depend on the substrate, it is known that the generic equilibrium state for a given dilution rate consists in the survival of only one of the species. In [30], we propose a model of competition of $n$ species in a chemostat, where we add constant inputs of some species. We achieve a thorough study of all the situations that can arise when having an arbitrary number of species in the chemostat inputs; this always results in a Globally Asymptotically Stable equilibrium where all input species are present with at most one of the other species.

The competition of several microalgal species was also studied in order to determine conditions that may give a competitive advantage to a species of interest. We study the competition for two species subject to photoinhibition at high light. This leads to a closed loop control strategy based on the regulation of the light intensity at the bottom of the reactor. The winning species is the one with the highest growth rate at high light. Then we show that the proposed controller allows the selection of a species of interest among n species [102].

### 6.1.2. Model design, identification and validation

**Participants:** Olivier Bernard, Francis Mairet.

One of the main families of biological systems that we have studied involves mass transfer between compartments, whether these compartments are microorganisms or chemical species in a bioreactor, or species populations in an ecosystem. We have developed methods to estimate the models of such systems [2]. These systems can be represented by models having the general structure popularized by [69], [74], and based on an underlying reaction network:

$$\frac{d\xi}{dt} = K\,r(\xi,\psi) + D(\xi_{in} - \xi) - Q(\xi)$$

We address two problems: the determination of the pseudo-stoichiometric matrix $K$ and the modeling of the reaction rates $r(\xi,\psi)$.

In order to identify $K$, a two-step procedure has been proposed. The first step is the identification of the minimum number of reactions to be taken into account to explain a set of data. If additional information on the process structure is available, we showed how to apply the second step: the estimation of the pseudo-stoichiometric coefficients.

This approach has been applied to various bioproduction processes, among which activated sludge processes [68], anaerobic digestion [87], [114] and anaerobic digestion of microalgae [20]. Recently it was also used to reduce the ADM1 model in the case of winery effluent wastewater [88].

### 6.1.3. Nonlinear observers

**Participants:** Jean-Luc Gouzé, Olivier Bernard, Francis Mairet.

*Interval observers*

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [7]. We have extended the interval observer design to new classes of systems. First, we designed interval observers, even when it was not possible in the original basis, by introducing a linear, time-varying change of coordinates [105]. This approach was then extended to $n$-dimensional linear systems, leading to the design of interval observers in high dimensions [106]. Interval observers for non linear triangular systems satisfying Input to State Stability has been proposed [22]. Extension to time-delay systems have also been proposed [23]. The efficiency of the interval observer design, even with chaotic systems has been developed and applied considering parameters uncertainties of the system and biased output [108], [105].

The combination of the observers has also been improved in the case where various types of interval observers are run in parallel in a so-called "bundle of observers" [73]. These algorithms have been improved by the estimation of the observer gain providing the best estimate [40], [21]. The approach has been applied to estimation of the microalgae growth and lipid production [101].

These works are done in collaboration with Frédéric Mazenc (DISCO, Inria) and Marcelo Moisan (EMEL S.A., Chile).

### 6.1.4. Metabolic and genomic models

**Participants:** Jean-Luc Gouzé, Madalena Chaves, Alfonso Carta, Ismail Belgacem, Xiao Dong Li, Olivier Bernard, Wassim Abou-Jaoudé, Luis Casaccia, Caroline Baroukh, Rafael Muñoz-Tamayo, Jean-Philippe Steyer.

*Multistability and oscillations in genetic control of metabolism*

Genetic feedback is one of the mechanisms that enables metabolic adaptations to environmental changes. The stable equilibria of these feedback circuits determine the observable metabolic phenotypes. Together with D. Oyarzun from Imperial College, we considered an unbranched metabolic network with one metabolite acting as a global regulator of enzyme expression. Under switch-like regulation and exploiting the time scale separation between metabolic and genetic dynamics, we developed geometric criteria to characterize the equilibria of a given network. These results can be used to detect mono- and bistability in terms of the gene regulation parameters for any combination of activation and repression loops. We also find that metabolic oscillations can emerge in the case of operon-controlled networks; further analysis reveals how nutrient-induced bistability and oscillations can emerge as a consequence of the transcriptional feedback [27].

*Global stability for metabolic models and unreduced Michaelis-Menten equations*

We are interested in the uniqueness and stability of the equilibrium of reversible metabolic models. For biologists, it seems clear that realistic metabolic systems have a single stable equilibrium. However, it is known that some types of metabolic models can have no or multiple equilibria. We have made some contribution to this problem, in the case of a totally reversible enzymatic system. We prove that the equilibrium is globally asymptotically stable if it exists; we give conditions for existence and behavior in a more general genetic-metabolic loop [26]. Moreover, with the same techniques, we studied full (i.e. not reduced by any time-scale argument) Michelis-Menten reactions or chains of reactions: we prove global stability when the equilibrium exists, and show that it may not exist. This fact has important consequences for reduction of metabolic systems in a coupled genetic-metabolic system [34], [45], [70].

*Interconnections of Boolean modules: asymptotic and transient behavior*

A biological network can be schematically described as an input/output Boolean module: that is, both the states, the outputs, and the inputs are Boolean. The dynamics of a Boolean network can be represented by an asynchronous transition graph, whose attractors describe the system's asymptotic behavior. We have shown that the attractors of the feedback interconnection of two Boolean modules can be fully identified in terms of cross-products of the semi-attractors (states of the attractor with same output) of each module. In [82], the *asymptotic graph* was proposed, which is quite fast to compute and identifies all attractors of the interconnected system, but may also generate some spurious attractors. In [31] the *cross graph* is proposed, which exactly identifies the attractors of the interconnected system but is slower to compute. The asymptotic dynamics of high-dimensional biological networks can thus be predicted through the computation of the dynamics of two isolated smaller subnetworks. An application is, for instance, to interconnect four individual "cells" to obtain all the attractors of the segment polarity genes model in *Drosophila*.

*Probabilistic approach for predicting periodic orbits in piecewise affine differential models*

The state space of a piecewise affine system is partitioned into hyperrectangles which can be represented as nodes in a directed graph, so that the system's trajectories follow a path in a transition graph. Using this property we defined a *transition probability* between two nodes $A$ and $B$ of the graph, based on the volume of the initial conditions on the hyperrectangle $A$ whose trajectories cross to $B$ [15]. The parameters of the system can thus be compared to the observed or experimental transitions between two hyperrectangles. This definition is useful to identify sets of parameters for which the system yields a desired periodic orbit with a high probability, or to predict the most likely periodic orbit given a set of parameters, as illustrated by a gene regulatory system composed of two intertwined negative loops.

*Structure estimation for unate Boolean models of gene regulation networks*

Estimation or identification of the network of interactions among a group of genes is a recurrent problem in the biological sciences. Together with collaborators from the University of Stuttgart, we have worked on the reconstruction of the interaction structure of a gene regulation network from qualitative data in a Boolean framework. The idea is to restrict the search space to the class of unate functions. Using sign-representations, the problem of exploring this reduced search space is transformed into a convex feasibility problem. The sign-representation furthermore allows to incorporate robustness considerations and gives rise to a new measure which can be used to further reduce the uncertainties. The proposed methodology is demonstrated with a Boolean apoptosis signaling model [35].

*E. coli modeling and control*

In the framework of ANR project Gemco, we developed and analyzed a model of a minimal synthetic gene circuit, that describes part of the gene expression machinery in *Escherichia coli*, and enables the control of the growth rate of the cells during the exponential phase.

This model is a piecewise non-linear system with two variables (the concentrations of two gene products) and an input (an inducer). We studied the qualitative dynamics of the model and the bifurcation diagram with respect to the input. Moreover, an analytic expression of the growth rate during the exponential phase as function of the input has been derived. A relevant problem was that of parameters identifiability of this expression supposing noisy measurements of exponential growth rate. We presented such an identifiability study that we validated in silico with synthetic measurements [36].

We also studied a model of the global cellular machinery designed by D. Ropers and collaborators (IBIS team, Grenoble). This model has 11 variables and many parameters ; we explored different techniques for reduction and simplification [56], [57].

*Transition graph and dynamical behavior of piecewise affine systems*

We investigated the links between the topology of the transition graph and the number and stability of limit cycles in a class of two-dimensional piecewise affine biological models. To derive these structure-to-dynamics principles, we use the properties of continuity, monotonicity and concavity of Poincare maps associated with transition cycles of the transition graph [64].

*Robust estimation for a hybrid model of genetic networks*

State estimation problems with Boolean measurements for a classical negative loop genetic network governed by a piecewise affine (PWA) model have been studied in [39]. Observers are proposed for the cases where either full state or only partial state Boolean measurements are available. In the first case, sliding modes may occur, which leads to finite time convergence for the observer. In the second case, an algebraic computation is proposed to solve the initial condition inverse problem. The robustness of the observer for a parametric uncertain model is investigated, and we show that the error bound is proportional to the magnitude of the uncertainty.

*Modeling the metabolic network in non balanced growth conditions*

We have developed a new approach to represent the metabolic network of organisms for which the hypothesis of balanced growth is not satisfied [67]. This is especially true for microalgae which store carbon during the day and nitrogen during the night [44]. The proposed formalism is based on the assumption that some parts of the metabolic network satisfy the balance growth conditions, *i.e.* there is no accumulation of intermediate compounds. This hypothesis specifically applies to the main functions in the cell (respiration, photophosphorylation,...). Between two functions, some compounds can accumulate with storage/reuse kinetics. The resulting system is thus a slow-fast system.

## 6.2. Fields of application

### *6.2.1. Bioenergy*

#### *6.2.1.1. Modeling of microalgae production*

**Participants:** Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Rafael Muñoz-Tamayo, Ghjuvan Grimaud, Charlotte Combe, 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 [18], [17], [48] and support model development. These experiments have been carried out in the Lagrangian simulator (SEMPO), 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 observed.

Other experiments were carried out to reproduce the light percept by a cell in a raceway [58]. An electronic platform was developed to reproduce the flashing light which, from the hydrodynamical studies, is likely to happen in a raceway at the cell scale. The experiments show that the microalgae adapt their pigments to the average light that they have received.

The effect in the cell cycle of both the light periodic signal and a nitrogen limitation were studied. The strong interactions of the interactions between the different phases of the cell cycle through checkpoints was highlighted [24].

This work is done in collaboration with Amélie Talec, Thomas Lacour, and Christophe Mocquet (CNRS-Océanographic Laboratory of Villefranche-sur-Mer).

*Modeling the effect of temperature*

The effect of temperature on microalgae has been represented by adapting the CTMI model developed for bacteria [115]. The proposed model [14], [28], was able to correctly represent the growth response to temperature for 15 different species. A procedure for model calibration and estimation of the parameter uncertainties was specially developed, allowing to gather experimental data from various sources. It was shown that different strains of the same species have a very similar response to temperature fluctuations. Moreover, for low light intensities, a simple model can represent both effects of light and temperature [14].

*Modeling light distribution within a photobioreactor*

The light distribution within a photobioreactor was estimated thanks to a multi photon Monte-Carlo simulation. From measurements of absorption and scattering properties, it was thus possible to extrapolate and validate the light distribution within a photobioreactor or a raceway.

*Modeling lipid accumulation*

We have proposed a new model for lipid production by microalgae which describes the fate of the $CO_2$ incorporated during photosynthesis [10]. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. It has been calibrated and validated with experimental data. This model highlights and explains the phenomenon of hysteresis in lipid production which has been experimentally verified. It has been extended to account for light/dark cycles [96].

*Modeling a microalgae production process*

The integration of different models developed in the group [72], [96], [10],[14] 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 in a photobioreactor under light and nitrogen limitations [72]. 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 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 [109], [110]. This model is used to predict lipid production in the perspective of large scale biofuel production.

#### 6.2.1.2. Coupling growth of microalgae with hydrodynamics

**Participants:** Olivier Bernard, Antoine Sciandra, Philipp Hartmann, Charlotte Combe.

*Modeling 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 [13]. This model is supported by the work of ANGE aiming at reproducing the hydrodynamics of the raceway, with a specific attention to the effect of the paddle wheel on the fluid.

*Modeling the photosynthesis response to fast fluctuating light*

The impact of the hydrodynamics on the light percept by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [37] [92]. 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 [93].

#### 6.2.1.3. Optimization of microalgae production

**Participants:** Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Rafael Muñoz-Tamayo.

*Numerical optimization*

Using the detailed model for raceway systems, we assessed strategies for optimal operation in continuous mode [109]. Two strategies were developed. The first one resides in solving 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. In the second strategy, we aimed at translating the optimization problem into a regulation problem. We proposed a simple operational criterion that when integrated in a strategy of closed-loop control allows to reach biomass productivities very near to the productivities obtained with the optimal control. We demonstrated that the practical advantages for real implementation makes our proposed controller a suitable control strategy for optimizing microalgae production in raceways.

*Analytical optimization*

Optimization strategies were based on simple microalgae models : first, biomass production has been optimized in a constant light environment [104], yielding results emphasizing the importance of the optical depth of the reactor. In a second work, we focused on the optimal operating conditions for the biomass productivity under day/night cycles using Pontryiagin's maximum principle (assuming a periodic working mode) [61] [90].

### 6.2.2. $CO_2$ fixation by microalgae

**Participants:** Olivier Bernard, Antoine Sciandra, Ghjuvan Grimaud.

*Experimental work*

We have run experiments to observe the response of a population of microalgal cells to various periodic light/dark or nitrate signals. The measurements show the synchronicity of the cells for some conditions. These experiments support the hypothesis that uptake of nitrogen stops during cell division [24].

*Modeling cell cycle*

On this basis, we have developed a structured model representing the development of microalgal cells through three main phases of their cell cycle: G1, G2 and M. The model is made of three interdependent Droop models [107]. The model was validated through extensive comparison with experimental results in both condition of periodic light forcing and nitrogen limitation. The model turns out to accurately reproduce the experimental observations [107].

*Calcification of coccolithophorids*

The effect of $CO_2$ partial pressure increase on photosynthesis and calcification of the calcareous microalgae *Emiliania huxleyi* have been experimentally observed. It results in an increase of the coccolith size together with a decrease in the calcification rate [25].

Three models accounting for the possible coupling between photosynthesis and calcification [75] were included in an ocean model; they account for settling and predation by grazers, and a bloom of coccolithophorids was simulated [76], [77].

*Nitrogen fixation by nitrogenotrophs*

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [60] [89]. The main fluxes of carbon and nitrogen are represented in the cell. The accumulation of starch during the day to fuel the nitrogenase working in the absence of oxygen during the night was the key process to explain the nitrogen fixation. The strong influence of the cell cycle was also included in the model. Finally, the model was calibrated and validated with the data of 3 experiments carried out with different duration of the light period and daily dose. The model succeded to efficiently reproduce the experimental data.

This work is done in collaboration with Sophie Rabouille (CNRS-Océanographic Laboratory of Villefranche-sur-Mer).

*Including phytoplankton photoadaptation into biogeochemical models*

The complexity of the marine ecosystem models and the representation of biological processes, such as photoadaptation, remain an open question. We compared several marine ecosystem models with increasing complexity in the phytoplankton physiology representation in order to assess the consequences of the complexity of photoadaptation models in biogeochemical model predictions. Three models of increasing complexity were considered, and the models were calibrated to reproduce ocean data acquired at the Bermuda Atlantic Time-series Study (BATS) from in situ JGOFS (Joint Global Ocean Flux Study) data. It turns out that the more complex models are trickier to calibrate and that intermediate complexity models, with an adapted calibration procedure, have a better prediction capability [12], [43], [42].

This work is done in collaboration with Sakina Ayata (UPMC-Océanographic Laboratory of Villefranche-sur-Mer).

### *6.2.3. Design of ecologically friendly plant production systems*

#### *6.2.3.1. Controlling plant pests*

**Participants:** Frédéric Grognard, Ludovic Mailleret, Mickaël Teixeira-Alves, 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 natural enemies introductions optimization has been investigated for several years [9] [111]. It had allowed to unveil the crucial influence of within-predator density dependent processes, and especially negative density dependence. In particular, we concluded that pest control is more efficiently achieved through the frequent introduction of small populations of natural enemies as compared to larger and rarer ones. Because contrarily to predatory biocontrol agents, parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, the current modeling effort concentrates on studying the impact of positive predator-predator interactions on the optimal introduction strategies [55].

Connected experimental research is also being pursued in the laboratory on *trichogramma spp.* which tends to show positive density dependence because of demographic stochasticity [32], and the PhD thesis of Thibaut Morel Journel (UMR ISA) has just started on this topic.

*Food source diversity and classical biological control efficiency using generalist natural enemies*

Because generalist biocontrol agents can feed on different food sources like, e.g. a given pest and pollen, they are capable of surviving pest absence within crops [118]. From the biological control point of view, this makes it possible to sustain natural enemies populations able to fight pests at the onset of pest attacks. Moreover, when supplied with different food types, generalists organisms are expected to thrive. Alternative prey, banker plants or more generally habitat enhancement based biological control strategies are thus becoming popular IPM (Integrated Pest Management) methods [112]. Although it has clear advantages, the simultaneous presence of various food sources also has important drawbacks: feeding on different food sources means that a given individual cannot feed on each food source at the same moment. This distraction effect thus potentially reduces the overall predation pressure imposed by the natural enemy population, and the interaction between the demographic response of the predator population and individual behavior is complex. To investigate such questions, we developed and analyzed behavioral-demographic population models taking into account the negative density dependent character of most generalist biocontrol agents. We found out that predator distraction effects can dominate the demographic response of the predator populations, potentially disrupting pest control [120]. An additional conclusion of our study, is that higher predator densities can actually bring about lesser pest suppression. Such results question current biological control practices, and show that, counter-intuitively, recording a lot of predators within fields does not ensure efficient pest control.

*Plant compensation, pest control and plant-pest dynamics*

Plant compensation is the process by which plants respond positively to recover from the effects of pest injury on plant growth. It is a common phenomenon, which has been repeatedly reported in various plant taxa during the last thirty years. Of special interest is the overcompensation phenomenon: consecutively to a pest attack, a plant may reach a higher biomass or have a better fitness compared to the no-pest-attack situation [65]. Although this phenomenon has mainly been documented in wild plants [65] it has also been observed on agricultural plants [121], [113]. To understand better this plant-herbivore interaction and to assess the efficacy of different pest control strategies we built a plant-pest model of plant compensatory growth. We have shown that depending on plants and pests characteristics, plant overcompensation may or may not happen. Moreover, because the model undergoes a backward bifurcation, it is shown that plant overcompensation is also dependent on the level of pest attacks and does not necessarily show up even when the plant-pest couple do have the potential to produce overcompensation [38].

This work is part of the PhD thesis of Audrey Lebon (Cirad), and done in collaboration with Yves Dumont (Cirad).

#### 6.2.3.2. Controlling plant pathogens

**Participants:** Frédéric Grognard, Ludovic Mailleret, Elsa Rousseau.

*Sustainable management of plant resistance*

The introduction of plant strains that are resistant to one pathogen often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants. The resistance strain then becomes useless. It is therefore necessary to develop ways of introducing such resistance into crop production without jeopardizing its future efficiency. We did so by choosing the proportion of resistant plants that are mixed with the non-resistant ones. We studied a vector borne pathogen in a seasonal environment, with healthy crop being planted at the beginning of each season and cropped at its end, the pathogen surviving in the environment during the 'winter'. Two strategies have been proposed, one that aims at minimizing the cumulated damage over a 15 years horizon and one that aims at preventing the virulent strain outbreak. We showed that pathogen's fitness cost associated with resistance breakdown was one of the main factors governing damage reduction at the landscape scale, although the optimal resistance deployment also strongly relied on epidemic characteristics and landscape coonnectivity [16], [51]. The capacity for a virulent virus to establish itself in such an environment, as well as the evolution of the virus characteristics have been studied [59].

This work is done in collaboration with Frédéric Fabre and Benoit Moury (INRA Avignon).

*Eco-evolutionary dynamics of plant pathogens in seasonal environments*

The coexistence of closely related plant parasites is widespread. Yet, understanding the ecological determinants of evolutionary divergence in plant parasites remains an issue. Niche differentiation through resource specialization has been widely researched, but it hardly explains the coexistence of parasites exploiting the same host plant. Most agricultural systems in temperate environments are characterized by the cyclical presence and absence of the crop, due to cropping practices such as harvest and planting. The seasonal character of agrosystems can induce complex plant-pathogens dynamics [19] and is an important force promoting evolutionary diversification of plant pathogens [91]. Plant parasites reproduction mode may strongly interact with seasonality. In this context, we investigated the influence of cyclical parthenogenesis, i.e. the alternation of sexual and asexual reproduction phases, on the eco-evolutionary dynamics of plant parasites [80]. By means of a theoretical approach, we show that an obligate sexual event prior to overseasoning promotes evolutionary divergence in terms of investment into asexual reproduction in plant parasites. Yet, polymorphism may be transient; namely, morphs mostly investing into sexual reproduction may eventually exclude morphs mostly investing into asexual reproduction. Our findings nicely echo with recent population genetics results on *Leptosphaeria maculans*, the causal agent of the blackleg disease of canola, reporting differential investments into sexual and asexual reproduction both at the global and continental scales.

This work is part of the PhD thesis of Magda Castel (Agrocampus Ouest) and is done in collaboration with Frédéric Hamelin (Agrocampus Ouest).

### 6.2.4. Biological depollution - Anaerobic digestion

#### 6.2.4.1. Coupling microalgae to anaerobic digestion

**Participants:** Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Philipp Hartmann.

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 Symbiose project is aiming at evaluating the potential of this process [117], [116].

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 optimization [97], [79], [20]. Considering three main reactions, this model fits adequately the experimental data of an anaerobic digester fed with *Chlorella vulgaris* (data from INRA LBE). On the other hand, we have tested the ability of ADM1 [119] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [99], [100], [98] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

#### 6.2.4.2. *Life Cycle Assessment of microalgae production*

**Participants:** Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Laurent Lardon and Arnaud Helias of INRA-LBE through the co-supervision of Pierre Collet's PhD thesis [83].

An analysis of the potential environmental impacts of biodiesel production from microalgae has been carried out using the life cycle assessment (LCA) methodology [94]. This study has allowed to identify the obstacles and limitations which should receive specific research efforts to make this process environmentally sustainable. This study has been updated and the effects of technological improvements (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 electricty is produced on the plant [86], [85]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage.

A LCA has been carried out to assess the environmental impact of methane production by coupling microalgae and anaerobic digestion. The study highlights the limitation derived by the low biodegradability of the considered microalgae [84] which induces a large digester design and thus more energy to mix and heat it.

### 6.2.5. *Models of ecosystems*

#### 6.2.5.1. *Optimality in consumer-resource dynamics*

**Participants:** Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

*Adaptive behavior in seasonal consumer-resource dynamics*

In this work we studied the evolution of a consumer-resource (or predator-prey) system with seasonal character of the dynamics. We specified two main parts of the process. First, we considered the system during one season with a fixed length: the prey lay eggs continuously and the predators lay eggs or hunt the prey (choose their behavior) according to the solution of an optimal control problem [66]. We then examined how (resident) predators adopting this optimal behavior would fare when faced with a small population of selfish mutants that would be identical to the resident but would have the freedom to choose a different behavior. We studied the resulting optimal control problem where the mutants maximize their own number of offspring using the knowledge of the resident's behavior, and showed that, in most situations, mutants can take advantage of their low frequency and fare better than the residents. Over the course of a large number of seasons, the mutants replace the residents, only to find themselves applying the original resident behavior [52].

*Optimal foraging and residence times variations*

Charnov's marginal value theorem (MVT) [81] is a central tenet of ecological theory. In fragmented environments, the MVT connects the quality and distribution of patches to the optimal time an individual should spend on any patch, and thus the rate of movement in the habitat. Unfortunately, it does not offer explicit predictions regarding how changing habitat quality would affect residence times. In this work, we answer that question in a very general setting, for habitats with homogeneous or heterogeneous patches and with general fitness functions. We then particularize it to the resource consumption framework and indicate how the residence times variations relate to the curvatures of the functional responses [49], [78].

This last work is done in collaboration with Vincent Calcagno and Eric Wajnberg (INRA Sophia Antipolis)

#### 6.2.5.2. *Growth models of zooplankton*

**Participants:** Jean-Luc Gouzé, Jonathan Rault, Eric Benoît.

The model built to describe a zooplankton community is some variant of the McKendrick-Von Foerster Equation. The model includes cannibalism within zooplankton communities and predation on phytoplankton. Dynamic mass budget theory is used in order to describe individual behavior and allows mass conservation. Also we have added phytoplankton dynamics, and we use environmental data as an input for the model. The aim is to compare simulations with data provided by the Laboratoire d'Océanographie de Villefranche (Lars Stemmann). We have also built a discrete size-structured model. Discrete models are less numerically demanding and so can be more easily incorporated into bigger models. Moreover the study of discrete models are often easier than that of continuous ones. We focus our study on the impact of cannibalism within the zooplankton community and show that under some hypotheses, cannibalism can stabilize the equilibrium of the model [29], [11]. We also address the problem of control of such models (by harvesting or biological control); we obtain results for stabilization of the equilibrium [41], [11].

## 6.3. Software design

**Participants:** Olivier Bernard, Mélaine Gautier.

Over the years, BIOCORE has been developing a software framework for bioprocess control and supervision called ODIN [71]. This C++ application (working under Windows and Linux) enables researchers and industrials to easily develop and deploy advanced control algorithms through the use of a Scilab interpreter [46], [47]. It also contains a Scilab-based process simulator which can be harnessed for experimentation and training purposes. ODIN is primarily developed in the C++ programming language and uses CORBA to define component interfaces and provide component isolation. ODIN is a distributed platform, enabling remote monitoring of the controlled processes as well as remote data acquisition. Recently, a software development effort has been directed to the graphical user interface, a synoptic view component, new drivers for the experimental hardware and integration of the PlantML data exchange format. ODIN has been tested on four different processes and has been set up with Eric Latrille to supervise the 66m2 high rate pond at the LBE, INRA Narbonne.

# CARMEN Team

# 5. New Results

## 5.1. Models

- [12]: we explain the links between the solutions of the bidomain and monodomain models using some analytical arguments. The result is partially based on the theory of the bidomain operator explained in [11].
- [23]: Fibre structure and anisotropy is a determinant issue to provide accurate simulations of the electrical activity of atrial tissue. Though, atrial fibre architecture remains unreachable to standard imagery techniques on patients. A method to construct models of the fibre architecture on patient-specific geometries is then a key for numerical simulations of atrial tissues. Such a method is proposed. Pathological and non pathological patient specific surface models of the left atria (LA) are defined. Hence, a pathological scenario is explored : a mechanism of micro- reentry in the left superior pulmonary vein (LSPV) and its interaction with the sinus rhythm (SR).

## 5.2. Numerical techniques

- [19]: In this paper we propose a preconditioning for the bidomain model either for an isolated heart or in an extended framework including a coupling with the surrounding tissues (the torso). The preconditioning is based on a formulation of the discrete problem that is shown to be symmetric positive semi-definite. A block LU decomposition of the system together with a heuristic approximation (referred to as the monodomain approximation) are the key ingredients for the preconditioning definition. Numerical results are provided for two test cases: a 2D test case on a realistic slice of the thorax based on a segmented heart medical image geometry, a 3D test case involving a small cubic slab of tissue with orthotropic anisotropy. The analysis of the resulting computational cost (both in terms of CPU time and of iteration number) shows an almost linear complexity with the problem size, i.e. of type nlog$\alpha$(n) (for some constant $\alpha$) which is optimal complexity for such problems.

## 5.3. Medical applications of numerical models

- [26]: We computed some bidomain solutions for use by M. Pop and M. Sermesant in the STACOM'11 challenge from the MICCAI 2011 conference and derived collaborative article [26].
- [18]: The aim of this study was to describe a new familial cardiac phenotype and to elucidate the electrophysiological mechanism responsible for the disease. Mutations in several genes encoding ion channels, especially SCN5A, have emerged as the basis for a variety of inherited cardiac arrhythmias. Three unrelated families comprising 21 individuals affected by multifocal ectopic purkinje-related premature contractions (MEPPC) characterized by narrow junctional and rare sinus beats competing with numerous premature ventricular contractions with right and/or left bundle branch block patterns were identified. All the affected subjects carried the same transition in the SCN5A gene. Patch-clamp studies revealed a net gain of function of the sodium channel, leading, in silico, to incomplete repolarization in Purkinje cells responsible for premature ventricular action potentials. In vitro and in silico studies recapitulated the normalization of the ventricular action potentials in the presence of quinidine.
- [22]: In some cases, the standard methods to construct activation maps based on the derivatives of the signals may lead to inaccurate results. In this paper, we evaluated a novel Directional Activation Algorithm (DAA) based on EGM analysis. The DAA calculates the time delays between adjacent EGMs and assigns to each a localized propagation vector. The accuracy of the proposed methodology is compared with known activities obtained from a monodomain, isotrope, Beeler-Reuter model of the atria.

- [20]: Although the ECG is a widely used tool, the ionic basis underlying its changes caused by drugs and diseases are often unclear. In this work we present a computational model of the human ECG capable of representing drug-induced effects from the ionic to the surface potential level. We use the state-of-the-art bidomain model coupled to a membrane kinetics model in the heart and the Laplace equation in the torso. The membrane kinetics are represented by a detailed physiological human action potential model. We modified the potassium (respectively sodium) representation in the model in order to introduce the ion channel/drug interactions representing classIII (respectively class I) drugs. The drug model is represented by an ion channel conduction block depending on the IC50 value and the drug dose. We conduct numerical simulation of the ECGs measured on the surface of the thorax and could assess each of the potassium and sodium block effects (for class I and class III drugs).

## 5.4. Inverse problems

- [24]: The treatment of atrial fibrillation has greatly changed in the past decade. Ablation therapy, in particular pul- monary vein ablation, has quickly evolved. However, the sites of the trigger remain very difficult to localize. In this study we propose a machine-learning method able to non-invasively estimate a single site trigger. The machine learning technique is based on a kernel ridge regression algorithm. In this study the method is tested on a simulated data. We use the monodomain model in order to simulate the electrical activation in the atria. The ECGs are computed on the body surface by solving the Laplace equation in the torso.
- [16]: In the present paper, an optimal control problem constrained by the tridomain equations in electrocardiology is investigated. The state equations consisting in a coupled reaction–diffusion system modeling the propagation of the intracellular and extracellular electrical potentials, and ionic currents, are extended to further consider the effect of an external bathing medium. The existence and uniqueness of solution for the tridomain problem and the related control problem is assessed, and the primal and dual problems are discretized using a finite volume method which is proved to converge to the corresponding weak solution. In order to illustrate the control of the electrophysiological dynamics, we present some preliminary numerical experiments using an efficient implementation of the proposed scheme.
- [17]: This note is devoted to the analysis of the null controllability of a nonlinear reaction–diffusion system, approximating a parabolic–elliptic system, modeling electrical activity in the heart. The uniform, with respect to the degenerating parameter, null controllability of the approximating system by a single control force acting on a subdomain is shown. The proof needs a precise estimate with respect to the degenerating parameter and it is done combining Carleman estimates and energy inequalities.

# DRACULA Project-Team

# 6. New Results

## 6.1. Modelling of Erythroblastic Islands (red blood cell production)

**Participants:** Fabien Crauste [Contact person], Olivier Gandrillon, Vitaly Volpert.

*In collaboration with N. Bessonov, S. Fischer and P. Kurbatova.*

The production and regulation of red blood cells (erythropoiesis) occurs in the bone marrow where erythroid cells proliferate and differentiate within particular structures, called erythroblastic islands. A typical structure of these islands consists of a macrophage (white cell) surrounded by immature erythroid cells (progenitors), with more mature cells on the periphery of the island, ready to leave the bone marrow and enter the bloodstream.

We proposed a hybrid model [11], coupling a continuous model (ordinary differential equations) describing intracellular regulation through competition of two key proteins, to a discrete spatial model describing cell-cell interactions, with growth factor diffusion in the medium described by a continuous model (partial differential equations), to investigate the role of the central macrophage in normal erythropoiesis. Intracellular competition of the two proteins leads the erythroid cell either to proliferation, differentiation, or death by apoptosis. This approach allows considering spatial aspects of erythropoiesis, involved for instance in the occurrence of cellular interactions or the access to external factors, as well as dynamics of intracellular and extracellular scales of this complex cellular process, accounting for stochasticity in cell cycle durations and orientation of the mitotic spindle. The analysis of the model showed a strong effect of the central macrophage on the stability of an erythroblastic island, when assuming the macrophage releases prosurvival cytokines. Even though it is not clear whether or not erythroblastic island stability must be required, investigation of the model concludes that stability improves responsiveness of the model, hence stressing out the potential relevance of the central macrophage in normal erythropoiesis.

## 6.2. Modelling of the CD8 T cell Immune Response

**Participants:** Fabien Crauste [Contact person], Olivier Gandrillon, Emmanuelle Terry.

*In collaboration with J. Marvel and C. Arpin.*

The CD8 immune response is a specific immune response triggered by the organism when the innate response is unable to fight a pathogen. We proposed a new model of the CD8 T cell immune response based on the description of feedback controls exerted by the cytotoxic CD8 T cell population on the pathogen and the population itself [14]. This model, a system of ordinary and age-structured partial differential equations, allows describing a classical response, characterized by a cellular expansion following the pathogen-mediated activation, then a contraction phase and the generation of memory CD8 T cells. Moreover, we showed the global asymptotic stability of this system corresponding to the elimination of the virus. This situation is expected and describes for instance what is observed with the flu virus.

A simpler version of this model (based on nonlinear ordinary differential equations) has then been confronted to experimental data generated by Jacqueline Marvel's team in Lyon (immunology team), with 3 different pathogens. A parameter sweep has been performed and some parameters of the model, specific of cellular processes, have been shown to characterize CD8 immune responses against either a virus or a bacterium. This work is in progress and should be submitted soon.

## 6.3. Modelling of Platelet Thrombus Formation

**Participants:** Alen Tosenberger, Vitaly Volpert [Contact person].

*In collaboration with F. Ataullakhanov, N. Bessonov, A. Butylin, G. Panasenko, M. Panteleev, E. Shnol, I. Sirakov and A. Tokarev.*

An injury of a blood vessel requires quick repairing of the wound in order to prevent a loss of blood. This is done by the hemostatic system. The key point of its work is the formation of an aggregate from special blood elements, namely, platelets. The construction of a mathematical model of the formation of a thrombocyte aggregate with an adequate representation of its physical, chemical, and biological processes is an extremely complicated problem. A large size of platelets compared to that of molecules, strong inhomogeneity of their distribution across the blood flow, high shear velocities, the moving boundary of the aggregate, the interdependence of its growth and the blood flux hamper the construction of closed mathematical models convenient for biologists. We proposed a new PDE-based model of a thrombocyte aggregate formation [21], [22]. In this model, the movement of its boundary due to the adhesion and detachment of platelets is determined by the level set method. The model takes into account the distribution inhomogeneity of erythrocytes and platelets across the blood flow, the invertible adhesion of platelets, their activation, secretion, and aggregation. The calculation results are in accordance with the experimental data concerning the kinetics of the ADP-evoked growth of a thrombus in vivo for different flow velocities. The model constructed here can be easily extended to the case of other hemostatic mechanisms and can be integrated into different continuous blood flow models.

## 6.4. Reaction-Diffusion Model of Atherosclerosis Development

**Participant:** Vitaly Volpert [Contact person].

*In collaboration with N. El Khatib, S. Genieys and B. Kazmierczak.*

Atherosclerosis begins as an inflammation in blood vessel walls (intima). The inflammatory response of the organism leads to the recruitment of monocytes. Trapped in the intima, they differentiate into macrophages and foam cells leading to the production of inflammatory cytokines and further recruitment of white blood cells. This self-accelerating process, strongly influenced by low-density lipoproteins (cholesterol), results in a dramatic increase of the width of blood vessel walls, formation of an atherosclerotic plaque and, possibly, of its rupture. We suggested a 2D mathematical model of the initiation and development of atherosclerosis which takes into account the concentration of blood cells inside the intima and of pro- and anti-inflammatory cytokines [18]. The model represents a reaction-diffusion system in a strip with nonlinear boundary conditions which describe the recruitment of monocytes as a function of the concentration of inflammatory cytokines. We proved the existence of travelling waves described by this system and confirmed our previous results which suggest that atherosclerosis develops as a reaction-diffusion wave.

## 6.5. Hematopoietic model with feedback control

**Participants:** Mostafa Adimy [Contact person], Lila Sebaa.

*In collaboration with O. Angulo and C. Marquet.*

We investigate a mathematical model of blood cell production in the bone marrow (hematopoiesis). The model describes both the evolution of primitive hematopoietic stem cells and the maturation of these cells as they differentiate to form the three types of blood cells (red blood cells, white cells and platelets). The primitive hematopoietic stem cells and the first generations of each line (progenitors) are able to self-renew, and can be either in a proliferating or in a resting phase ($G_0$-phase). These properties are gradually lost while cells become more and more mature. The three types of progenitors and mature cells are coupled to each other via their common origin in primitive hematopoietic stem cells compartment. Peripheral control loops of primitive hematopoietic stem cells and progenitors as well as a local autoregulatory loop are considered in the model. The resulting system is composed by eleven age-structured partial differential equations. To analyze this model, we don't take into account cell age-dependence of coefficients, that prevents a usual reduction of the structured system to an unstructured delay differential system. We investigate some fundamental properties of the solutions of this system, such as boundedness and positivity. We study the existence of stationary solutions: trivial, axial and positive steady states. Then we give conditions for the local asymptotic stability of the trivial steady state and by using a Lyapunov function, we obtain a sufficient condition for its global asymptotic stability [7]. In some particular cases, we analyze the local asymptotic stability of the positive steady state by using the characteristic equation. Finally, by numerical simulations, we illustrate our results and we show that a change in the duration of cell cycle can cause oscillations. This can be related to observations of some periodical hematological disease such as chronic myelogenous leukemia, cyclical neutropenia, cyclical thrombocytopenia, etc.

# MACS Project-Team

# 6. New Results

## 6.1. Asymptotic and multiscale modeling in biomechanics

### 6.1.1. Detailed validations of muscle model

**Participants:** Matthieu Caruel, Dominique Chapelle, Alexandre Imperiale, Philippe Moireau.

Until recently we had only considered simple isotropic passive laws of Mooney-Rivlin type in our muscle model, albeit with an overall behavior already highly non-isotropic due to the fiber-oriented active component. We have now implemented and calibrated a new visco-hyperelastic passive law of exponential and orthotropic type for the hyperelastic part, in better agreement with the models and data generally found in the literature. It should be noted that most experimental data available concern the passive behavior only, indeed. In addition, we implemented a new conservative numerical scheme for the time discretization of the contractile variables. Moreover an original boundary condition of contact type has been successfully applied on several detailed cardiac geometries to represent the interactions between the epicardium, pericardium and the surrounding structures.

Major advances in the understanding of heart contraction cycle can be achieved by testing papillary muscle preparations *in vitro*. Single papillary muscles have an essentially one-dimensional structure suitable for uniaxial mechanical testing, and therefore represent the simplest setup to test the robustness of a model of heart contraction against a vast set of experimental results available in the literature. In collaboration with Y. Lecarpentier (Institut du Coeur, Pitié-Salpêtrière Hospital, Paris) and R. Chabiniok (King's College London), we have further refined and calibrated the muscle mechanical model in order to quantitatively reproduce experimental data from rat cardiomyocytes. These results include the static stress-strain constitutive relation, kinetic response to isotonic loadings, and force-velocity relation see Fig.1 .

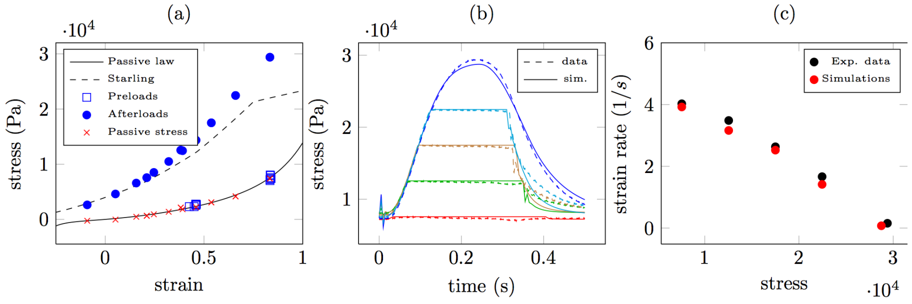


*Figure 1. Results of the model compared with experimental data*

### 6.1.2. Multi-scale mechanics of muscle contraction

**Participant:** Matthieu Caruel.

Muscles are an active tissue material capable of producing force. At the microscale, force is the result of complex interactions between two types of proteins, namely, actin and myosin, which work coherently in very large assemblies ($\sim 10^9$). The passive mechanical response of so-called striated muscles at fast time scales is dominated by long range interactions inducing cooperative behavior without breaking the detailed balance. This leads to such unusual material properties as negative equilibrium stiffness and drastically different behavior in force and displacement controlled loading conditions. Analysing experimental data strongly suggests that muscles are finely tuned to perform close to a critical point (see Fig.2 ). This work in collaboration with Jean-Marc Allain and Lev Truskinovsky (LMS, Ecole Polytechnique) is the subject of a paper submitted to Physical Review Letters (see [22]).

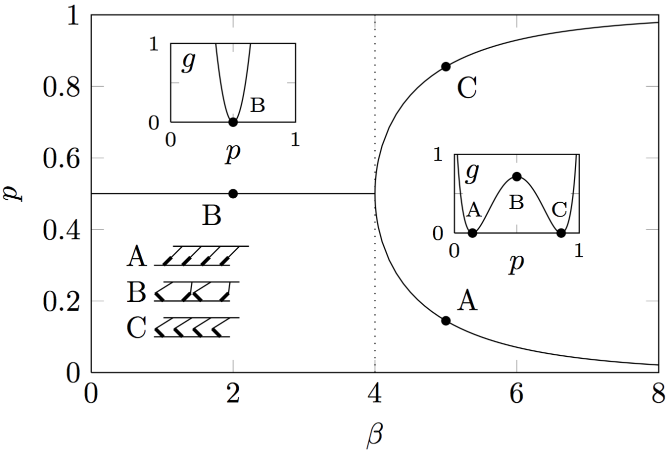


*Figure 2. Bifurcation diagram of a model of coupled molecular motors. $p$ is the fraction of motors in the stress generating configuration (post-power-stroke). $\beta$ is a non dimensional parameter representing the intensity of thermal fluctuations ($\beta \to 0$ represents infinitely strong thermal forces). For $\beta < 4$, the system lives in a mixed configuration (B): the free energy is convex with a minimum at $p = 1/2$ (see the left inset showing the energy landscape g). For $\beta > 4$ the system is organized in two distinct populations (A and C) corresponding to the 2 minima of a non-convex energy landscape (see the right inset). One population is mainly pre-power-stroke (A) while the other is post-power-stroke (C). This is a signature of mechanical cooperativity.*

### 6.1.3. Asymptotic analysis applied to cardiac electrophysiology modeling

**Participants:** Dominique Chapelle, Annabelle Collin, Jean-Frédéric Gerbeau [(REO team)].

Computational electrophysiology is a very active field with tremendous potential in medical applications, albeit leads to highly intensive simulations. We here propose a surface-based electrophysiology formulation, motivated by the modeling of thin structures such as cardiac atria, which greatly reduces the size of the computational models. Moreover, our model is specifically devised to retain the key features associated with the anisotropy in the diffusion effects induced by the fiber architecture, with rapid variations across the thickness which cannot be adequately represented by naive averaging strategies. Our proposed model relies on a detailed asymptotic analysis in which we identify a limit model and establish strong convergence results. We also provide detailed numerical assessments which confirm an excellent accuracy of the surface-based model – compared with the reference 3D model – including in the representation of a complex phenomenon, namely, spiral waves, see Figure 3 . This work was submitted for publication in "M3AS: Mathematical Models and Methods in Applied Sciences".

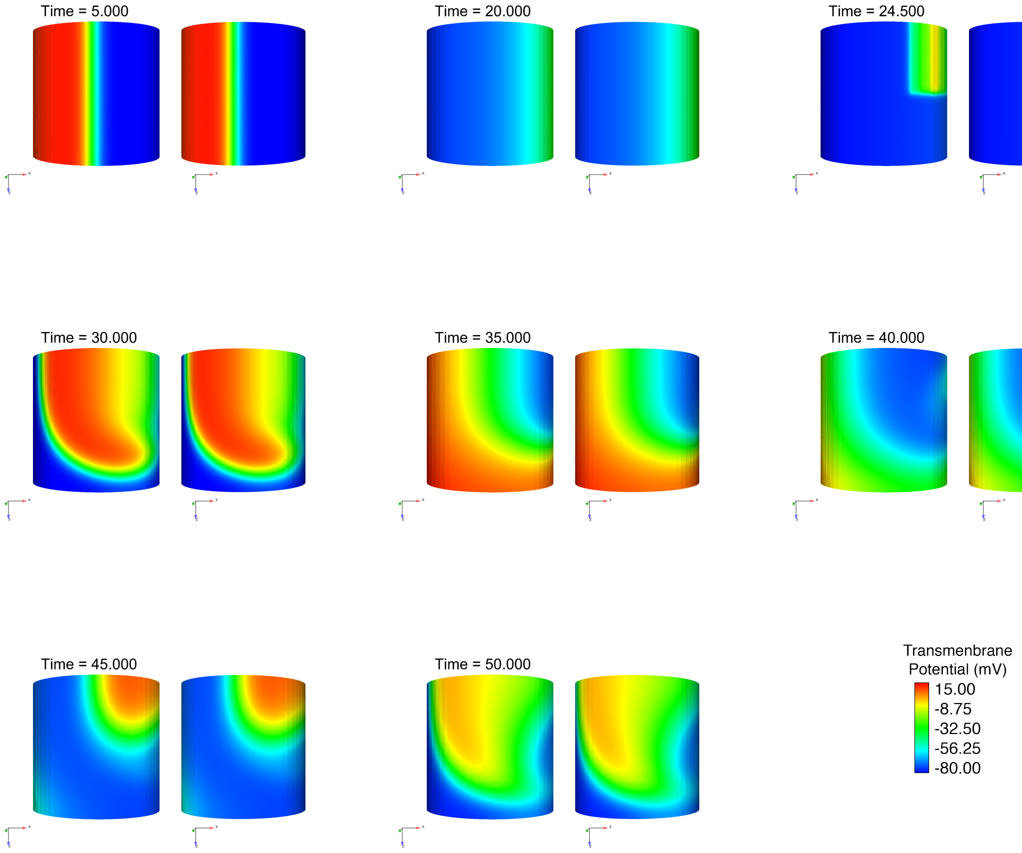


*Figure 3. Spiral wave on cylinder – Comparison of asymptotic surface model (left), 3D model (center) and naive 2D model (right) on the midsurface at 8 consecutive times*

### 6.1.4. Cardiac atria electrophysiology surface-based modeling and assessment of physiological simulations

**Participants:** Dominique Chapelle, Annabelle Collin, Jean-Frédéric Gerbeau [(REO team)].

We aim at validating the 2D (namely, surface-based) electrophysiology model designed for thin cardiac structures with strongly heterogeneous anisotropy, presented in Paragraph 6.1.3 with a real model of the atria. We produced a surface mesh representing the mid-surface of the two atria. We used the bibliography to identify and prescribe the fibers directions at the endocardium and epicardium. Figure 4 displays the simulation results obtained with the surface-based model.

### 6.1.5. Strong convergence results in the asymptotic behavior of the 3D-shell model

**Participants:** Dominique Chapelle, Annabelle Collin.

The objective of this work is to establish the strong convergence for the asymptotic analysis of the so-called 3D-shell model presented in [2]. We apply similar methods to those used in the work on "Asymptotic analysis applied to cardiac electrophysiology modeling".

## 6.2. Estimation in biomechanics

### 6.2.1. Exponential convergence of an observer based on partial field measurements for the wave equation

**Participants:** Dominique Chapelle, Philippe Moireau.

We analyze an observer strategy based on partial—that is, in a subdomain—measurements of the solution of a wave equation, in order to compensate for uncertain initial conditions. We prove the exponential convergence of this observer under a nonstandard observability condition, whereas using measurements of the time derivative of the solution would lead to a standard observability condition arising in stabilization and exact controllability. Nevertheless, we directly relate our specific observability condition to the classical geometric control condition. Finally, we provide some numerical illustrations of the effectiveness of the approach. This work in collaboration with M. de Buhan (Univ. Paris V) and N. Cîndea (Univ. Clermont-Ferrand) is published in [13].

### 6.2.2. Sequential identification of boundary support parameters in a fluid-structure vascular model using patient image data

**Participants:** Dominique Chapelle, Philippe Moireau.

This work [17] is in collaboration with C. Bertoglio and J.-F. Gerbeau (REO team) and N. Xiao, C.A. Figueroa and C.A. Taylor (Stanford University), where we propose a complete methodological chain for the identification of the corresponding boundary support parameters, using patient image data. We consider distance maps of model to image contours as the discrepancy driving the data assimilation approach, which then relies on a combination of (1) state estimation based on the so-called SDF filtering method, designed within the realm of Luenberger observers and well-adapted to handling measurements provided by image sequences, and (2) parameter estimation based on a reduced-order UKF filtering method which has no need for tangent operator computations and features natural parallelism to a high degree. Implementation issues are discussed, and we show that the resulting computational effectiveness of the complete estimation chain is comparable to that of a direct simulation. Furthermore, we demonstrate the use of this framework in a realistic application case involving hemodynamics in the thoracic aorta. The estimation of the boundary support parameters proves successful, in particular in that direct modeling simulations based on the estimated parameters are more accurate than with a previous manual expert calibration. This paves the way for complete patient-specific fluid-structure vascular modeling in which all types of available measurements could be used to estimate additional uncertain parameters of biophysical and clinical relevance.

This work published in BMMB (impact factor 3.192) can be considered as the first trial of data assimilation using real data in hemodynamics.

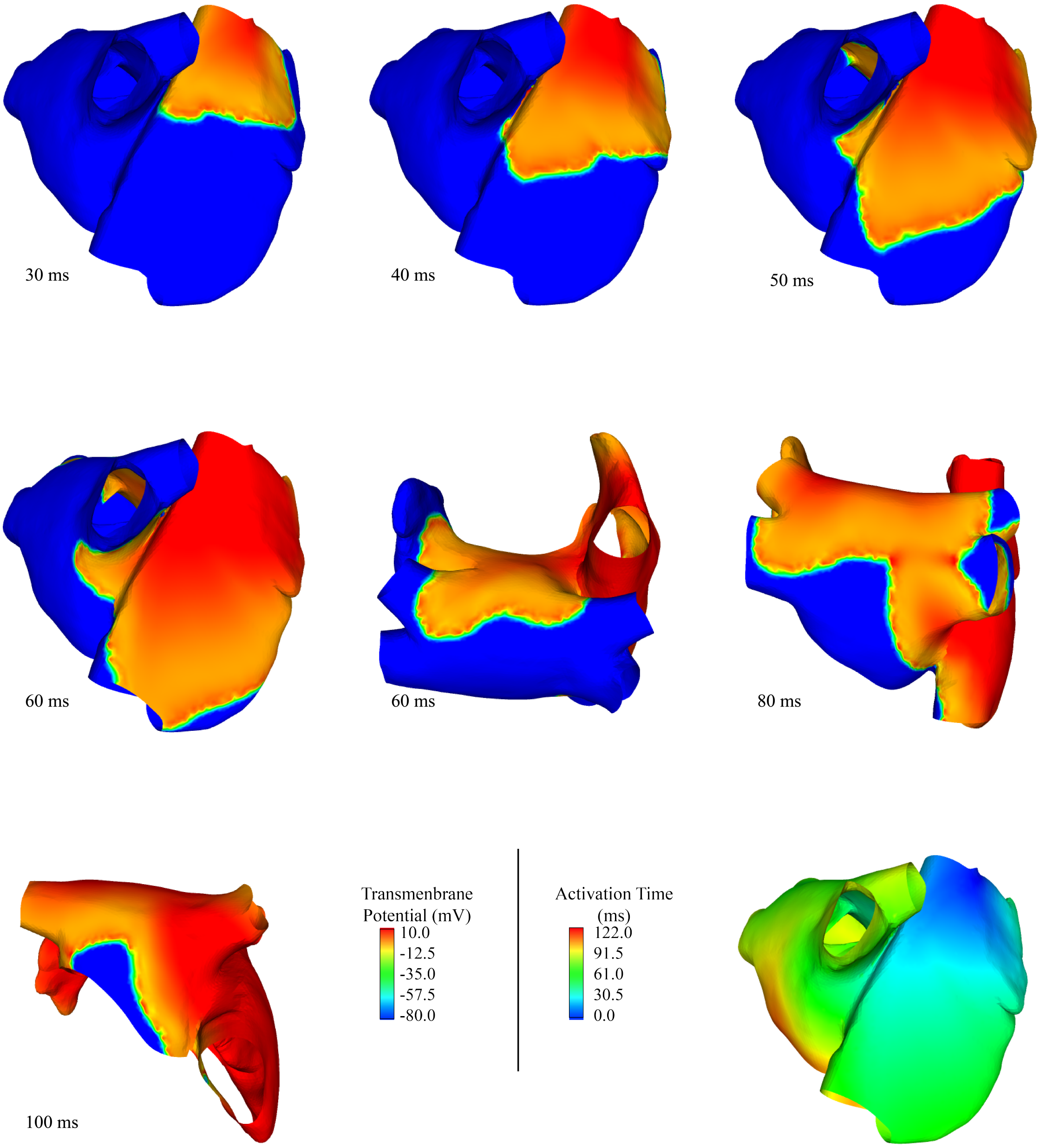


*Figure 4. Simulation of atrial depolarization*

### 6.2.3. Filtering strategies using image data

**Participants:** Alexandre Imperiale, Philippe Moireau, Alexandre Routier.

Some progress has been achieved concerning the Luenberger filtering procedure – also known as nudging – for the cardiovascular system in several directions. We have studied the impact of data interpolation (in time and space) on the method performance (a paper on this subject is being prepared) and, during Alexandre Routier end-of-curriculum internship from INSA Rouen, we have adapted the formalism of currents (inspired by a collaboration with S. Durrleman). This formalism in an elegant way to represent geometric objects (endo- and epicardium surfaces for example) as operators on a test vector space defined on the ambient space. From this key idea the main work was to define a numerical tractable norm on the space of surfaces and derive it with respect to the Lagrangian displacement of the solid domain in order to incorporate such a representation of surfaces into our filtering technique. Among other advantages this new observer requires significantly less prior efforts in terms of image processing.

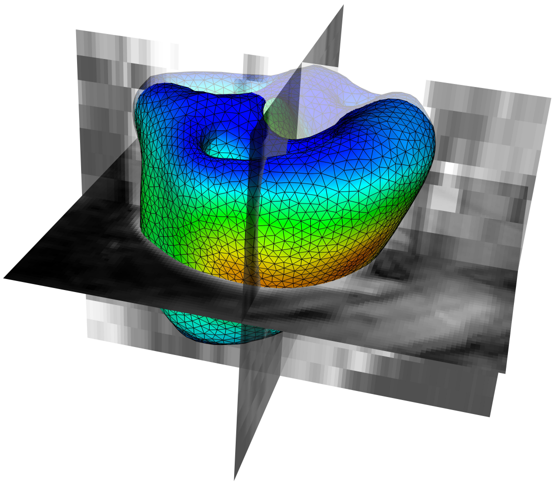


*Figure 5. Results of the model compared with experimental data*

### 6.2.4. Formulation of observers for parabolic equations

**Participants:** Karine Mauffrey, Philippe Moireau.

We are currently working on optimal filtering using observers for a class of evolution PDEs including heat-like equations. As for the optimal control issue,the optimal filtering issue is related to the resolution of a differential Riccati equation. In [25] or [29], the link between the optimal filtering formulations and the derived Riccati equation is done by finite dimension arguments. There exist also other results on the linear quadratic optimal control that are based on infinite dimensional considerations (see, for example, [26] and [28]). A work in progress consists in presenting a direct approach for the optimal filtering issue, using infinite dimension considerations only. Then we should be able to introduce reduced-rank considerations to be able to stabilize only the low frequencies part of the parabolic system, and therefore offer a discretization strategy. This discretization will the be analyzed in details.

## 6.3. Other topics

### 6.3.1. Sail modeling

**Participants:** Dominique Chapelle, Daniele Trimarchi.

This is a collaboration with Marina Vidrascu (REO team) and Stephen Turnock and Dominic Taunton (Southampton University), as part of the recently completed PhD of Daniele Trimarchi. We propose a method of modelling sail type structures which captures the wrinkling behaviour of such structures. The method is validated through experimental and analytical test cases, particularly in terms of wrinkling prediction. An enhanced wrinkling index is proposed as a valuable measure characterizing the global wrinkling development on the deformed structure. The method is based on a pseudo-dynamic finite element procedure involving non-linear MITC shell elements. The major advantage compared to membrane models generally used for this type of analysis is that no ad hoc wrinkling model is required to control the stability of the structure. We demonstrate our approach to analyse the behaviour of various structures with spherical and cylindrical shapes, characteristic of downwind sails over a rather wide range of shape and constitutive parameters. In all cases convergence is reached and the overall flying shape is most adequately represented, which shows that our approach is a most valuable alternative to standard techniques to provide deeper insight into the physical behaviour. Limitations appear only in some very special instances in which local wrinkling-related instabilities are extremely high and would require specific additional treatments, out of the scope of the present study. This work has been published in [20].

### 6.3.2. PODs for parameter-dependent problems and estimation

**Participants:** Dominique Chapelle, Philippe Moireau.

This work – submitted to M2AN [24] – is derived from the latest part of Asven Gariah's PhD, jointly supervised by Jacques Sainte-Marie (Bang team) and D. Chapelle, and defended in late 2011. We address the issue of parameter variations in POD approximations of time-dependent problems, without any specific restriction on the form of parameter dependence. Considering a parabolic model problem, we propose a POD construction strategy allowing us to obtain some *a priori* error estimates controlled by the POD remainder – in the construction procedure – and some parameter-wise interpolation errors for the model solutions. We provide a thorough numerical assessment of this strategy with the FitzHugh-Nagumo 1D model. Finally, we give detailed illustrations of the approach in two parameter estimation applications, the first in a variational estimation framework with the FitzHugh-Nagumo model, and the second with a beating heart mechanical model for which we employ a sequential estimation method to characterize model parameters using real image data in a clinical case.

# MASAIE Project-Team

# 5. New Results

## 5.1. Robustness and $\mathcal{R}_0$

We have obtained new results about the relation between Robustness and the basic reproduction number $\mathcal{R}_0$. It is now well admitted that the basic reproduction ratio $\mathcal{R}_0$ is a key concept in mathematical epidemiology and the literature devoted to this concept is now quite important, see [20], [40], [19], [22], [23], [24], [26], [28], [30], [34] and references therein.

This number is a threshold parameter for bifurcation of an epidemic system : for a general compartmental disease transmission model, if $\mathcal{R}_0 < 1$, the disease free equilibrium (DFE) is locally asymptotically stable; whereas, if $\mathcal{R}_0 > 1$, the DFE is unstable.

It is said in some papers that $\mathcal{R}_0$ is a measure to gauge the amount of uniform effort needed to eliminate infection from a population [22], [24], [25], [31], [30].

The concept of robustness, coming from control theory, is associated to uncertainty. Usually the parameters of a system are known within a certain margin. A question is, how some properties, e.g. stability, can be ascertained with uncertainty on the parameters. In control theory “stability margin” is an important concept. Another way to formulate this problem is to analyze the effect of perturbations, unstructured or structured. This problem is also related to the so-called pseudo-spectrum [36], [37], [35].

We found that the basic reproduction number of an epidemic system is not an accurate gauge of the distance from the Jacobian $J$ of this system, computed at the disease free equilibrium, to the set of stable matrices (if $J$ in unstable), respectively to the set of unstable matrices (if $J$ is stable). The same conclusion arises for another indicator, introduced by Heestebeck et al. [24], [31], [30], the type-reproduction number.

## 5.2. Wolbachia and Dengue

*Wolbachia* is a genus of bacteria which infects arthropod species, including a high proportion of insects. It is one of the world’s most common parasitic microbes and is possibly the most common reproductive parasite in the biosphere. *Wolbachia* is a maternally inherited endosymbiont of a large number of insects and other arthropods that induces various effects on host reproductive biology. Estimated to infect more than 60% of all insect species *Wolbachia* species are present in mature eggs, but not mature sperm. Only infected females pass the infection on to their offspring. Another consequence of infection is cytoplasmic incompatibility, i.e., the inability of *Wolbachia*-infected males to successfully reproduce with uninfected females.

The successful introduction of a life-shortening strain of *Wolbachia* into the dengue vector *Aedes aegypti* that halves adult lifespan has recently been reported.

Mosquitoes carrying this *Wolbachia* strain show around a 50% reduction in adult female lifespan compared to uninfected mosquitoes. It has been reported that wMel and wMelPop-CLA strains block transmission of dengue serotype 2 (DENV-2) in *Aedes aegypti*, forming the basis of a practical approach to dengue suppression. Infection by *Wolbachia* has a triple effect : reduction of recruitment, increasing of mortality for the mosquitoes and reduction of dengue transmission.

With our colleague of Brazil (see International cooperation) we built and study different models for the introduction of *Wolbachia* in a population of *Aedes aegypti*. These models are epidemiological models with vertical transmission only, which is quite new. We found that bistabilty does exist : three equilibria are present. We show that the coexistence equilibrium is unstable. We show that the equilibrium without infection and the equilibrium with the whole population infected are asymptotically stable. Numerical experimentation shows that the basin of the second equilibrium is appreciable. This indicates that introduction of *Wolbachia* is feasible. The connection of theses models with transmission models of dengue is under investigation by the French-Brazilian team.

## 5.3. Bilharzia

Schistosomiasis or bilharzia is a water-borne parasitic disease that affects 200 million people and poses a treat to 600 million in more than 76 countries [39]. It is caused by blood-dwelling fluke worms of the genus *Schistosoma*. The transmission cycle requires contamination of surface water by excreta, specific freshwater snails as intermediate hosts, and human water contact [21]. Schistosome are transmitted via contact with contaminated water containing cercaria the infective stage of the parasite [39], [32].

In connection with EPLS, a research NGO based in Saint-Louis (Senegal), and Pasteur Institute of Lille, we investigate a spatially deterministic metapopulation model in which infectious agents persist within a network of connected environments. This model accounts for human population age and behavior structure. We completely analyses the asymptotic behavior of this model. We give a formula for computing the basic reproduction ratio $\mathcal{R}_0$. If $\mathcal{R}_0 \leq 1$ we prove that the disease free equilibrium is globally asymptotically stable. If $\mathcal{R}_0 > 1$, with an hypothesis on connectedness, we prove that there exists a unique positive endemic equilibrium, which is globally asymptotically stable.

The validation of this model, using data of EPLS, is under investigation and is the subject of a Phd thesis. The defense will occur at the beginning of 2013. We explore the identification of key parameters using different kind of observers.

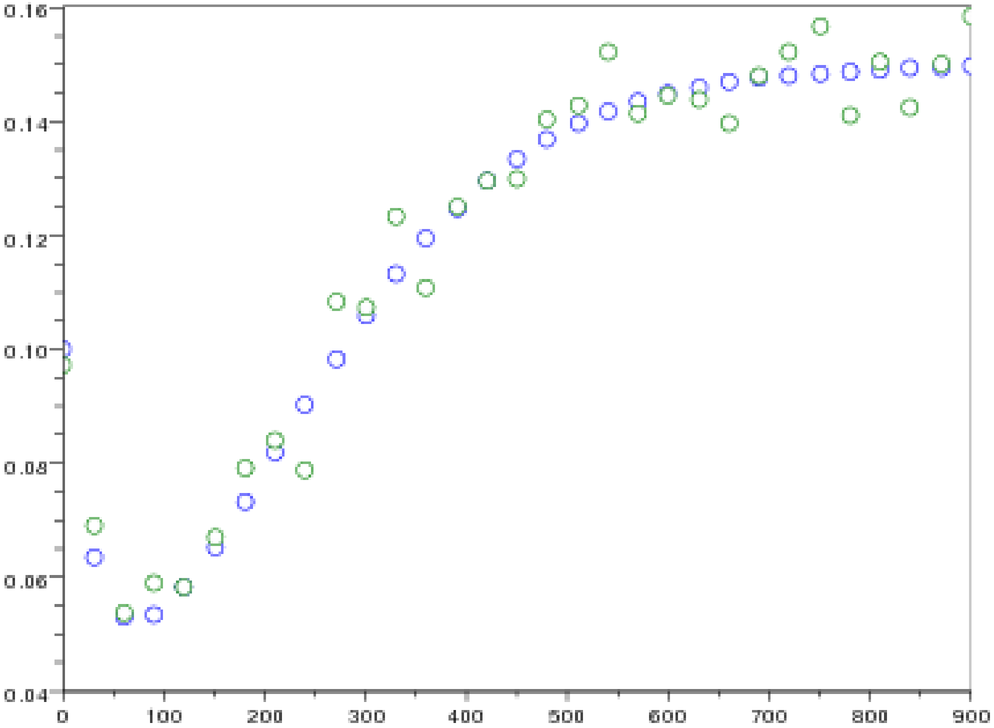


*Figure 1. Noisy and discrete measure of host prevalence*

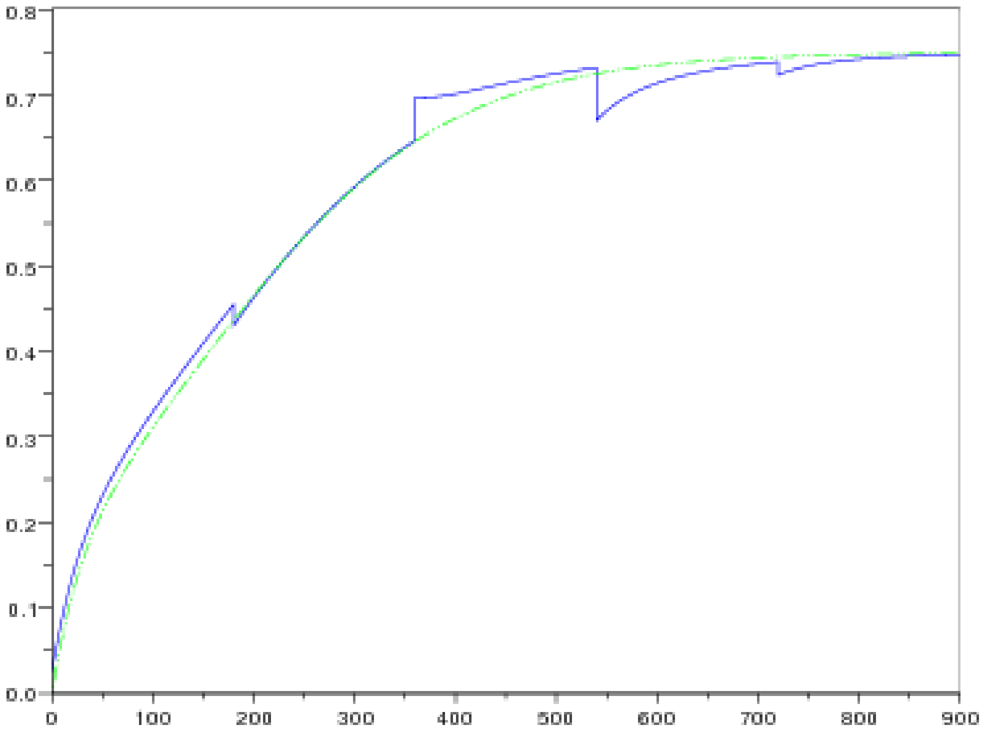


*Figure 2. Reconstruction of the snail prevalence from preceding data*

# MODEMIC Project-Team

# 6. New Results

## 6.1. Theoretical results

### 6.1.1. Models resource/consumer

The team maintains a significant activity about the theory of the chemostat model, proposing and studying extensions of the classical models.

#### 6.1.1.1. Theory of competition and coexistence

**Participants:** Jérôme Harmand, Claude Lobry, Tewfik Sari.

In the papers [41], [50] we consider deterministic models of competition. We study the persistence of species. In [25] we study a syntrophic relation between microbial species. In [26], we give a global asymptotic stability result for a mathematical model of competition between several species in a chemostat, by using a new Lyapunov function. The model includes both monotone and non-monotone response functions, distinct removal rates for the species and variable yields, depending on the concentration of substrate.

#### 6.1.1.2. Study of interconnected chemostats

**Participants:** Jérôme Harmand, Alain Rapaport.

We have shown how a particular spatial structure with a buffer globally stabilizes the chemostat dynamics with non-monotonic response function, while this is not possible with single, serial or parallel chemostats of the same total volume and input flow. We give a characterization of the set of such configurations that enjoy this property, as well as the configuration that ensures the best nutrient conversion. Furthermore, we characterize the minimal buffer volume to add to a single chemostat for obtaining the global stability. These results are illustrated with the Haldane function that models inhibition in micro-organisms growth [67].

In industrial applications, the attraction of the wash-out equilibrium is undesired because it presents a risk that may ruin the culture in case of disturbance, temporarily pump breakdown or presence of toxic material that could drive the state in the attracting basin of the wash-out equilibrium. This approach has led to a patent deposit by INRA [59] during the PhD of H. Haidar, a former PhD student of the team [80].

#### 6.1.1.3. Aggregation models in the chemostat

**Participants:** Radhouene Fekih-Salem, Jérôme Harmand, Claude Lobry, Alain Rapaport, Tewfik Sari.

We have studied a model of the chemostat where the species are present in two forms, isolated and aggregated individuals, such as attached bacteria in biofilm or bacteria in flocks. We show that our general model contains a lot of models that were previously considered in the literature. Assuming that flocculation and deflocculation dynamics are fast compared to the growth of the species, we construct a reduced chemostat-like model in which both the growth functions and the apparent dilution rate depend on the density of the species. We also show that such a model involving monotonic growth rates may exhibit bi-stability, while it may occur in the classical chemostat model, but when the growth rate is non monotonic [21], [54]. This work is part of the PhD of R. Fekih-Salem co-supervised by A. Rapaport and T. Sari.

This research subject has been mainly motivated by the DISCO project (see Section 7.3 ).

#### 6.1.1.4. Overyielding in continuous bioprocesses

**Participants:** Denis Dochain, Alain Rapaport.

We have shown that for certain configurations of two chemostats fed in parallel, the presence of two different species in each tank can improve the yield of the whole process, compared to the same configuration having the same species in each volume. This leads to a (so-called) “transgressive over-yielding” due to spatialization [35].

This work has been achieved during the stay of Prof. P. de Leenheer (Univ. Florida).

### 6.1.2. Measuring taxonomic diversity of microbial communities

**Participant:** Bart Haegeman.

Diversity is considered to be a main determinant of the behavior of microbial communities. However, measuring microbial diversity is challenging. Although metagenomic techniques allow us to sample microbial communities at unprecedented depths, the disparity between community (e.g., $10^{15}$ organisms) and sample (e.g., $10^{5}$ organisms) remains large. We have studied what the diversity observed in a sample tells us about the real diversity of the community.

For a given empirical sample the aim is to construct the community from which this sample was taken. It turns out that a large set of community structures are consistent with the sample data. Some diversity metrics vary widely over this set of consistent communities, and are therefore difficult to infer from the sample data. Other diversity metrics are approximately constant over the set of consistent communities, and are therefore much easier to infer from the sample data.

The analysis of the set of consistent communities has yielded the following insights. First, it is impossible to robustly estimate the number of species from sample data. This is easy to understand. Microbial communities typically contain a large number of rare species, and these rare species are unlikely to be present in the sample. Hence, sample data are lacking crucial information to estimate species richness. Second, other diversity metrics, in particular Shannon and Simpson diversity, can be robustly estimated from sample data. We have constructed lower and upper estimates for a general class of diversity metrics, and we have shown that the difference between the extremal estimators, that is, the estimation uncertainty, is small for Shannon and Simpson diversity.

### 6.1.3. A theory of genetic diversity within bacterial species

**Participant:** Bart Haegeman.

With the wide availability of DNA sequencing, microbiologists are now able to rapidly sequence entire bacterial genomes. Comparison of these genomes has revealed a large genetic diversity within bacterial species. For example, one genome of the bacteria *E. coli* has about 4000 different genes, but a set of 10 genomes of *E. coli* has typically over 10000 different genes. Some of these genes are shared by all or almost all of the genomes, but many other genes are only present in one or a few of the genomes. This observation has important implications for the definition of bacterial species and for the description of the functional characteristics of bacteria.

In [23] we propose a theory for the frequency distribution of genes within a set of genomes. The model describes the genetic diversity as a balance between two forces. Demographic processes such as division and death tend to reduce the genetic diversity; horizontal gene transfer from other species can increase the genetic diversity. Our model predicts that the gene frequency distribution is U-shaped, meaning that there are a large number of genes present in only a few genomes, a small number of genes present in about half of the genomes, and a large number of genes present in almost all genomes. This prediction is consistent with the gene frequency distributions of six bacterial species we have analyzed (about 100 sequenced genomes in total). Importantly, the model does not assume any functional difference between the genes, that is, genes are considered to be selectively neutral. By showing that empirical gene frequency distributions can be reproduced by a neutral genome model, we contend that the frequency of a gene should not be interpreted as an indication of its essentiality or importance.

### 6.1.4. Individual-based modelling for bacterial ecosystems

**Participants:** Fabien Campillo, Chloé Deygout, Coralie Fritsch, Jérôme Harmand, Marc Joannides, Claude Lobry.

In terms of computational modelling of ecosystems, individual-based models (IBMs) are an interesting path to explore. We can outline two types of IBMs. On the one hand "detailed IBM" attempt to integrate in an ad-hoc way all the knowledge available about an ecosystem. On the other hand, "simplified IBM" are limited to one or several mechanisms to simplify the analysis. The former may be more realistic but are often difficult to analyze. Although the latter are too simplistic in realistic situations they lend themselves to the analysis and numerical analysis. We focus on the latter.

The IBMs offer an interdisciplinary language between biologists, biotechnologists, mathematicians, and computer scientists, to develop models in the form of relatively simple rules. In the case of simplified IBMs it is possible to translate these rules in the form of a branching Markov process with values in a space of measures. Using scaling methods, the IBMs can be approximated by integro-differential equations; using model simplification methods IBMs can be reduced to stochastic or ordinary differential equations. The mathematical interpretation of the IBMs and their analysis is relatively recent and still very few studies exist [78]. The numerical analysis of these models is yet to be built. Under certain conditions, IBMs themselves can be simulated through adapted Monte Carlo procedures.

The MODEMIC project-team develops many studies in the field of IBMs. The first is part of the ANR MODECOL on the modelling of clonal plant growth (see Section 7.4 ); the second is part of the ANR DISCO on modelling of biofilms (see Section 7.3 ); the third is also part of the ANR DISCO is dedicated to the modelling of biofilms in plug-flow reactors (see Section 6.2.2 ); the last one is the ongoing thesis of Coralie Fritsch at the École Doctorale I2E of the University of Montpellier 2, under the supervision of Fabien Campillo and Jérôme Harmand. The thesis aims at developing and analyzing "simple" individual-based microbial ecosystems models.

In all cases, we aim at developing the Monte Carlo simulation of the IBM as well as analyzing their links with integro-differential models. We also seek to make connections with non-IBM models proposed in Section 6.1.5 .

### 6.1.5. Stochastic/discrete and stochastic/continuous modelling for biotechnology and population dynamics

**Participants:** Fabien Campillo, Marc Joannides, Claude Lobry.

In [14], we consider a stochastic model of the two-dimensional chemostat as a diffusion process for the concentration of substrate and the concentration of biomass. The model allows for the washout phenomenon: the disappearance of the biomass inside the chemostat. We establish the Fokker-Planck equation associated with this diffusion process, in particular we describe the boundary conditions that modelize the washout. We propose an adapted finite difference scheme for the approximation of the solution of the Fokker-Planck equation.

In [15], we consider a hybrid version of the classical predator-prey differential equation model. The proposed model is hybrid: continuous/discrete and deterministic/stochastic. It contains a parameter $\omega$ which represents the number of individuals for one unit of prey – if $x$ denotes the quantity of prey in the differential equation model $x = 1$ means that there are $\omega$ individuals in the discrete model – is derived from the classical birth and death process. It is shown by the mean of simulations and explained by a mathematical analysis based on results in singular perturbation theory (the so called theory of Canards) that qualitative properties of the model like persistence or extinction are dramatically sensitive to $\omega$. This means that we must be very cautious when we use continuous variables in place of jump processes in dynamic population.

### 6.1.6. Optimal control of continuous bioprocesses

**Participants:** Walid Bouhafs, Amel Ghouali, Jérôme Harmand, Alain Rapaport.

In continuous bioprocesses, a usual objective is to stabilize the output of the bioreactors about a desired steady state (in wastewater industry, this value is typically chosen under the norm of authorized discharge). It happens more and more frequently that transient trajectories are expected also to maximize a product of interest.

We have begun to study the maximization of the gaseous production of methane in anaerobic processes over a given period of time on specific problems. For the moment we have proven that the optimal trajectory consists in approaching a unique singular arc as fast as possible when only one limiting substrate has to be converted, but the problem is still open when involving several substrates [39]. Another problem arises for alternating aerobic-anoxic systems. Revisiting the results obtained several years ago within the framework of Djalel Mazouni's thesis, we aims at proposing optimal time control policies for sequencing batch reactors in which simultaneous nitrification and denitrification take place. The solution for the original problem is a difficult task but we have already proposed a solution for a slightly modified model [30]. These last results have been obtained within the framework of the PhD thesis of W. Bouhafs.

Reference points in batch processes can be mimicked by a series of continuously stirred bioreactors in series at steady state (see applications 6.2.4 and 7.1 ). We study the minimal time problem to drive the nutrients concentrations of a cascade of chemostats. The control variable is the dilution rates of each tank, under the constraint that each dilution rate is bounded by the one of the previous tank, that makes the system not locally controllable. For the particular case of two tanks with total mass at steady state, the planar feedback synthesis has been found but the problem is still under investigation for the general case.

One important issue in bioprocesses is to find optimal feedback control laws in order to steer a system describing a perfectly-mixed bioreactor to a given target value in a minimal amount of time. Finding adequate feeding strategies can significantly improve the performance of the system. A typical target (for fed-batch bioreactors) is to consider the substrate concentration to be lower than a given reference value at the end of the process. Other criterium can be studied such as the maximization of the production of biomass in a given time period. Singular strategies (in reference to the theory of singular arcs in optimal control theory) can be defined in this context and are natural due to the constraints on the system. One objective of our research is to characterize optimal feedback control laws using mathematical tools from optimal control laws, and also to develop numerical methods that can handle problems with a large number of parameters.

### 6.1.7. Modelling the functioning of soil ecosystems

**Participants:** Céline Casenave, Jérôme Harmand, Alain Rapaport.

In ecology, one of the important challenges is the understanding of the biodiversity observed in the natural ecosystems. Several models have been proposed to explain this biodiversity, and in particular the coexistence of different species. In these models, it is often assumed that, when they die, the micro-organisms are directly converted in an assimilable resource, that is a resource that alive organisms can consume. However, we know that it is not the case in reality. Indeed, the organic matter stemmed from the dead organisms has to be transformed in assimilable resource before it can be consumed. This transformation is performed by some micro-organisms which have this special ability.

We have proposed a new model of soil ecosystems, of chemostat type. This model is rather simple, but also original because it takes into account several mechanisms:

- the growth, the mortality and the respiration,
- the ability of some organisms to transform the non assimilable resources in assimilable ones,
- the advantage that an organism can gain from this ability of transformation.

For the moment, we have considered the case where only one or two types of organisms are present. The model is finally composed of 3 (or 4) nonlinear ordinary differential equations: one per type of organisms, one for the assimilable resource and one for the non assimilable one. The study of the equilibrium points has first shown the possibility of coexistence, at equilibrium, of the two organisms. Then, in numerical simulations, we have observed the possible existence of limit cycles, which can also explain the coexistence of organisms observed in the nature.

This problem is still under study; a working group (in particular with researchers of the UMR Eco & Sols, conducted by B. Jaillard) meet regularly to discuss about the problems of modelling in ecology.

We have also investigated the *sampling effect* that occurs when micro-biologists select randomly species in a natural ecosystem for studying reconstituted ecosystems in a controlled environment. We have proposed a very simple probabilistic model, that shows that observing average increases or decreases on the performances of these reconstitute ecosystems when modifying the size of the sampling, allows to infer kinds and proportions of the interactions among species present in the original ecosystem [65]. This research is conducted with the UMR Eco & Sols, Montpellier, (B. Jaillard) and the UMR BIOEMCO, Grignon (N. Nunan).

### 6.1.8. Equivalence between models of fractured porous media

**Participants:** Céline Casenave, Jérôme Harmand, Alain Rapaport, Alejandro Rojas-Palma.

In geosciences, models of fractured porous media are often described as a *mobile* zone driven by advection, and one or several *immobile zones* directly or indirectly connected to the mobile zone by diffusion terms. We believe that these models are also relevant to describe flows in soil or in porous media such as biofilms. They are very close from the spatial representations used in Section 6.2.6 . We have shown, using a transfer function approach, that two schemes often used in the literature: the MINC (Multiple INteractive Continua) where diffusive compartments are connected in series, and the MRMT (Multiple Rate Mass Transfer) where diffusive compartments are connected in star around the mobile zone, are equivalent input-output representations [56], and providing formulas (up to three compartments) to pass from one representation to another. This result means that one can simply choose the most convenient approach when dealing with control or optimization without any loss of generality. We are currently working on the general case of $n$ compartments with $n$ larger than three, and looking for equivalent classes of configurations that could be half way between MINC and MRMT and fit better the spatial representations of real world.

This work is performed with the UMR GéoSciences Rennes (J.R. de Dreuzy), and has led to the internship of a Chilean MsC student (A. Rojas-Palma).

### 6.1.9. Non-linear filtering for the chemostat

**Participants:** Boumediene Benyahia, Amine Boutoub, Fabien Campillo, Jérôme Harmand.

Monte Carlo-based inference methods like particle filtering are bound to develop in the context of biotechnology. In contrast with the classical observer approach, inference through Monte Carlo methods can handle measurements in discrete time in low frequency and with low signal-to- noise ratio. Based on the stochastic modeling of the chemostat, these approaches may also be used for model selection and hypothesis testing.

In a preliminary work [28] we consider the bootstrap particle filter applied to a 2-dimensional chemostat model. The internship of Amine Boutoub dedicated to the study of particle filtering for more realistic chemostat models has recently started.

### 6.1.10. Functional identification of growth functions in the chemostat

**Participant:** Alain Rapaport.

We have proposed an adaptive control law that allows one to identify unstable steady states of the open-loop system in the single-species chemostat model without the knowledge of the growth function. We have then shown how to use a continuation technique to reconstruct the whole graph of the growth function, providing a new method for identifying non-monotonic growths [42], [27]. Two variants, in continuous and discrete time, have been studied. An analysis of the case of two species in competition has shown the ability of the method to detect a mixed culture for which dominance depends on the dilution rate, due to a property of stability loss in slow-fast dynamics. This method leaves open future extensions for extremum seeking problems.

This work has been conducted in cooperation with Universities of Exeter (J. Sieber) and Plymouth (S. Rodrigues), and the EPI SISYPHE (M. Desroches).

### 6.1.11. Diffusive representation of integro-differential models

**Participant:** Céline Casenave.

This work is done in collaboration with Emmanuel Montseny (LAAS/CNRS), Gérard Montseny (LAAS/CNRS), and Christophe Prieur (LIAFA/CNRS).

In lots of dynamic systems of Physics or others scientific fields such as Biology (Volterra models), dynamic integral operators, often of convolution type, are involved. Problems relating to integro-differential models are often difficult to solve, especially because these models are not time-local. In this context, the methodology called "diffusive representation" presents some interests: an integral operator is represented by its gamma-symbol, directly deduced from its transfer function. It can be formulated by means of a state realization whose dimension is numerically reasonable whatever the size of the system may be. In addition to this interesting practical side, the diffusive representation offers a unified mathematical framework, well adapted to analysis of integral convolution operators.

Several dynamic problems can be tackled in an original and quite simple way by using the diffusive representation. In fact, all the operatorial problems of modeling, simulation, control, model identification, model reduction, etc. can be formulated in such a way that the object of the problem is the gamma-symbol of the operator solution.

Several problems are under study:

- the identification of integro-differential models [66],
- the controllability of some SISO Volterra models [63],
- the simulation and the analysis of a model of porous media[64].

These works follow up on the ones developed during the PhD thesis of Céline Casenave, which deals with the problem of the operator inversion for the application to non local dynamic problems.

## 6.2. Applications

### *6.2.1. Modelling and control of Anaerobic Digestion processes*

**Participants:** Boumediene Benyahia, Amine Charfi, Radhouene Fekih-Salem, Jérôme Harmand, Guilherme Pimentel, Tewfik Sari.

We consider the AM2 or AMOCO model developed in [72] and extend both the model in itself and its analysis to the following cases:

- Depending on the AM2 model parameters, the steady states were analytically characterized and their stability were analyzed [12]. Following this study, it was shown that the overloading tolerance, a parameter proposed in [81] to on-line monitoring anaerobic processes, may be not adapted under certain operating conditions and even lead to bad operating decisions.
- Within the framework of the PhD theses of Amine Charfi and Boumediene Benyahia, we have included the fouling dynamics of membranes into the AM2 and we have analyzed the resulting model (called the AM2b) [16], [29].
- We actually work towards two directions: (i) we are extending these results in including into the AM2 an additional process, *i. e.* the hydrolysis step in order to study bioprocesses treating solid waste (the resulting model being called the AM3) [36], [37]; (ii) we try to find links between complex models such as the ADM1 model and simple models such as the AM2b or the AM3 [40].

Apart from this work on the modelling of anaerobic digesters and membrane bioreactors, we have developed a number of specific simple models for control design accounting for the coupling of such processes with membrane modules in the chemostat (PhD thesis of G. Pimentel). This work aims at contributing to the efficient treatment of wastewaters produced in fish production farms. The work of G. Pimentel aims at studying the coupling of simple fouling models with the model of the chemostat in order to propose new simple models for control design.

### *6.2.2. Hybrid modelling of biofilms in plug-flow reactors*

**Participants:** Fabien Campillo, Chloé Deygout, Annick Lesne, Alain Rapaport.

We have proposed a multi-scaled modelling that combines three scales: a microscopic one for the individual bacteria, a mesoscopic or "coarse-grained" one that homogenises at an intermediate scale the quantities relevant to the attachment/detachment process, and a macroscopic one in terms of substrate concentration.

Such a “hybrid” approach allows for modelling and understanding in plug-flow reactors the interplay between

- the formation of the biofilm at a microscopic scale, that starts from a small number of bacteria (thus a stochastic individual based description),
- the limitation of the biofilm, due the carrying capacity of the wall attachment, at a mesoscopic scale,
- the consumption of nutrient along the flow at a macroscopic level, as a solution of a coupled transport-reaction partial differential equation.

The numerical computation of such a model requires a software architecture that allows the simultaneous simulation of stochastic events at the bacteria scale and the continuous evolution (in space and time) of the substrate density.

This work has been conducted within the DISCO project (see Section 7.3 ) and the postdoctoral stay of C. Deygout hired by the project, in close collaboration with A. Lesne (LPTMC, Univ. Paris VI). A first paper on the simulation model has been published [17].

Within the DISCO project, experiments on real tubular plug-flow reactors have been simultaneously driven at IRSTEA Antony with the perspective of comparison with numerical simulations.

The multi-species case with different bacteria specialized in different environments (poor or rich in nutrient) is a work in progress.

### *6.2.3. Individual-based models for the bacterial degradation of the cellulose*

**Participants:** Fabien Campillo, Chloé Deygout.

We propose an individual-based model for the degradation of one cellulose bead (dozens of micrometers in diameter) by cellulolytic bacteria. This model accounts for biofilm formation with minimal hypotheses: soluble substrate diffusion combined with bacterial chemotaxis-like movement in the liquid phase, lack of bacterial movement in the solid phase. The IBM results are qualitatively different from the main macroscopic degradation models previously used for cellulose degradation. It suggests that random and discrete processes could significantly impact the cellulose degradation dynamics by their effect on the spatial structuration of the colonized cellulose particles [44].

### *6.2.4. Modelling and control of cascade biosystems to mimic batch wine making processes*

**Participants:** Térence Bayen, Céline Casenave, Jérôme Harmand, Alain Rapaport, Matthieu Sebbah.

An experimental setup of four tanks connected in series has been designed by the research unit SPO (Montpellier) for studying four physiological stages of yeast as steady state. The manipulated variables are the flow rates $Q_i$ of each tank with the constraint $Q_i \geq Q_{i-1} \geq 0$, and the objective is to reach simultaneously four set-points in the four tanks. We are studying two kinds of control strategies:

- a linearizing feedback law that drives exponentially the dynamics to the target. This is not the fastest strategy but is has good robustness properties. Nevertheless, the inputs constraint imposes to use saturation functions that provide satisfactory convergence in simulations but that is hard to prove mathematically.
- a minimal time feedback. Due to lack of local controllability imposed by the constraint on the inputs, the optimal synthesis is not smooth with the presence of “barriers”. The input constraint $Q_i \geq Q_{i-1} \geq 0$ is unusual in optimal control problems that are linear w.r.t. to the control. The optimality of candidate singular arcs is still open for this problem.

This summer, some experiments have been made to test the first feedback law on the experimental setup. The control law seems to perform work, but other experiments should be made next year with more reliable input flow pumps.

This work was conducted as a part of the European CAFE project (Computer-Aided Food processes for control Engineering) described in Section 7.1 .

### 6.2.5. Modelling and control of an ice cream crystallization process

**Participants:** Céline Casenave, Denis Dochain.

In the ice cream industry, the type of final desired product (large cartons or ice creams on a stick) determine the viscosity at which the ice cream has to be produced. The control the viscosity of the ice cream at the outlet of a continuous crystallizer is therefore an important challenge. The problem has been studied in two steps. First, we have completed and validated on experimental data the reduced order model of the system. This model has been obtained by application of the method of moments on a population balance equation describing the evolution of the crystal size distribution. Then, we have proposed a nonlinear control strategy to control of the viscosity of the ice cream with the temperature of the refrigerant fluid of the crystallizer. This control strategy is based on a linearizing control law coupled with a Smith predictor to account for the measurement delay. The control has been validated on an experimental pilot plant located at IRSTEA (Antony, France).

This work was conducted as a part of the European CAFE project (Computer-Aided Food processes for control Engineering) described in Section 7.1 .

### 6.2.6. Bioremediation of natural resources

**Participants:** Sébastien Barbier, Jérôme Harmand, Alain Rapaport, Antoine Rousseau.

The objective of this work is to provide efficient strategies for the bioremediation of natural water resources. The originality of the approach is to couple minimal time strategies that are determined on a simplified model with a faithful numerical model for the hydrodynamics. This work has been carried out in close cooperation with A. Rousseau. Based on a previous paper that deals with an implicit representation of the spatial inhomogeneity of the resource with a small number of homogeneous compartments (with a system of ODEs), we have implemented a coupled ODE-PDE system that accounts for the spatial non-homogeneity of pollution in natural resources. The main idea is to implement a Navier-Stokes model in the resource (such as a lake), with boundary conditions that correspond to the output feedback that has been determined to be optimal for the simple ODEs model of a (small) bioreactor. A first mathematical model has been introduced and numerical simulations have been performed in academic situations. During the internship of S. Barbier (co-advised by A. Rousseau and A. Rapaport) we built a reduced model that approximates the reference PDE model thanks to a set of ODEs with parameters. Numerical optimization is performed on these parameters in order to better fit the reference model. This will lead to a publication.

The study of the minimal time strategies on the system of ODEs has been mainly achieved in cooperation with Chilean researchers (P. Gajardo, Universidad Tecnica Federico Santa Maria, and H. Ramirez, Centro de Modelamiento Matemático) and a Chilean PhD student (V. Riquelme, Depto. Ingenieria Matematica, Universidad de Chile) within the associated team DYMECOS [57].

### 6.2.7. Modelling and simulating terrestrial plant ecological dynamics

**Participant:** Fabien Campillo.

This study is part of the ANR Syscomm MODECOL that is done in collaboration particularly with the University of Rennes I, the University of La Rochelle and Inria. The first semester of 2012 was the last part of the project. We propose a very original individual-based model for clonal plant dynamics in continuous time and space that focuses on the effects of the network structure of the plants on the reproductive strategy of ramets. The model is coupled with a PDE dynamics for resources. The basic structure of the IBM encompass a population of “ramets” (the individuals) connected by “stolons or rhizomes” (the network) [13], [22]. See http://www-sop.inria.fr/members/Fabien.Campillo/software/ibm-clonal/ for more details.

### 6.2.8. Modelling and inferring agricultural dynamics

**Participants:** Fabien Campillo, Angelo Raherinirina.

The International Laboratory LIRMA supports this work that is done in collaboration with the University of Fianarantsoa in Madagascar and with Dominique Hervé (IRD, Fianarantsoa, Madagascar). The aim is to study the dynamics of agricultural plots on the edge of primary forest. In [32] a simple Markov model has been successfully confronted to a first data set with the help of maximum likelihood and Bayesian approaches. On a new data set developed by IRD, the Markov hypothesis has been rejected and we proposed to use semi-Makov models: for this new dataset the law of the sojourn time on certain states will depend on the next state visited, which is incompatible with the Markov hypothesis and which is consistent with the semi-Markov hypothesis.

# NUMED Project-Team

# 6. New Results

## 6.1. New result 1

Numed has developped a general strategy and generic softwares (to be released soon) to allow populational parametrization on complex models like PDEs.

# REO Project-Team

# 6. New Results

## 6.1. Mathematical and numerical analysis of fluid-structure interaction problems

**Participants:** Cristóbal Bertoglio Beltran, Muriel Boulakia, Miguel Ángel Fernández Varela, Sébastien Martin, Jean-Frédéric Gerbeau, Jimmy Mullaert, Marina Vidrascu.

- In [26], we study a three-dimensional fluid-structure interaction problem. The motion of the fluid is modeled by the Navier-Stokes equations and we consider for the elastic structure a finite-dimensional approximation of the equation of linear elasticity. The time variation of the fluid domain is not known a priori, so we deal with a free boundary value problem. Our main result yields the local in time existence and uniqueness of strong solutions for this system.
- In [31], a robust finite volume method for the solution of high-speed compressible flows in multi-material domains involving arbitrary equations of state and large density jumps is presented. One of the main contributions of this paper is a tabulation method based on a sparsegrid approximation to solve very efficiently two-phase Riemann problems for arbitrary equations of state. The proposed computational method is illustrated with the three-dimensional simulation of the dynamics of an underwater explosion bubble.
- In [52] we analyze the performances of several Luenberger observers to estimate the state of a fluid-structure interaction model for hemodynamics, when the measurements are assumed to be restricted to displacements or velocities in the solid. The present framework establishes that these methods are very attractive strategies (compared, e.g., to classical variational techniques) to perform state estimation.
- In [51] we analyze two 3D-0D coupling approaches in which a fractional-step projection scheme is used in the fluid. We introduce and analyze an implicitly 3D-0D coupled formulation with enhanced stability properties and which requires a negligible additional computational cost. The theoretical stability results are confirmed by meaningful numerical experiments in patient specific geometries coming from medical imaging.
- In [55] we introduce a class of explicit Robin-Neumann schemes for the explicit coupling of a general thin-structure (e.g., viscoelastic and non-linear) with an incompressible fluid. These methods generalize the displacement correction schemes introduced in [32]. A priori stability and convergence error estimates show that optimal first-order accuracy can be achieved with appropriate extrapolation and without compromising stability. A deep numerical study confirms the theoretical findings.
- In [64] we present two-dimensional simulations of chemotactic self-propelled bacteria swimming in a viscous fluid. Self-propulsion is modelled by a couple of forces of same intensity and opposite direction applied on the rigid bacterial body and on an associated region in the fluid representing the flagellar bundle. The orientations of the individual bacteria are subjected to random changes, with a frequency that depends on the surrounding oxygen concentration, in order to favor the direction of the concentration gradient.
- In [40] we propose a method of modeling sail structures which captures the wrinkling behavior of such structures. The method is validated through experimental and analytical test cases, particularly in terms of wrinkling prediction. An enhanced wrinkling index is proposed as a valuable measure characterizing the global wrinkling development on the deformed structure. The method is based on a pseudo-dynamic finite element procedure involving non-linear MITC shell elements. The major advantage compared to membrane models generally used for this type of analysis is that no ad hoc

wrinkling model is required to control the stability of the structure. We demonstrate our approach to analyse the behavior of various structures with spherical and cylindrical shapes, characteristic of downwind sails over a rather wide range of shape constitutive parameters. In all cases convergence is reached and the overall flying shape is most adequately represented, which shows that our approach is a most valuable alternative to standard techniques to provide deeper insight into the physical behaviour. Limitations appear only in some very special instances in which local wrinkling-related instabilities are extremely high and would require specific additional treatments.

## 6.2. Numerical methods for fluid mechanics and application to blood flows

**Participants:** Grégory Arbia, Jean-Frédéric Gerbeau, Sébastien Martin, Saverio Smaldone, Marc Thiriet, Irène Vignon-Clementel.

- In [18], a procedure for modeling the heart valves is presented. Instead of modeling complete leaflet motion, leaflets are modeled in open and closed configurations. This method enables significant computational savings compared to complete fluid-structure interaction and contact modeling, while maintaining realistic three-dimensional velocity and pressure distributions near the valve, which is not possible from lumped parameter modeling. To illustrate the versatility of the model, realistic and patient-specific simulations are presented, as well as comparison with complete fluid-structure interaction simulation.
- [37] paves the way for a complete patient-specific fluid-structure vascular modeling in which all types of available measurements could be used to estimate uncertain parameters of biophysical and clinical relevance. We propose a complete methodological chain for the identification of the parameters involved in a model for external tissue support of blood vessels, using patient image data. We demonstrate the use of this framework in a realistic application case involving hemodynamics in the thoracic aorta. The estimation of the boundary support parameters proves successful, in particular in that direct modeling simulations based on the estimated parameters are more accurate than with a previous manual expert calibration.
- In [27] we study the image-based blood flow in the first generation of the pulmonary arterial tree. This patient-specific study is aimed at assessing effects of lung deformation and vascular resistance on the pulmonary blood flow, especially during the acute phase of a pneumothorax and after recovery. Arterial geometry was extracted up to the fifth generation from computed tomography images, and reconstructed. An unsteady laminar flow with a given set of resistances at outlets was modeled. Adaptation is set to match perfusion to ventilation.
- In [44], [36] we study the reciprocal effect of blood circulation and high-intensity focused ultrasound on the temperature field in the liver. High-intensity focused ultrasound (HIFU) is used as a thermal ablation process to eliminate tumors in different body's organs. Blood flow has a cooling effect. Conversely, ultrasounds are responsible for acoustic streaming. A three-dimensional acoustics-thermal-fluid coupling model is carried out to compute the temperature field a given hepatic cancerous region.
- The use of elaborate closed-loop lumped parameter network (LPN) models of the heart and the circulatory system as boundary conditions for 3D simulations can provide valuable global dynamic information, particularly for patient specific simulations. In [30], we have developed and tested a numerical method to couple a 3D Navier-Stokes finite-element formulation and a reduced model of the rest of the circulation, keeping the coupling robust but modular. For Neumann boundaries, implicit, semi-implicit, and explicit quasi-Newton formulations are compared within the time-implicit coupling scheme. The requirements for coupling Dirichlet boundary conditions are also discussed and compared to that of the Neumann coupled boundaries. Both these works were key for applications where blood flows in different directions during the cardiac cycle and where coupling with the rest of the circulation is instrumental (see the shunt optimization application [29]).

- Boundary conditions in patient-specific blood flow simulations is key because pressure and flow within the modeled domain are driven by the interplay between the local 3D hemodynamics and the rest of the circulation. However, these boundary conditions are rarely the measured variables. In [45], we showed how one can go from patient-specific clinical data (MRI and catheterization) to simulation input parameters, including modeling assumptions and the impact of both on simulation results. We explained how Windkessel models and more involved LPN can be calibrated.

  In [34], we developed two multi-scale models, each including the 3D model of the surgical junction constructed from MRI, and a closed-loop LPN derived from pre-operative data obtained from two patients prior to Stage 2 Fontan palliation of single ventricle congenital heart disease. "Virtual" surgeries were performed and a corresponding multi-scale simulation predicted the patient's post-operative hemodynamic conditions, tested under different physiological conditions. The impact of the surgical junction geometry on the global circulation was contrasted with variations of key physiological parameters.
- In [19], a similar 3D multiscale model was used but for the Stage 3 Fontan palliation. Several studies have been done to optimize the geometry of the surgical connection, to minimizing energy losses and improving surgical outcomes, but usually without taking into account respiration or exercise. A respiration model that modulates the extravascular pressures in the thoracic and abdominal cavities was implemented. Results showed that the preoperative model is able to realistically capture cardiac and respiratory oscillations compared to the venous Doppler velocity tracings. Three virtual surgical alternatives were coupled to the LPN and then investigated under rest and exercise conditions.
- In [29], such a 3D-closed loop LPN model was integrated with an automated derivative-free optimization algorithm in an idealized systemic-to-pulmonary shunt anatomy (Stage 1 Fontan palliation). The goal was to optimize shunt geometries. Clinicians selected three objective functions to be maximized: (1) systemic, (2) coronary, and (3) combined systemic and coronary oxygen. Results showed the geometries associated with the favored delivery, the origin of coronary artery flow being driven by the shunt position as well. The results made only sense when the 3D domain was connected to a closed-loop model of the circulation.
- A novel Y-shaped baffle was proposed for the Stage 3 Fontan operation achieving overall superior hemodynamic performance compared with traditional designs. Previously, we investigated if and how the inferior vena cava flow (which contains an important biological hepatic factor) could be best distributed among both lungs. In [41] we proposed a multi-step method for patient-specific optimization of such surgeries to study the effects of boundary conditions and geometry on hepatic factor distribution (HFD). The resulting optimal Y-graft geometry largely depended on the patient left/right pulmonary flow split. Unequal branch size and constrained optimization on energy efficiency were explored. Two patient-specific examples showed that optimization-derived Y-grafts effectively improved HFD.

## 6.3. Numerical methods for cardiac electrophysiology

**Participants:** Muriel Boulakia, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Vincent Martin, Elisa Schenone.

- In [62], we propose a surface-based electrophysiology formulation, motivated by the modeling of thin structures such as cardiac atria, which greatly reduces the size of the computational models. Our model is specifically devised to retain the key features associated with the anisotropy in the diffusion effects induced by the fiber architecture, with rapid variations across the thickness which cannot be adequately represented by naive averaging strategies. The model relies on a detailed asymptotic analysis in which we identify a limit model and establish strong convergence results. We also provide detailed numerical assessments which confirm an excellent accuracy of the surface-based model – compared with the reference 3D model – including in the representation of a complex phenomenon, namely, spiral waves.

## 6.4. Lung and respiration modeling

**Participants:** Laurent Boudin, Paul Cazeaux, Bérénice Grec, Muriel Boulakia, Anne-Claire Egloffe, Benoit Fabreges, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Stéphane Liwarek, Sébastien Martin, Ayman Moussa.

- [59], [60]:We are concerned here with identifiability, stability properties and estimates for the inverse problem of identifying a Robin coefficient on some non accessible part of the boundary from available data on the other part of boundary corresponding to solutions of the Stokes equations. In [59], we first consider a steady state two-dimensional Stokes problem and study the identifiability of Robin coefficient and then we establish a stability estimate of logarithm type using a global Carleman inequality. We then consider the unsteady problem. In [60]:We prove hölderian and logarithmic stability estimates associated to the unique continuation property for the Stokes system. The proof of these results is based on local Carleman inequalities. In the second part, these estimates on the fluid velocity and on the fluid pressure are applied to solve the inverse problem of identifying a Robin coefficient. For this identification parameter problem, we obtain a logarithmic stability estimate under the assumption that the velocity of a given reference solution stays far from 0 on a part of the boundary where Robin conditions are prescribed.
- In [61] we are interested in the mathematical modeling of the propagation of sound waves in the lung parenchyma, which is a foam–like elastic material containing millions of air– filled alveoli. In this study, the parenchyma is governed by the linearized elasticity equations and the air by the acoustic wave equations. The geometric arrangement of the alveoli is assumed to be periodic with a small period $\varepsilon > 0$. We consider the time–harmonic regime forced by vibrations induced by volumic forces. We use the two–scale convergence theory to study the asymptotic behavior as $\varepsilon$ goes to zero and prove the convergence of the solutions of the coupled fluid–structure problem to the solution of a linear–elasticity boundary value problem.
- In [53] we develop and study numerically a model to describe some aspects of sound propagation in the human lung, considered as a deformable and viscoelastic porous medium (the parenchyma) with millions of alveoli filled with air. Transmission of sound through the lung above 1 kHz is known to be highly frequency–dependent. We pursue the key idea that the viscoelastic parenchyma structure is highly heterogeneous on the small scale $\varepsilon$ and use two–scale homogenization techniques to derive effective acoustic equations for asymptotically small $\varepsilon$. This process turns out to introduce new memory effects. The effective material parameters are determined from the solution of frequency–dependent micro–structure cell problems. We propose a numerical approach to investigate the sound propagation in the homogenized parenchyma using a Discontinuous Galerkin formulation. Numerical examples are presented.
- In [22], we consider the Maxwell-Stefan model of diffusion previously introduced. We provide a qualitative and quantitative mathematical and basic numerical analysis of the model.
- In [65] we propose an integrated model for oxygen transfer into the blood, coupled with a lumped mechanical model for the ventilation process. We aim at investigating oxygen transfer into the blood at rest or exercise. The first task consists in describing nonlinear effects of the oxygen transfer under normal conditions. We also include the possible diffusion limitation in oxygen transfer observed in extreme regimes involving parameters such as alveolar and venous blood oxygen partial pressures, capillary volume, diffusing capacity of the membrane, oxygen binding by hemoglobin and transit time of the red blood cells in the capillaries. The second task consists in discussing the oxygen concentration heterogeneity along the path length in the acinus.
- In [43] we presented preliminary work on a multiscale 3D-0D airflow model to study differences between healthy and emphysema rats. The 0D model parameters were estimated from experimental data. 3D Navier-Stokes simulations were performed in healthy lungs, and in homogenous and heterogeneous emphysema lungs.

## 6.5. Miscellaneous

**Participants:** Laurent Boudin, Jean-Frédéric Gerbeau, Damiano Lombardi, Sébastien Martin, Marina Vidrascu, Irène Vignon-Clementel.

- In [56], a reduced-order model algorithm, based on approximations of Lax pairs, is proposed to solve nonlinear evolution partial differential equations. Contrary to other reduced-order methods, like Proper Orthogonal Decomposition, the space where the solution is searched for evolves according to a dynamics specific to the problem. It is therefore well-suited to solving problems with progressive waves or front propagation. Numerical examples are shown for the KdV and FKPP (nonlinear reaction diffusion) equations, in one and two dimensions.
- In [21], we investigate the asymptotic behaviour of the solutions to the non-reactive fully elastic Boltzmann equations for mixtures in the diffusive scaling. We deal with cross sections such as hard spheres or cut-off power law potentials. We use Hilbert expansions near the common thermodynamic equilibrium granted by the H-theorem. The lower-order non trivial equality obtained from the Boltzmann equations leads to a linear functional equation in the velocity variable which is solved thanks to the Fredholm alternative. Since we consider multicomponent mixtures, the classical techniques introduced by Grad cannot be applied, and we propose a new method to treat the terms involving particles with different masses. The next-order equality in the Hilbert expansion then allows to write the macroscopic continuity equations for each component of the mixture.
- In [58], we discuss some numerical properties of the viscous numerical scheme introduced in [23] to solve the one-dimensional pressureless gases system, and study in particular, from a computational viewpoint, its asymptotic behavior when the viscosity parameter $\varepsilon > 0$ used in the scheme becomes smaller.
- In [33] we study a network-based model for rubber. Since the pioneering work by Treloar, many models based on polymer chain statistics have been proposed to describe rubber elasticity. Recently, Alicandro, Cicalese, and the first author rigorously derived a continuum theory of rubber elasticity from a discrete model by variational convergence. The aim of this paper is twofold. First we further physically motivate this model, and complete the analysis by numerical simulations. Second, in order to compare this model to the literature, we present in a common language two other representative types of models, specify their underlying assumptions, check their mathematical properties, and compare them to Treloar's experiments.
- In [63] our aim is to demonstrate the effectiveness of the matched asymptotic expansion method in obtaining a simplified model for the influence of small identical heterogeneities periodically distributed on an internal surface on the overall response of a linearly elastic body. The results of some numerical experiments corroborate the precise identification of the different steps, in particular of the outer/inner regions with their normalized coordinate systems and the scale separation, leading to the model.
- In cancer modeling, to be able to capture the full in-vivo scale, tumors have to be modeled with continuum models. An important step consists in qualitatively and quantitatively comparing agent-based models (which parameters can generally be identified by experiments in vitro) and continuum models. We derived a first 1D continuum model for tumor growth from the cell based model (Drasdo and Hoehme, 2005): it results in a fluid-type model which capture tumor expansion in both diffusive and compact phenotypes. The tumor expands based on the pressure gradient generated by cell proliferation, the latter being hindered by high density or pressure. In [39] this modeled is analyzed mathematically, showing the existence of traveling waves in the different regimes (with or without internal friction and diffusion due to active movement). In particular the incompressible cells limit is very singular and relates to the Hele-Shaw equation. Numerical results confirm the analysis.

# SISYPHE Project-Team

# 5. New Results

## 5.1. Modeling, observation and control: systems modeled by ordinary differential equations

### *5.1.1. Nonlinear system identification*

**Participants:** Pierre-Alexandre Bliman, Boyi Ni, Michel Sorine, Qinghua Zhang.

In the framework of the joint Franco-Chinese ANR-NSFC EBONSI project, in collaboration with the Laboratory of Industrial Process Monitoring and Optimization of Peking University, and with Centre de Recherche en Automatique de Nancy (CRAN), the topics studied this year on nonlinear system identification are mainly on extended Hammerstein system identification with hysteresis nonlinearity and on continuous time block-oriented nonlinear system identification.

Motivated by the modeling of control valves with significant stiction, we have studied extended Hammerstein systems composed of a hysteresis nonlinearity followed by a linear dynamic subsystem. The joint characterization of the control valve and of the controlled process is formulated as the identification of an extended Hammerstein system. A point-slope based hysteresis model is used to describe the input hysteresis nonlinearity of the control valve. An iterative algorithm is proposed to solve the identification problem. The basic idea is to separate the ascent and descent paths of the input hysteresis nonlinearity subject to oscillatory excitations. Industrial examples are tested to verify the effectiveness of the proposed identification algorithm for characterizing complex behavior of control valve stiction in practice. This work has been presented at the 16th IFAC Symposium on System Identification [66].

A Hammerstein-Wiener system is composed of a dynamic linear subsystem preceded and followed by two static nonlinearities. Typically, the nonlinearities of such a system are caused by actuator and sensor distortions. The identification of such systems with a continuous time model has been studied this year in collaboration with colleagues of CRAN. Based on previously developed simplified refined instrumental variable method, and by making use of an adaptive observer for data filtering, a combined approach, referred to as Kalman pre-filtered instrumental variable based method, is developed. By taking advantages of the two aforementioned methods, the new method is faster and has a naturally stabilizing Kalman filter that does not color white noises. This work has been presented at the 16th IFAC Symposium on System Identification [62].

### *5.1.2. Model-based fault diagnosis*

**Participants:** Abdouramane Moussa Ali, Qinghua Zhang.

The increasing requirements for higher performance, efficiency, reliability and safety of modern engineering systems call for continuous research investigations in the field of fault detection and isolation. This year we have studied algebro-differential systems through an adaptive observer based approach, and linear time varying systems through a Kalman filter based statistical testing approach.

In the framework of the MODIPRO project funded by Paris Region ASTech, the monitoring of the air conditioning system of an aircraft has been studied this year. Part of this system is modeled by nonlinear algebro-differential equations. A method for fault diagnosis of such systems has been developed in our study. Through a particular discretization method and under realistic assumptions, the considered continuous time DAE model is transformed to an explicit state space model in discrete time. An adaptive observer is then applied to the discretized system for monitoring faults possibly affecting the system and represented by changes in model parameters. This work will be presented at the 5th IFAC Symposium on System Structure and Control [61].

While the theory of fault diagnosis has been mostly developed for linear time invariant (LTI) systems, in many industrial applications it is important to take into account the nonlinear behavior of the monitored systems. One possible approach is to linearize a nonlinear system all along its state trajectory, resulting in linear time varying (LTV) or linear parameter varying (LPV) models. In collaboration with Michèle Basseville of IRISA (Institut de Recherche en Informatique et Systèmes Aléatoires), fault diagnosis for stochastic LTV systems has been studied this year. By applying the Kalman filter in a particular manner avoiding the difficulty related to unknown faults possibly affecting the system, the problem of fault diagnosis in a dynamic LTV system is transformed into a hypothesis testing problem in a simple linear regression model. Generalized likelihood ratio (GLR) tests are then applied to the resulting hypothesis testing problem. This work has been presented at the 16th IFAC Symposium on System Identification [67].

## 5.2. Observation, control and traveling waves in systems modeled by partial differential equations

### *5.2.1. Modeling of electric transmission networks*

**Participants:** Mohamed Oumri, Michel Sorine, Qinghua Zhang.

The increasing number and complexity of wired electric networks in modern engineering systems is amplifying the importance of the reliability of electric connections. In the framework of the ANR 0-DEFECT project, we have studied mathematical models of complex electric networks with the aim of designing algorithms for fault diagnosis. The well known Baum-Liu-Tesche (BLT) equation is a powerful model for describing quite general networks and allows to compute the current and voltage waves at the nodes of a network from the specifications of its nodes and connecting cables [63]. This year we have studied the inverse problem: what can we know about the properties of the cables connecting the nodes of a network from experiments made at the nodes of the network? A convenient model for this purpose is formulated with admittance matrices. It is essentially equivalent to the BLT equation, hence can describe quite general networks. The inverse problem is then solved through a decomposition of the admittance matrix of the entire network.

### *5.2.2. Diagnosis of insulator degradation in long electric cables*

**Participants:** Leila Djaziri, Michel Sorine, Qinghua Zhang.

For the diagnosis of insulator degradation in long electric cables, the estimation of the shunt conductance of such cables have been studied, in the framework of the ANR INSCAN project. The shunt conductance of a healthy electric cable is usually very weak. Even when the insulator in the cable is significantly degraded, the shunt conductance can still remain at a quite low level. The main difficulty in this study is due to the fact that the measurements made at the ends of a cable are hardly sensitive to the variations of the shunt conductance. To overcome this difficulty, two methods have been studied. One of them is based on the analysis of the sensitivity of the wave phase shift to the shunt conductance. The efficiency of this method has been demonstrated through extensive tests on cables of SNCF (Société Nationale des Chemins de Fer français). Another method is based on the processing of long time data records. It is designed for the estimation of distributed shunt conductance, in order to detect and to locate inhomogeneous degradation of the insulator. The main idea of this method is to compensate the weak sensitivity of the measurement by long time data records. The results of this method evaluated by numerical simulations have been reported at the 16th IFAC Symposium on System Identification [68].

## 5.3. System theory approach of some quantum systems

**Participants:** Hadis Amini, Zaki Leghtas, Mazyar Mirrahimi, Pierre Rouchon.

Most of this work is done in close collaboration with the Pierre Aigrain laboratory (LPA) at ENS Paris and the Quantronics Laboratory (Qlab) of Michel Devoret and the Rob Schoelkopf Lab at Yale University.

Modern scientific and technologic requirements have led the theoretical and experimental research toward an engineering of quantum systems. The technologies that are proposed or developed include nano-scale electromechanical devices, tools for implementing quantum computation and quantum communication, NMR applications, quantum chemistry synthesis, high-resolution sensors, etc. The recent theoretical and experimental researches have shown that the quantum dynamics can be studied in the framework of the theory of estimation and control of systems, but give place to models that are not completely explored yet.

Our activities lie in the theoretical and experimental interface of this progressing field of research with an accent on the applications in quantum information and computation as well as high-precision metrology. By focusing on two different but similar types of experimental setups, consisting of cavity quantum electrodynamical systems and quantum Josephson circuits, we aim in preparing highly non-classical states of a microwave field and protect these states against decoherence. Two different approaches are considered: 1- real-time measurement, quantum filtering and feedback ; 2- dissipation engineering also called reservoir engineering. Through the first methodology, we try to propose new experimental feedback protocols based on a fast real-time processing of measurement signal, followed by a state estimation applying the filtered signal and finally designing simple feedback laws based on the estimated state. The second methodology consists of designing new quantum circuit schemes that allow to orient the system's coupling to its environment in such a way that evacuates the undesired entropy induced by un-controlled noise sources.

### 5.3.1. Measurement based feedback

We have developed the mathematical methods underlying a recent quantum feedback experiment stabilizing photon-number states [17], [30], [29], [24]. We consider a controlled system whose quantum state, a finite dimensional density operator, is governed by a discrete-time nonlinear Markov process. In open-loop, the measurements are assumed to be quantum non-demolition (QND) measurements. This Markov process admits a set of stationary pure states associated to an orthonormal basis. These stationary states provide martingales crucial to prove the open-loop stability: under simple assumptions, almost all trajectories converge to one of these stationary states; the probability to converge to a stationary state is given by its overlap with the initial quantum state. From these open-loop martingales, we construct a supermartingale whose parameters are given by inverting a Metzler matrix characterizing the impact of the control input on the Kraus operators defining the Markov process. This supermartingale measures the "distance" between the current quantum state and the goal state chosen from one of the open-loop stationary pure states. At each step, the control input minimizes the conditional expectation of this distance. It is proven that the resulting feedback scheme stabilizes almost surely towards the goal state whatever the initial quantum state. This state feedback takes into account a known constant delay of arbitrary length in the control loop. This control strategy is proved to remain also convergent when the state is replaced by its estimate based on a quantum filter. It relies on measurements that can be corrupted by random errors with conditional probabilities described by a known left stochastic matrix. Closed-loop simulations corroborated by experimental data illustrate the interest of such nonlinear feedback scheme for the photon box [29].

We have also investigated the stabilization of the dynamical state of a superconducting qubit. In a series of papers, A. Korotkov and his co-workers suggested that continuous weak measurement of the state of a qubit and applying an appropriate feedback on the amplitude of a Rabi drive, should maintain the coherence of the Rabi oscillations for arbitrary time. Here, in the aim of addressing a metrological application of these persistent Rabi oscillations, we explore a new variant of such strategies. This variant is based on performing strong measurements in a discrete manner and using the measurement record to correct the phase of the Rabi oscillations. Noting that such persistent Rabi oscillations can be viewed as an amplitude- to-frequency convertor (converting the amplitude of the Rabi microwave drive to a precise frequency), we propose another feedback layer consisting of a simple analog phase locked loop to compensate the low frequency deviations in the amplitude of the Rabi drive [60].

### 5.3.2. Dissipation engineering

We have introduced a new quantum gate that transfers an arbitrary state of a qubit into a superposition of two quasi-orthogonal coherent states of a cavity mode, with opposite phases. This qcMAP gate is based on conditional qubit and cavity operations exploiting the energy level dispersive shifts, in the regime where they are much stronger than the cavity and qubit linewidths [77], [26]. The generation of multi-component superpositions of quasi-orthogonal coherent states, non-local entangled states of two resonators and multi-qubit GHZ states can be efficiently achieved by this gate. We also propose a new method, based on the application of this gate, to autonomously correct for errors of a logical qubit induced by energy relaxation. This scheme encodes the logical qubit as a multi-component superposition of coherent states in a harmonic oscillator. The error correction is performed by transferring the entropy to an ancila qubit and reseting the qubit. We layout in detail how to implement these operations in a practical system [78]. This proposal directly addresses the task of building a hardware-efficient and technically realizable quantum memory [78].

We have also studied the application of dissipation engineering techniques to perform a high-performance and fast qubit reset. Qubit rest is crucial at the start of and during quantum information algorithms. Our protocol, nicknamed DDROP (Double Drive Reset of Population) is experimentally tested on a superconducting transmon qubit and achieves a ground state preparation of at least $99.5\%$ in times less than $3\mu$s; faster and higher fidelity are predicted upon parameter optimization [74]. We are currently working on extending our protocol to prepare and protect two-qubit entangled states and to perform autonomous quantum error correction.

## 5.4. Modeling, observation and control in biosciences - Reproductive system

### *5.4.1. Numerical simulation of the selection process of the ovarian follicles*

**Participants:** Benjamin Aymard, Frédérique Clément.

Collaboration with Frédéric Coquel and Marie Postel.

*Implementation of a parallelized numerical scheme based on finite volumes.* We have designed and implemented a numerical method to simulate a multiscale model describing the selection process in ovarian follicles [11], [10]. The PDE model consists in a quasi-linear hyperbolic system of large size, namely $N_f \times N_f$, ruling the time evolution of the cell density functions of $N_f$ follicles (in practice $N_f$ is of the order of a few to twenty). These equations are weakly coupled through the sum of the first order moments of the density functions. The time-dependent equations make use of two structuring variables, age and maturity, which play the roles of space variables. The problem is naturally set over a compact domain of $\mathbf{R}^2$. The formulation of the time-dependent controlled transport coefficients accounts for available biological knowledge on follicular cell kinetics. We introduce a dedicated numerical scheme that is amenable to parallelization, by taking advantage of the weak coupling. Numerical illustrations assess the relevance of the proposed method both in term of accuracy and HPC achievements [32].

*A numerical method for cell dynamics: kinetic equations with discontinuous coefficients.* The motivation of this work is the numerical treatment of the mitosis in biological models involving cell dynamics. More generally we study hyperbolic PDEs with flux transmission conditions at interfaces between subdomains where coefficients are discontinuous. A dedicated finite volume scheme with a limited high order enhancement is adapted to treat the discontinuities arising at interfaces. The validation of the method is done on 1D and 2D toy problems for which exact solutions are available, allowing us to do a thorough convergence study. A simulation on the original biological model illustrates the full potentialities of the scheme [72].

### *5.4.2. Optimal control of cell mass and maturity in a model of follicular ovulation*

**Participants:** Frédérique Clément, Peipei Shang.

Collaboration with Jean-Michel Coron

We have studied some optimal control problems associated with a scalar hyperbolic conservation law modeling the development of ovarian follicles. Changes in the age and maturity of follicular cells are described by a 2D conservation law, where the control terms act on the velocities. The control problem consists in optimizing the follicular cell resources so that the follicular maturity reaches a maximal value in fixed time. Formulating the optimal control problem within a hybrid framework, we have proved necessary optimality conditions in the form of Hybrid Maximum Principle [36]. We have then derived the optimal strategy and shown that there exists at least one optimal bang-bang control with one single switching time.

### *5.4.3. Multiscale analysis of mixed-mode oscillations in a phantom bursting model*

**Participants:** Frédérique Clément, Mathieu Desroches, Maciej Krupa, Alexandre Vidal.

We have studied mixed mode oscillations in a model of secretion of GnRH (gonadotropin releasing hormone). The model is a phantom burster consisting of two feedforward coupled FitzHugh-Nagumo systems, with three time scales. The forcing system (Regulator) evolves on the slowest scale and acts by moving the slow null-cline of the forced system (Secretor). There are three modes of dynamics: pulsatility (transient relaxation oscillation), surge (quasi steady state) and small oscillations related to the passage of the slow null-cline through a fold point of the fast null-cline. We have derived a variety of reductions, taking advantage of the mentioned features of the system. We have obtained two results; one on the local dynamics near the fold in the parameter regime corresponding to the presence of small oscillations and the other on the global dynamics, more specifically on the existence of an attracting limit cycle. Our local result is a rigorous characterization of small canards and sectors of rotation in the case of folded node with an additional time scale, a feature allowing for a clear geometric argument. The global result gives the existence of an attracting unique limit cycle, which, in some parameter regimes, remains attracting and unique even during passages through a canard explosion [43].

### *5.4.4. A network model of the periodic synchronization process in the dynamics of calcium concentration in GnRH neurons*

**Participants:** Frédérique Clément, Maciej Krupa, Alexandre Vidal.

Mathematical neuroendocrinology is a branch of mathematical neurosciences that is specifically interested in endocrine neurons, which have the uncommon ability of secreting neurohormones into the blood. One of the most striking features of neuroendocrine networks is their ability to exhibit very slow rhythms of neurosecretion, on the order of one or several hours. A prototypical instance is that of the pulsatile secretion pattern of GnRH (gonadotropin releasing hormone), the master hormone controlling the reproductive function, whose origin remains a puzzle issue since its discovery in the seventies. We have investigated the question of GnRH neuron synchronization on a mesoscopic scale and study how synchronized events in calcium dynamics can arise from the average electric activity of individual neurons. We have used as reference seminal experiments performed on embryonic GnRH neurons from rhesus monkeys, where calcium imaging series were recorded simultaneously in tens of neurons, and which have clearly shown the occurrence of synchronized calcium peaks associated with GnRH pulses, superposed on asynchronous, yet oscillatory individual background dynamics [100]. We have designed a network model by coupling 3D individual dynamics of FitzHugh-Nagumo type. Using phase-plane analysis, we have constrained the model behavior so that it meets qualitative and quantitative specifications derived from the experiments, including the precise control of the frequency of the synchronization episodes. In particular, we have shown how the time scales of the model can be tuned to fit the individual and synchronized time scales of the experiments. Finally, we have illustrated the ability of the model to reproduce additional experimental observations, such as partial recruitment of cells within the synchronization process or the occurrence of doublets of synchronization [76].

## 5.5. Clinical and physiological applications

### *5.5.1. DynPeak: An algorithm for pulse detection and frequency analysis in hormonal time series*

**Participants:** Frédérique Clément, Claire Médigue, Alexandre Vidal, Qinghua Zhang.

Collaboration with Stéphane Fabre (UMR CNRS-INRA 6175).

The endocrine control of the reproductive function is often studied from the analysis of luteinizing hormone (LH) pulsatile secretion by the pituitary gland. Whereas measurements in the cavernous sinus cumulate anatomical and technical difficulties, LH levels can be easily assessed from jugular blood. However, plasma levels result from a convolution process due to clearance effects when LH enters the general circulation. Simultaneous measurements comparing LH levels in the cavernous sinus and jugular blood have revealed clear differences in the pulse shape, the amplitude and the baseline. Besides, experimental sampling occurs at a relatively low frequency (typically every 10 min) with respect to LH highest frequency release (one pulse per hour) and the resulting LH measurements are noised by both experimental and assay errors. As a result, the pattern of plasma LH may be not so clearly pulsatile. Yet, reliable information on the InterPulse Intervals (IPI) is a prerequisite to study precisely the steroid feedback exerted on the pituitary level. Hence, there is a real need for robust IPI detection algorithms. We have designed an algorithm for the monitoring of LH pulse frequency, basing ourselves both on the available endocrinological knowledge on LH pulse (shape and duration with respect to the frequency regime) and synthetic LH data generated by a simple model [54]. We make use of synthetic data to make clear some basic notions underlying our algorithmic choices. We focus on explaining how the process of sampling affects drastically the original pattern of secretion, and especially the amplitude of the detectable pulses. We then describe the algorithm in details and perform it on different sets of both synthetic and experimental LH time series. We further comment on how to diagnose possible outliers from the series of IPIs which is the main output of the algorithm.

# VIRTUAL PLANTS Project-Team

# 5. New Results

## 5.1. Analysis of structures resulting from meristem activity

### 5.1.1. Acquisition and design of plant geometry

**Participants:** Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier.

*This research theme is supported by RTRA project named PlantScan3D.*

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (light, rain, wind, pests, ...), to set up initial conditions of growth models or to assess their output against real data. In the past decade, methods have been developed to digitize plant architectures in 3D [48], [42]. 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 pictures, laser scanner, video and direct sketching.

- *Automated reconstruction of plant architecture* (Chakkrit Preuksakarn, Mathilde Balduzzi, Julien Diener, Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Bernard Mourrain [Inria, Galaad], Franck Hetroy [Inria, Morpheus], Marie-Paule Cani [Inria, Imagine], Pascal Ferraro [Labri, Bordeaux] )

  We investigate the possibility to use 3D laser scanners to automate plant digitizing. We are developping 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 this we collaborate with the EPI Galaad from Sophia-Antipolis, the EPI Imagine from Grenoble, different INRA teams, UMR PIAF in Clermont Ferrand, UMR LEPSE and AFEF team in Montpellier and Lusignan, the University of Helsinki, Finland and the CFCC in England. We developed a reconstruction pipeline composed of several procedures. A contraction procedure, first aggregates points at the center of the point cloud. The team proposed a simple adaptive scheme to contract points. In a second step, a skeleton procedure uses a Space Colonization Algorithm [47] to build the skeleton of the shape from the contracted point set. This method is adaptive to the local density of the point set. Then a pipe-model based procedure makes it possible to estimate locally diameters of the branches. Finally, an evaluation procedure has been designed to assess the accuracy of the reconstruction and a comparison with alternative methods has been carried out. Publication of this work is in progress.

  An automated reconstruction pipeline is also developed for processing 2D images of root system architecture (RSA) in the context of the Rhizopolis project. The analysis of these data is currently a major challenge in understanding root development. Existing tools either focus on specific applications, on simple structures (for example one root segment) or require long manual work. Here, we develop a processing pipeline that takes as an input 2D high resolution images of petri plates containing root systems. The pipeline makes it possible to extract from the images the whole architecture of root systems, with minimal or no user intervention. In order to obtain this result, the problem was decomposed in several steps: filter and label the input image, extract the image skeleton as a general graph structure and then convert it into a tree structure representing the visualised RSA, using a priori knowledge to solve inconsistencies. The pipeline has been added to the OpenAlea platform, thus allowing resulting data to be directly processed by other advanced high-level computational or statistical tools. The developed pipeline is currently being tuned and tested on several databases of 2D images with varying complexities of both arabidopsis and rice.

*Figure 3. Reconstruction of a cherry tree. Left: photograph of the original tree. Right: 3D reconstruction from a laser scan rendered and integrated on the same background.*

Additionnally, we also investigate the reconstruction of tree foliage from laser scanners in the context of Mathild Balduzzi's PhD thesis. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers on the silhouette of the scans that make the meshing of the point set difficult. New generation of laser scanners provides intensity of the laser reflected) on the surface of scanned objects. This intensity is dependant of the distance to the object, its optical property and the incidence angle. A first work on this topic shows that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we develop a reconstruction technique using the scan intensities and based on Shape-From-Shading approaches. The idea is to generate a new point set from the intensities and a set of seed points. This new point set has the property of being smooth but is not necessary the exact representation of the scanned object. To consolidate the reconstruction, we are working on merging it with the orignal noisy point set coming from the scans using Kalman filtering. As a result, a final point set will be obtained without noise and with outiers naturally removed.

- *Sketching of plants.* (Frédéric Boudon, Christophe Godin, Steven Longuay [University of Calgary, Canada], Przemyslaw Prusinkiewicz [University of Calgary, Canada])

  Modeling natural elements such as trees in a plausible way, while offering simple and rapid user control, is a challenge. In a first collaboration with the EPI Imagine (ex-Evasion) we developed a method based on the design of plants from silhouettes [50]. This sketching paradigm allows quick and intuitive specification of foliage at multiple scales. On this topic, we started a collaboration with S. Longuay and P. Prusinkiewicz who develop iPad tools to design plants based on SCA. Combination of multitouch interface, sketching paradigm and powerfull adaptive procedural model that generate realistic trees offer intuitive and flexible design tools. This work is part of the Inria associated team with the University of Calgary. It has been published Eurographics Symposium on Sketch-Based Interfaces and Modeling [32].

- *Reconstruction from video.* (Frédéric Boudon, Jérome Guenard [IRIT, Toulouse], Géraldine Morin [IRIT, Toulouse], Pierre Gurdjos [IRIT, Toulouse], Vincent Charvillat [IRIT, Toulouse])

  Even if mature computer vision techniques allow the reconstruction of challenging 3D objects from images, due to high complexity of plant topology, dedicated methods for generating 3D plant models must be devised. We propose an analysis-by-synthesis method which generates 3D models of a plant from both images and a priori knowledge of the plant species.

  Our method is based on a skeletonisation algorithm that allows to generate a possible skeleton from a foliage segmentation. Then, a 3D generative model, based on a parametric model of branching systems that takes into account botanical knowledge is built. This method extends previous works by constraining the resulting skeleton to follow hierarchical organisation of natural branching structure.

3D models are then generated. A reprojection of the models can be compared with the original image to assess the visual accuracy of the reconstruction. We optimise the parameter values of the generative model based on the reprojection criterion. Realistic results are obtained on different species of plants, in particular vineyards. Publication of this work is in progress.

- *Reconstruction of virtual fruits from pictures.* (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 Fruit3D.*

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function and used to investigate such effects. To this end, we have developed a modeling pipeline in the OpenAlea platform that involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex network of vasculature that is embedded within this mesh. To create the 3D volumetric mesh, we use reconstruction algorithms from the 3D mesh generation package of the Computational Geometry Algorithms Library (CGAL). To generate the pattern of vasculature within this volumetric mesh, we use a Space Colonisation Algorithm that populates the volume of the fruit by simulating competition for space of the vasculature. We have applied our modeling pipeline to generate the internal and external geometry of a cherry tomato fruit using Magnetic Resonance Imaging data as input. These studies demonstrate the possibility to create species-specific models of fruit structure with relatively low effort [26]. These volumetric meshes are then combined with models of function to form integrative computational fruit models, which will help to investigate the effects of fruit structure on quality (see section 5.3.2 ).

- *Reconstruction of gramineous leaves.* (Christian Fournier, Christophe Pradal )

  *This research theme is supported by the Agropolis project OpenAlea.*

  Unlike trees, the 3D architecture of gramineous plants is much more related to the shapes of its leaves than the arrangement of its branches. Many modeling efforts have thus concentrated on correctly capturing its complex shape at different stages and use them as scalable geometric primitives. Still, additional control of such objects is needed in the context of Functional Structural Modeling. The objective of this work is to propose a plastic and dynamic 3D leaf model that is well suited for such uses, still able to capture a variety of observed static shapes. Leaf shape is modeled by a parametric surface describing leaf midrib curvature, leaf width variation, undulation of leaf margins and twist along the midrib. Meshes can be generated form these surfaces, and reduced using a decimation algorithm. The model can be fed with data or with curves drawn by user interaction. Morphological operators are defined and allows for plastic deformation of the control curves. The dynamics of shape acquisition can also be specified, and combined with morphological operators to simulate various scenarii of evolution and responses to stresses. The capabilities of the model are demonstrated through several cases of use. Future directions of research are thought to be a better integration of mechanical or physiological constraints that would reduce the model plasticity but avoid user-induced unrealistic simulation. [28].

### 5.1.2. Modeling the plant ontogenic program

**Participants:** Christophe Godin, Yann Guédon, Evelyne Costes, Jean-Baptiste Durand, Anaëlle ambreville, Pierre Fernique, Christophe Pradal, Jean Peyhardi, Catherine Trottier, Yassin Refahi, Etienne Farcot.

*This research theme is supported by two PhD programs.*

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. [37], [43], [44], [41], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as "morphogenetic program" [45], "age state" [40] or "physiological age" [38]. 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 [40], [38]. Here we develop computational methods to decipher these rules.

- *Relating branching structure to the shoot properties* (Jean Peyhardi, Yann Guédon, Evelyne Coste, Catherine Trottier, 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 is 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. maximum growth rate of the leaf, surface of the leaf, length of the internode, 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 in our setting, we chose to develop a new family of multinomial GLMs called multi-step multinomial GLMs that enable to tackle partially ordered categorical response variables. Typically, we need to distinguish different timing of branching (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 multi-step multinomial GLMs and are applied to different data sets corresponding mainly to fruit tree branching structures.
- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Jean Peyhardi, Baptiste Guitton [AGAP, AFEF team], Yan Holtz [AGAP, AFEF team] Catherine Trottier, Evelyne Costes, Yann Guédon)

  The aim of this work was to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot or AS, which is the 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 allows for quantifying alternation from the yearly numbers of inflorescences at tree scale.

  However, phenotyping subsamples of AS sequences is more realistic in the framework of genotype selection. To model alternation of flowering at AS scale, a second-order Markov tree model was built. Its transition probabilities were modeled as generalized linear mixed models, to incorporate the effects of genotypes, year and memory of flowering for the Markovian part, with interactions between these components. Asynchronism of flowering at AS scale was also assessed using an entropy-based criterion.

  This work started during the PhD's work of Baptiste Guitton. It was then extended in 2012 by Yan Holtz during this Master 2 internship, supervised by Evelyne Costes and Jean-Baptiste Durand. New progenies were considered, as well as the performance of approximating the descriptors at whole tree scale with those at AS scale. These descriptors allowed the identification of QTL zones involved in the control of flowering in apple trees.

  As a perspective of this work, patterns in the production of children ASs (numbers of flowering and vegetative children) depending on the type of the parent AS must be analyzed using branching processes and different types of Markov trees, in the context of Pierre Fernique's PhD Thesis (see next item in Section 5.1.2 ).

- *Modeling branching patterns in fruit tree shoots through the characterization of their demographic properties* (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).

  To test the effect of some properties of a given parent shoot on the properties of its children shoots, statistical models based on multitype branching processes were developed. This kind of dependence between parent and children shoots is frequently at stake in fruit trees, for which the number of flowering or vegetative children of a parent shoot depends on its nature, with potential interactions with other factors. Thus, controlling demographic patterns of the shoots (through varietal selection or crop management strategies) is expected to bring substantial improvements in the quantity and quality of yields.

  Formally, the shoot properties are summed up using the notion of shoot state. The number of children shoots in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious 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. A parametric distribution was associated with each graph. It was obtained by combining families of univariate and multivariate distributions or regression models. These were chosen by selection model procedures among different parametric families.

  This work was carried out in the context of Pierre Fernique's first year of PhD (Montpellier 2 University and CIRAD). It was applied to model dependencies between short or long, vegetative or flowering shoots in apple trees. The results highlighted contrasted patterns related to the parent shoot state, with interpretation in terms of alternation of flowering (see previous item in Section 5.1.2 ). It was also applied to the analysis of the connections between cyclic growth and flowering of mango trees. This work will be continued during Pierre Fernique's PhD thesis, with extensions to other fruit tree species and other parametric discrete multivariate families of distributions, including covariates and mixed effects.
- *Analyzing fruit tree phenology* (Anaëlle Dambreville, Jean-Baptiste Durand, Pierre Fernique, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand, Catherine Trottier) Mango is a tropical tree characterized by strong asynchronisms within and between trees. Causation networks explaining the vegetative and reproductive growths within and between growing cycles were studied on the basis of generalized linear models. We highlighted in this way marked interplays between structural and temporal components of tree structure development at three scales. At growth unit scale, a growth unit appeared early in the growing cycle had higher rate of burst compared to late appeared growth units. At growing cycle scale, a growth unit which flowered delayed its future vegetative growth compared to a vegetative growth unit. At tree scale, a fruiting tree delayed further vegetative growth and flowering compared to a non-fruiting tree. These results evidenced that tree phenology is strongly affected by structural components and not only by the environment. We are now investigating jointly structure development and phenology of mango using statistical models for trees in particular hidden Markov tree models and multitype branching processes.
- *Integrative developmental growth stages of shoots* (Anaëlle Dambreville, Yann Guédon, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand) Growth and development are often studied as two separated processes. Our aim is to investigate the coordination between growth and development in mango shoots. We considered three types of organ, namely the shoot axis, its attached leaves and the inflorescence. Two types of data were collected during the shoot and inflorescence follow-up: developmental stages determined in an expert way and organs sizes determined from measurements. To give an integrative view of the shoot and inflorescence growth and development, we adopted the following strategy. For a given cultivar, we first built a multi-state model on the basis of absolute growth

rate sequences deduced from the measurements. Using these models, we computed growth stages. These growth stages highlighted growth asynchronisms between two topologically-connected organs: the axis and its leaves. Then, we compared these growth stages with the developmental ones and we obtained strong matches between them. The integrated developmental growth stages emphasized that the developmental stages are markedly related to growth rates an can be interpreted in terms of physiological (hydraulics, carbohydrates partitioning) and developmental (organs preformation versus neoformation) processes.

- *Self-nested structure of plants.* (Christophe Godin, Farah Ben Naoum) In a previous work [6], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG, 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. This method thus compresses a tree in width, but not in height. In this 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 is being written.

### 5.1.3. Analyzing the influence of the environment on the plant ontogenic program

**Participants:** Frédéric Boudon, Jean-Baptiste Durand, Christophe Godin, Yann Guédon, Jean Peyhardi, Pierre Fernique, Maryline Lièvre, Christine Granier, Evelyne Costes, Pascal Ferraro, Catherine Trottier.

*This research theme is supported by three PhD programs.*

The ontogenetic program 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 5.1.2 ), 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 program 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 environment 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 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])

  In a forest ecology context, we identified robust indicators that summarize the balance between tree ontogeny and environmental constraints (mainly related to light environment). In this context, tree growth data typically correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models (hidden Markov and semi-Markov chains) that enable to identify tree growth phases. This statistical modeling approach was applied to both deciduous (sessile oak and Persian walnut) and evergreen (Corsican pine and silver fir) tree species growing in contrasted conditions ranging from managed forest stands to unmanaged understoreys. The growth phase duration distributions estimated within these segmentation models characterize the balance between tree ontogeny and the environmental constraints in tree development at the population scale. These distributions had very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. The characteristics of growth phase duration distributions may change over tree life reflecting changes in tree competition.

- *Investigating the influence of geometrical traits on light interception efficiency of trees and grass* (Liqi Han [AFEF Team, AGAP], Christophe Pradal, Frédéric Boudon, Christophe Godin, David Da Silva [UC Davis], Evelyne Costes, Philippe Balandier [PIAF], André Marquier [PIAF], Gaëtan Louarn [URP3F], Didier Combes [URP3F], Christian Fournier)

  Light availability in forest understory is essential for many processes. It controls for instance the growth potential of species and individuals in plant communities. It is, therefore, a valuable information regarding forest and crop management. However, the effects of competition for light on short term vegetation dynamics are still poorly understood. This is in part due to a lack of tractacle and precise methods to estimate light resource within a canopy. To alleviate this difficulty, models can be used to compute light interception. At a detailed scale, they often require a lot of field data to accurately predict light distribution, particularly in the case of heterogeneous canopies.

  To investigate this issue, we first analyzed the deterioration of the prediction quality of light distribution to the reduction of inputs by comparing simulations to transmitted light measurements in forests of increasing complexity in three different locations [17]. With a full set of parameters to describe the tree crown (i.e., crown extension in at least eight directions, crown height and length), the model accurately simulated the light distribution. Simplifying crown description by a geometric shape with a mean radius of crown extension led to deteriorated but acceptable light distributions. Allometric relationships used to calculate crown extension from trunk diameter at breast height seriously reduced light distribution accuracy.

  We also studied the light interception of herbaceous plants with contrasting architectures (monocultures and binary mixtures) grown at high or low density and sought to determine the important architectural features necessary to account for light partitioning among individual plants [21]. It was shown that the studied plant populations were typical of a wide range of competition intensities, ranging from sparse plants to dense size-structured populations. Plant representations using whole plant envelopes with homogeneous leaf area density (LAD) were not reliable to estimate light partitioning, irrespective of the accuracy of envelope definition. Accounting for heterogeneous LAD within plants helped to solve this problem in both sparse and dense canopies. The relative importance of traits however changed with competition intensity and was different from reports made on isolated plants. Simple envelope-based reconstructions were finally shown robust enough to support parameterisation from a tractable set of traits measured in the field provided that height and vertical LAD gradient were characterised.

  Using virtual growth simulation tools, a detailled analysis was also carried out to study more precisely the influence of architectural variability of apple trees on their light interception efficiency [30]. For this we used MAppleT, an in silico functional-structural plant model that has been built for simulating architectural development of apple trees. The STAR, namely the silhouette to total area ratio, of leaves, was chosen to evaluate the level of such efficiency. The strategy was to integrate MAppleT with the light interception model provided by the fractalysis module of the VPlants software library. Target values of four major traits (internode length, leaf area, branching angle and top shoot diameter), are varied in range previously observed in a segregating population of apple hybrids. A sensitivity analysis based on polynomial and generalised additive models was performed for highlighting the most influential trait on light interception and suggesting the optimal combination(s) of traits leading to the highest STAR. The contribution of stochastic processes that pilot tree topology in MAppleT is also investigated in the sensitivity analysis. This study not only provides a time- and resource-saving alternative for data collection, but also sets a methodology for ideotype definition and further genetic improvement of apple trees.

## 5.2. 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.

### *5.2.1. Data acquisition and design of meristem models*

**Participants:** Frédéric Boudon, Christophe Godin, Christophe Pradal, Vincent Mirabet [RDP, ENS], Jan Traas, Grégoire Malandain, Jean-Luc Verdeil [PHIV, AGAP].

*This research theme is supported by the iSam and Morphogenetics projects.*

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging
  **Participants:** Léo Guignard, Christophe Godin, Grégoire Malandain, Jan Traas, Pradeep Das [RDP, ENS], Vincent Mirabet [RDP, ENS].

  The MARS-ALT (Multi-Angles Registration and Segmentation - Automatic Lineage Tracking) software pipeline 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 [5]. A new version of this pipeline is currently being developed. MARS-ALT Version 2 is based on the same algorithms and methods and is intended to improve the overall robustness of the pipeline (protocol, noise in the input image) and automate completely the process. To test the new pipeline, we use different acquisition protocols and different organisms (floral and apical meristems and the early stages of development of a marine animal *Phallusia mammillata*). The segmentation is corrected a posteriori to deal with imaging artifacts due to uncertainties of acquisition. The image data set on which we develop the methods consists of :

  - *Arabidopsis thaliana* shoot apical meristem and primordia with around 6000 cells. The organ is captured from three different angles every 20 hours 3 or 4 times with a confocal microscope (Collaboration Sainsbury lab, Cambridge)
  - *Arabidopsis thaliana* flower meristems with around 2000 cells. The organ is also captured from three different angles every 20 hours 3 to 5 times with a confocal microscope (Collaboration RDP Lyon)
  - *Phallusia mammillata* and *Ciona intestinalis* embryos with from 32 cells to around 1000 cells. The organism is captured from four different angles every 2 minutes during 2 to 3 hours with a SPIM (Single Plane Illumination Microscope) (Collaboration CRBM Montpellier / EMBL Heidelberg)

  The pipeline provides as an output segmented images on which metrics for each cells can be extracted such as volume, principal components, convex hull and so on. A new non-linear registration algorithm developed by G. Malandain (MORPHEME team, Inria Sophia-Antipolis) is now available and will lead to an improvement of ALT algorithm. Redesign and improvement of the lineage tracking pipeline will be the next step.

- *Design of 3D virtual atlases for specifying gene expression patterns* (Jérôme Chopard, Christophe Godin, Jan Traas, Françoise Monéger [RDP, ENS])

  *This research theme is supported the ANR GeneShape and iSam projects.*

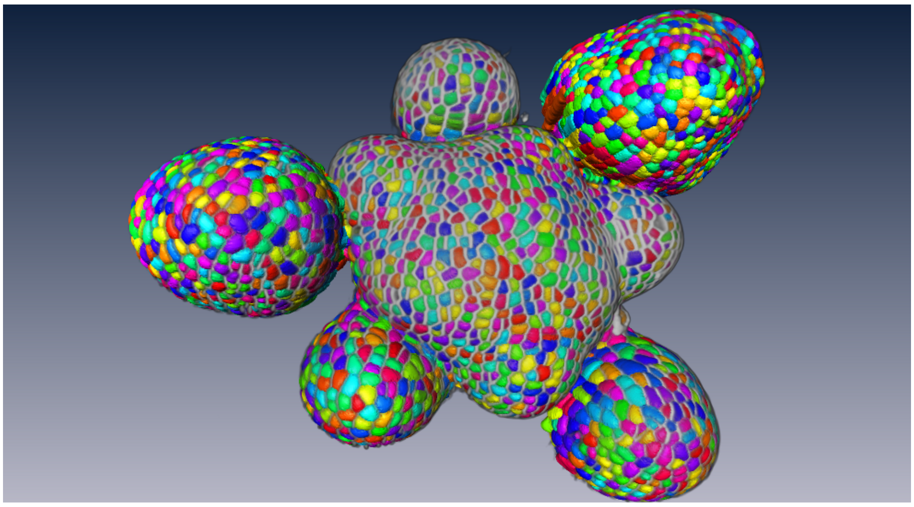


*Figure 4. Superimposition of an automatic cell segmentation of an arabidopsis flower meristem using the new MARS pipeline with the original confocal image stack where the membranes are marked.*

To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database. 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 virtual atlas is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the virtual 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.

A prototype version of the 3D virtual atlas was built last year [8]. Further developments of this tool will rely on the segmented images produced from microscopy, as presented in the previous section. In particular, a common underlying data structure has to be developed transversally to these two scientific developments. The definition of this data structure has been initiated last year through several team meetings, and should lead to a revised implementation next year.

### 5.2.2. Shape analysis of meristems

(Jonathan Legrand, Clémence Hatt [BURST, AGAP], Jean-Baptiste Durand, Frédéric Boudon, Christophe Godin, Yann Guédon, François Mankessi [BURST, AGAP], Olivier Monteuuis [BURST, AGAP], Jean-Luc Verdeil [PHIV, AGAP] )

Plants that grow several forms or type of leaves along a shoot, depending on age or shoot length, are called heteroblastic.The influence of heteroblasty on morphological and histocytological characteristics of Acacia mangium shoot apical meristems (SAMs) was assessed comparing materials with mature and juvenile leaf morphology in natural and in vitro conditions. For this we introduced a workflow for characterizing dome shape with few parameters (SAM dome heigth (H), basal diameter (D) and shape factor (S)) and their joint statistical analysis to assess influence of conditions on SAM shape. In particular, a new statistical test is introduced here for multivariate analysis. This is a generalization of univariate ANOVA that takes into account statistical dependencies between the shape parameters. As a result, we found that SAM dome height (H) and basal diameter (D) were highly correlated. The joint analysis revealed that H, D, and shape (S) varied significantly according to the four plant origins investigated, with the higher scores for the outdoor mature source "Mat". Overall, heteroblasty induced more conspicuous differences of SAM characteristics for the outdoor than for the in vitro materials. A paper presenting these results has been published in Trees [20]

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 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). We are now designing models and algorithms for finding patterns in time courses of meristems. In particular, we are investigating spectral clustering methods in order to define homogeneous regions in term of cell identities.

### *5.2.3. Transport models*

**Participant:** Michael Walker.

*This research theme is supported by the ANR GeneShape and ERASysBio+ iSAM projects and Morphogenetics.*

Active transport of the plant hormone auxin has been shown to play a key role in the initiation of organs at the shoot apex, and vein formation in both leaves and the shoot apical meristem. Polar localized membrane proteins of the PIN1 and AUX/LAX family facilitate this transport and observations and models suggest that the coherent organization of these proteins in the L1 layer is responsible for the creation of auxin maxima (surrounded by a depletion zone), which in turn triggers organ initiation close to the meristem center [46] [1]. Furthermore, canalized PIN allocations are thought to play a crucial role in vein formation in the leaf and in the L2. Previous studies have typically modeled the L1 and L2 with different models to explain different patterns of PIN allocations. In the last two years, we developed a unifying model showing that a unique flux-based model could be sufficient to explain PIN patterns in both L1 and L2 [25]. Contrary to our previous study [11], here no change in the model parameters is needed for this. Our approach is based on inherent topological and geometrical differences between the L1 and L2, specifically their dimensionality and the distribution of sources and sinks.

In a different perspective, another study on auxin transport models have been submitted this year. In this work, a generic, adimensional flux-based model of auxin transport was studied using a combination of analytic and numeric approach. The steady-states with uniform auxin distribution where characterised for arbitrary tissues, and some of their bifurcations (loss of stability and Hopf) were described. This work, initiated during an "Explorateur" project funded by Inria during the period October 2012-January 2013, was submitted by E. Farcot and Y. Yuan (Memorial University of Newfoundland, Canada) in September and is still under review.

### *5.2.4. Mechanical model*

**Participants:** Jérôme Chopard, Olivier Ali, Christophe Godin, Frédéric Boudon, Jan Traas, Olivier Hamant [ENS-Lyon], Arezki Boudaoud [ENS-Lyon].

*This research theme is supported by the ANR VirtualFlower and Geneshape projects together with the Inria project Morphogenetics and the ERC from Jan Traas.*

The rigid cell walls that surround plant cells is responsible for their shape. These structures are under constraint due to turgor pressure inside the cell. To study the changes of shape in plant tissues during organogenesis, we need a mechanical model of tissue development at cellular resolution. We developed such a model, in which walls are characterized by their mechanical properties like the Young modulus which describes the elasticity of the material. Wall deformation results from forces due to turgor pressure. Growth results from cell wall synthesis that is triggered when wall deformation exceeds a particular threshold. The final shape of the tissue integrates mechanically all the local deformations of each cell.

To model this process, we used a tensorial approach to describe both tissue deformation and stresses. Deformations were decomposed into elementary transformations that can be related to underlying biological processes. However, we showed that the observed deformations does not map directly local growth instructions given by genes and physiology in each cell. Instead, the growth is a two-stage process where genes are specifying how cell walls should yield to mechanical stresses. In this way, different regions in the tissue with different cell identities can have different growth properties. The final shape of the tissue results from the integration of all these mechanical properties and stresses at organ level under the growth force due to turgor pressure at tissue scale.

A paper describing the mechanical model and its application to model primorium formation in the shoot apical meristem is currently being written. Additionally, a redesign of our mechanical model using the SOFA framework is in progress.

### 5.2.5. Gene regulatory networks

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene expression. This fact is becoming more and more acknowledged, and has led this year to the publication of several review articles including members of Virtual Plants as authors [22], [23], [24], [13].

- *The auxin signaling pathway* (Etienne Farcot, Yann Guédon, Christophe Godin, Yassin Refahi, Jonathan Legrand, Jan Traas, Teva Vernoux, Stéphane Robin [AgroParisTech], Jean-Benoist Leger [AgroParisTech])

  The auxin signalling network involves about 50 potentially interacting factors. We applied a graph clustering method [12] that relies on 0/1 interactions between factors deduced from yeast two-hybrid (Y2H) data. The Y2H analysis involves two independent tests (X-gal and HIS3 tests). Each possible interaction was tested in the two possible configurations, where each protein was alternatively the bait and the prey protein. A binary interaction is thus a summary of the four outputs of the X-gal and HIS3 tests. In order to limit the loss of information, we designed a standardization procedure to summarize the outputs of the X-gal and HIS3 tests as a distance defined on a continuous scale. This opens the possibility to study the influence of phylogenetic distances between factors on their interactions using an extension of the mixture model for random graphs that incorporate explanatory variables. This new model evidences different behaviors between the ARF+ and the Aux/IAA factors.

  Extensions of this ODE model are necessary to better understand this system in more general contexts than the shoot apical meristem development, which was the framework of our previous study. This work involves defining and studying mathematically a series of distinct network topologies for the auxin signalling pathway. This was the topic of a Master's internship for the studend Cyril Lavedrine, from September to December 2012, co-supervised by E. Farcot and T. Vernoux. This work led to qualitative predictions which remain to be tested experimentally.

- *Complex dynamics and spatial interactions in gene networks* (Yassin Refahi, Etienne Farcot, Christophe Godin)

  Complex computational and mathematical questions arise in the study of gene networks at two levels: (i) the single cell level, due to complex, nonlinear interactions, (ii) the tissue level, where multiple cells interact through molecular signals and growth, so that even simple local rules can challenge our intuition at higher scales.
  At the single cell level, new results were obtained in the framework of piecewise-linear models. Since their introduction in the late 1960's, these models have been believed to present chaotic behavior in some parameter regimes. However, this was mostly observed numerically, based on intensive generation of random networks. In a long lasting collaboration between E. Farcot and R. Edwards (Univ. Victoria, Canada), with more recent input from one of his students, E. Foxall, we have introduced a method to explicitly build piecewise affine models having a return map which is conjugate to a topological horseshoe. A paper presenting these results has appeared this year [18].

For the same class of piecewise-linear models, it is in general very difficult to entirely characterize the attractors of a given system. In an attempt to improve our ability on this question, a probabilistic approach has been proposed in [15], in which it is shown that a Markov chain can built as an approximation of a given piecewise-linear system, and actually used to make predictions about its periodic attractors.

At a higher scale, we have also continued the study of gene regulation in meristematic tissues. In the context of Y. Refahi's post-doc between Virtual Plants and the group of Henrik Jönsson in Cambridge (Sainsbury Laboratory), we have continued a work that was initiated in Y. Refahi's thesis. This work is motivated by recent biological results, indicating that gradient-like patterns originating from the external layers of meristems may play a decisive role in the specification of the pool of stem cells in a central position. Using the methods in [5], and their on-going improvements, we have acquired new 3D and 4D images that were then segmented. These structures will be used in the next few months to investigate generic patterning properties of gradient like morphogen patterns. This will require a thorough analysis of free diffusion in realistic geometries, as made possible by the newly acquired images. As a preliminary work, we are also currently investigating the formation of gradient patterns in idealised tissues, allowing for deeper analytic treatment than the complex structures obtained by microscopy.

### 5.2.6. Model integration

**Participants:** Mikaël Lucas [IRD], Michael Walker, Jérôme Chopard, Frédéric Boudon, Christophe Godin, Laurent Laplaze, Jan Traas, François Parcy.

*This research theme is supported by the ANR/BBSRC project iSam.*

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). 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 submodel 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 grou at ENS-Lyon [23] and the CPIB group in Nottingham, UK [13].

- *Development of a computer platform for the 'programmable tissue'.* (Michael Walker, Frédéric Boudon, Etienne Farcot, Christophe Godin)

  One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development. This platform will be used to carry out integration of the different models developed in this research axis. The platform is based on *OpenAlea*. Partner models can be integrated in the platform in a non-intrusive way (the code of their model need not be rewritten). In this context, model integration will i) consist of designing adequate data-structures at different levels that will be exchanged and reused among the different plug-in models and ii) defining control flows at adequate levels to avoid the burden of excessive interaction between components. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform, through several group meetings along the year. A redesign of the structure is in progress.

- *Design of a genetic model of inflorescence development.* (Etienne Farcot, Christophe Godin, François Parcy)

  We studied the regulatory network that control flower development 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 repeatedly meristem fails in making a complete transition to the flower. This work couples models at different scales, since gene regulation is described by a minimal gene network, which 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 controling the passage to flower could be verified. However, a complete integrated picture of flower development could not be reached yet. After several unsuccessful attempts, further experiments are currently being made to verify the scenario predicted by the model.

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

### *5.3.1. Transport model in roots*

**Participants:** Mikaël Lucas [IRD], Christophe Pradal, Christophe Godin, 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 developped 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 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 implement 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.

### *5.3.2. Transport in fruits*

**Participants:** 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 Fruit3D.*

Understanding the controlling factors of fruit quality development is challenging, because fruit quality results from the interplay between physical and physiological processes that are under the control of genes and the environment. Although process-based models have been used to make significant progress in understanding these factors, they ignored to a large extent the shape and internal structure of the fruit.

Two essential functions in determining fruit quality are the transport and accumulation of water and dry matter to various fruit tissues. Since water and carbon are delivered to fruit tissues through a complex vasculature system, the internal fruit structure and pattern of vasculature may have a significant impact on their distribution within the fruit.

To help characterizing effects of fruit shape and internal structure on quality, the creation of a 3D virtual fruit model that integrates fruit structure and function with growth governed by environmental inputs has been investigated. For this, a modeling pipeline has been developed that includes the following steps: creation of a 3D volumetric mesh of the internal fruit structure, including vasculature (see section 4 ). Based on previous compartment models of fruit physiology developed at Avignon, we have then developed models of water and carbon transport that have been coupled with the 3D model of fruit. In the 3D model, different equations are describing the transport between adjacent regions of the fruit represented as a 3D mesh. The integration through space and time is carried out using a standard integration scheme (Runge-Kutta of order 4).

This approach has been applied to study tomato fruit (Solanum lycopersicum) by constructing 3D volumetric meshes from different sources (images of perpendicular fruit slices and MRI data), and integrating water and carbon transport processes into these meshes. To illustrate the tomato model, a simulation of one season of the fruit's growth has been performed and its results compared with an already published process-based tomato fruit model. We first showed that the classical results of the abstract process-based models could be also captured by the more detailed spatialized model. However, our model provides additional information on the internal heterogeneity of the fruit, such as a gradient in sugar concentration. Once the model is calibrated and evaluated, our approach will be suitable for studying the effects of internal fruit heterogeneity and overall shape on fruit quality development.

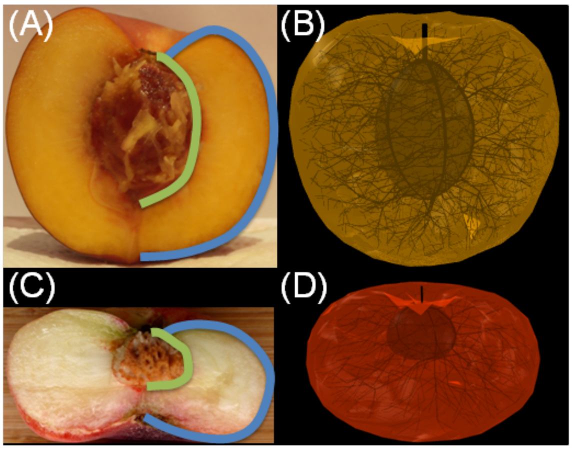


*Figure 5. Virtual models of peaches reconstructed from images with simulated vasculatur to simulate carbon and water transport in the fruit*

### 5.3.3. Analazing shoot and leaf elongation

**Participants:** Maryline Lièvre, Yann Guédon, Christine Granier.

The analysis of phenotyping data coming from automated platforms such as PHENOPSIS often focuses on the growth of a leaf at a given rank along the stem. We aim at developing a pipeline of methods for analyzing the growth of *arabidopsis* shoot at three scales:

1. tissular scale using a probabilistic model of endoreduplication for modeling the distribution of the leaf epidermis cell surfaces. Endoreduplication, which is a replication of the nuclear genome in the absence of cell division that leads to elevated nuclear gene content, strongly affects the leaf epidermis cells of *arabidopsis*.
2. organ scale using nonlinear regression model for analyzing the growth of each successive leaf.

3. shoot scale: The outputs of the analyses at the tissular and organ scales will be summarized as multivariate sequences along the shoots characterizing each successive leaf. These sequences will be augmented by supplementary morphological variables characterizing leaf shape and properties (e.g. presence/absence of trichomes). These sequences will be globally analyzed in order to take into account plant ontogeny and in particular the successive developmental stages before the floral transition for the wild type and selected mutants of *arabidopsis*.

### 5.3.4. Analyzing perturbations in Arabidopsis thaliana phyllotaxis

**Participants:** Christophe Godin, Yann Guédon, Yassin Refahi, Etienne Farcot.

*This research theme is supported by iSAM.*

The cytokinin hormones are known to play a significant role in the regulation of phyllotaxis. To investigate this, Fabrice Besnard and Teva Vernoux are studying *Arabidopsis thalianaahp6* mutants, AHP6 being a protein known for its inhibitory effect in the cytokinin signaling pathway. At the macroscopic scale, this mutation induces perturbations of the phyllotaxis, barely sensible on single plants. In order to characterize these perturbations, we designed a pipeline of models and methods which relies of combinatorial and statistical techniques. Using this pipeline of methods, we have shown that the perturbation patterns in both wild-type and mutant plants can be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs. The number of successive synchronized organs between two permutations reveals unexpected patterns that depend on the nature of the preceding permutation (2- or 3-permutation). We identified significant individual deviations of the level of baseline segments with reference to 137.5°, which confirms theoretical model predictions. Finally, we highlighted a marked relationship between permutation of organs and defects in the elongation of the internodes in between these organs. All these results can be explained by the absence of a strict coupling between the timing of organ development and their angular and longitudinal position on the stem. Two papers (one with biological aspects and the other about methodological developments) are currently in revision.