IN PARTNERSHIP WITH:
Centre CEA-Saclay

Activity Report 2018

Project-Team PARIETAL

Modelling brain structure, function and variability based on high-field MRI data.

IN COLLABORATION WITH: CEA Neurospin

RESEARCH CENTER
Saclay - Île-de-France

THEME
Computational Neuroscience and Medicine

Table of contents

Project-Team PARIETAL

Creation of the Project-Team: 2009 July 01

Keywords:

Computer Science and Digital Science:

A3.3. - Data and knowledge analysis
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A3.4.8. - Deep learning
A5.3.2. - Sparse modeling and image representation
A5.3.3. - Pattern recognition
A5.9.1. - Sampling, acquisition
A5.9.2. - Estimation, modeling
A5.9.6. - Optimization tools
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

B1.2. - Neuroscience and cognitive science
B1.2.1. - Understanding and simulation of the brain and the nervous system
B1.2.2. - Cognitive science
B2.2.6. - Neurodegenerative diseases
B2.6.1. - Brain imaging

1. Team, Visitors, External Collaborators

Research Scientists

Bertrand Thirion [Team leader, Inria, Senior Researcher, HDR]
Philippe Ciuciu [CEA, Researcher, HDR]
Denis Alexander Engemann [Inria, Starting Research Position]
Alexandre Gramfort [Inria, Senior Researcher, HDR]
Gaël Varoquaux [Inria, Researcher, HDR]
Demian Wassermann [Inria, Researcher]

Post-Doctoral Fellows

Darya Chyzhyk [Inria, until Jul 2018]

Elvis Dohmatob [Inria, until Feb 2018]
Thomas Moreau [Inria, from Mar 2018]

PhD Students

Pierre Ablin [Inria]
Thomas Bazeille [Inria, from Oct 2018]
Quentin Bertrand [Inria, from Oct 2018]
Patricio Cerda Reyes [Inria]
Hamza Cherkaoui [CEA]
Jerome-Alexis Chevalier [Inria]
Kamalaker Reddy Dadi [Inria]
Jérome Dockès [Inria]
Loubna El Gueddari [Ministère de l'Enseignement Supérieur et de la Recherche]
Valentin Iovene [Inria, from Jul 2018]
Hubert Jacob Banville [Interaxon, from Sep 2018]
Maeliss Jallais [Inria, from Nov 2018]
Hicham Janati [Inria]
Sylvain Lanuzel [CEA, from Feb 2018]
Carole Lazarus [CEA, until Sep 2018]
Antonia Machlouzarides Shalit [Inria]
Mathurin Massias [Inria]
Arthur Mensch [Univ Paris-Sud, until Oct 2018]
Tuan Binh Nguyen [Inria, from Oct 2018]
Nicolas Prost [Ecole polytechnique, from Oct 2018]
Hugo Richard [Univ Paris-Saclay, from Sep 2018]
David Sabbagh [INSERM, from Oct 2018]

Technical staff

Alexandre Abadie [Inria]
Jérémie Du Boisberranger [Inria, from Jun 2018]
Guillaume Favelier [Inria, from Dec 2018]
Ana Luisa Grilo Pinho [Inria]
Olivier Grisel [Inria]
Guillaume Lemaitre [Inria]
Joan Massich Vall [Inria]
Romuald Menuet [Inria, from Dec 2018]
Antonio Andre Monteiro Manoel [Inria, until Aug 2018]
Juan Jesus Torre Tresols [Inria, from Apr 2018]
Joris Van Den Bossche [Inria]
Kshitij Chawla [Inria, from Jul 2018]
Pierre Glaser [Inria, from Sep 2018]

Interns

L Emir Omar Chehab [Inria, from Mar 2018 until Aug 2018]
Maxime Cuny [Inria, from Oct 2018]
Omar Darwiche Domingues [Inria, from Apr 2018 until Sep 2018]
Antoine Gauthier [Inria, from Sep 2018]
Benoit Sarthou [CEA, from Mar 2018 until Aug 2018]
Meyer Scetbon [Ecole Normale Supérieure Cachan, from Apr 2018 until Aug 2018]
Matias Schmit [Inria, from Mar 2018 until Aug 2018]

Administrative Assistant

Corinne Petitot [Inria]

Visiting Scientists

Lang Chen [Univ. Stanford, from Dec 2018]
Guillermo Gallardo [Univ Côte d'Azur, from Dec 2018]
Matt Higger [Harvard Medical School, from Dec 2018]

External Collaborators

Sergul Aydore [Stevens Institute of Technology, until Apr 2018]
Alexandre Boucaud [CNRS, until Sep 2018]
Danilo Bzdok [Univ. Aachen, from Jul 2018]
Michael Eickenberg [Univ. Berkeley]
Mehdi Rahim [Inria, until Mar 2018]
Venkat Rajagoplan [Institut Telecom ex GET Groupe des Ecoles des Télécommunications , until Jun 2018]
Joseph Salmon [Institut Telecom ex GET Groupe des Ecoles des Télécommunications, HDR]
Moritz Boos [Univ. Oldenburg]

2. Overall Objectives

2.1. Overall Objectives

The Parietal team focuses on mathematical methods for modeling and statistical inference based on neuroimaging data, with a particular interest in machine learning techniques and applications of human functional imaging. This general theme splits into four research axes:

- Modeling for neuroimaging population studies,
- Encoding and decoding models for cognitive imaging,
- Statistical and machine learning methods for large-scale data,
- Compressed-sensing for MRI.

Parietal is also strongly involved in open-source software development in scientific Python (machine learning) and for neuroimaging applications.

3. Research Program

3.1. Inverse problems in Neuroimaging

Many problems in neuroimaging can be framed as forward and inverse problems. For instance, brain population imaging is concerned with the *inverse problem* that consists in predicting individual information (behavior, phenotype) from neuroimaging data, while the corresponding *forward problem* boils down to explaining neuroimaging data with the behavioral variables. Solving these problems entails the definition of two terms: a loss that quantifies the goodness of fit of the solution (does the model explain the data well enough?), and a regularization scheme that represents a prior on the expected solution of the problem. These priors can be used to enforce some properties on the solutions, such as sparsity, smoothness or being piece-wise constant.

Let us detail the model used in typical inverse problem: Let $\mathbf{X}$ be a neuroimaging dataset as an $(n_{subjects}, n_{voxels})$ matrix, where $n_{subjects}$ and n_{voxels} are the number of subjects under study, and the image size respectively, $\mathbf{Y}$ a set of values that represent characteristics of interest in the observed population, written as $(n_{subjects}, n_{features})$ matrix, where $n_{features}$ is the number of characteristics that are tested, and $\mathbf{w}$ an array of shape $(n_{voxels}, n_{features})$ that represents a set of pattern-specific maps. In the first place, we may consider the columns $\mathbf{Y}_1, .., \mathbf{Y}_{n_{features}}$ of Y independently, yielding $n_{features}$ problems to be solved in parallel:

$$\mathbf{Y}_i = \mathbf{X}\mathbf{w}_i + \epsilon_i, \forall i \in \{1, .., n_{features}\},$$

where the vector contains $\mathbf{w}_i$ is the i^{th} row of $\mathbf{w}$. As the problem is clearly ill-posed, it is naturally handled in a regularized regression framework:

$$\widehat{w}_i = \text{argmin}_{w_i} \|\mathbf{Y}_i - \mathbf{X}\mathbf{w}_i\|^2 + \Psi(\mathbf{w}_i), \tag{1}$$

where Ψ is an adequate penalization used to regularize the solution:

$$\Psi(\mathbf{w}; \lambda_1, \lambda_2, \eta_1, \eta_2) = \lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2 + \eta_1 \|\nabla \mathbf{w}\|_{2,1} + \eta_2 \|\nabla \mathbf{w}\|_{2,2} \tag{2}$$

with $\lambda_1,\ \lambda_2,\ \eta_1,\ \eta_2 \geq 0$ (this formulation particularly highlights the fact that convex regularizers are norms or quasi-norms). In general, only one or two of these constraints is considered (hence is enforced with a non-zero coefficient):

- When $\lambda_1 > 0$ only (LASSO), and to some extent, when $\lambda_1, \lambda_2 > 0$ only (elastic net), the optimal solution $\mathbf{w}$ is (possibly very) sparse, but may not exhibit a proper image structure; it does not fit well with the intuitive concept of a brain map.
- Total Variation regularization (see Fig. 1) is obtained for ($\eta_1 > 0$ only), and typically yields a piece-wise constant solution. It can be associated with Lasso to enforce both sparsity and sparse variations.
- Smooth lasso is obtained with ($\eta_2 > 0$ and $\lambda_1 > 0$ only), and yields smooth, compactly supported spatial basis functions.

Note that, while the qualitative aspect of the solutions are very different, the predictive power of these models is often very close.

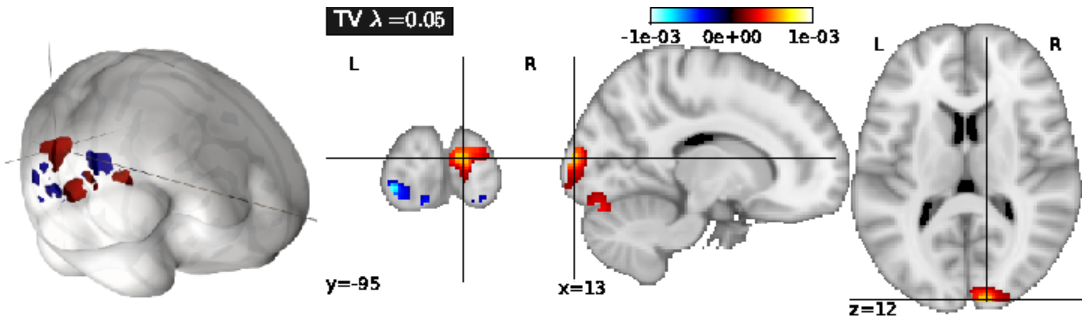


Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problem here is to predict the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the presentation of an image. Learning and test are performed across individuals. Unlike other approaches, Total Variation regularization yields a sparse and well-localized solution that also enjoys high predictive accuracy.

The performance of the predictive model can simply be evaluated as the amount of variance in $\mathbf{Y}_i$ fitted by the model, for each $i \in \{1, .., n_{features}\}$. This can be computed through cross-validation, by *learning* $\widehat{\mathbf{w}}_i$ on some part of the dataset, and then estimating $\|\mathbf{Y}_i - \mathbf{X}\widehat{w}_i\|^2$ using the remainder of the dataset.

This framework is easily extended by considering

- *Grouped penalization*, where the penalization explicitly includes a prior clustering of the features, i.e. voxel-related signals, into given groups. This amounts to enforcing structured priors on the solution.
- *Combined penalizations*, i.e. a mixture of simple and group-wise penalizations, that allow some variability to fit the data in different populations of subjects, while keeping some common constraints.
- *Logistic and hinge regression*, where a non-linearity is applied to the linear model so that it yields a probability of classification in a binary classification problem.
- *Robustness to between-subject variability* to avoid the learned model overly reflecting a few outlying particular observations of the training set. Note that noise and deviating assumptions can be present in both $\mathbf{Y}$ and $\mathbf{X}$
- *Multi-task learning*: if several target variables are thought to be related, it might be useful to constrain the estimated parameter vector $\mathbf{w}$ to have a shared support across all these variables.
  For instance, when one of the variables $\mathbf{Y}_i$ is not well fitted by the model, the estimation of other variables $\mathbf{Y}_j, j \neq i$ may provide constraints on the support of $\mathbf{w}_i$ and thus, improve the prediction of $\mathbf{Y}_i$.

$$\mathbf{Y} = \mathbf{X}\mathbf{w} + \epsilon, \tag{3}$$

then

$$\widehat{w} = \text{argmin}_{\mathbf{w}=(\mathbf{w}_i), i=1..n_f} \sum_{i=1}^{n_f} \|\mathbf{Y_i} - \mathbf{X}\mathbf{w_i}\|^2 + \lambda \sum_{j=1}^{n_{voxels}} \sqrt{\sum_{i=1}^{n_f} \mathbf{w_{i,j}^2}} \tag{4}$$

3.2. Multivariate decompositions

Multivariate decompositions provide a way to model complex data such as brain activation images: for instance, one might be interested in extracting an *atlas of brain regions* from a given dataset, such as regions exhibiting similar activity during a protocol, across multiple protocols, or even in the absence of protocol (during resting-state). These data can often be factorized into spatial-temporal components, and thus can be estimated through *regularized Principal Components Analysis* (PCA) algorithms, which share some common steps with regularized regression.
Let $\mathbf{X}$ be a neuroimaging dataset written as an $(n_{subjects}, n_{voxels})$ matrix, after proper centering; the model reads

$$\mathbf{X} = \mathbf{A}\mathbf{D} + \epsilon, \tag{5}$$

where $\mathbf{D}$ represents a set of n_{comp} spatial maps, hence a matrix of shape (n_{comp}, n_{voxels}), and $\mathbf{A}$ the associated subject-wise loadings. While traditional PCA and independent components analysis (ICA) are limited to reconstructing components $\mathbf{D}$ within the space spanned by the column of $\mathbf{X}$, it seems desirable to add some constraints on the rows of $\mathbf{D}$, that represent spatial maps, such as sparsity, and/or smoothness, as it makes the interpretation of these maps clearer in the context of neuroimaging. This yields the following estimation problem:

$$\text{min}_{\mathbf{D},\mathbf{A}} \|\mathbf{X} - \mathbf{A}\mathbf{D}\|^2 + \Psi(\mathbf{D}) \text{ s.t. } \|\mathbf{A}_i\| = 1 \; \forall i \in \{1..n_{features}\}, \tag{6}$$

where $(\mathbf{A}_i),\ i \in \{1..n_{features}\}$ represents the columns of $\mathbf{A}$. Ψ can be chosen such as in Eq. (2) in order to enforce smoothness and/or sparsity constraints.

The problem is not jointly convex in all the variables but each penalization given in Eq (2) yields a convex problem on $\mathbf{D}$ for $\mathbf{A}$ fixed, and conversely. This readily suggests an alternate optimization scheme, where $\mathbf{D}$ and $\mathbf{A}$ are estimated in turn, until convergence to a local optimum of the criterion. As in PCA, the extracted components can be ranked according to the amount of fitted variance. Importantly, also, estimated PCA models can be interpreted as a probabilistic model of the data, assuming a high-dimensional Gaussian distribution (probabilistic PCA).

Ultimately, the main limitations to these algorithms is the cost due to the memory requirements: holding datasets with large dimension and large number of samples (as in recent neuroimaging cohorts) leads to inefficient computation. To solve this issue, online methods are particularly attractive [24].

3.3. Covariance estimation

Another important estimation problem stems from the general issue of learning the relationship between sets of variables, in particular their covariance. Covariance learning is essential to model the dependence of these variables when they are used in a multivariate model, for instance to study potential interactions among them and with other variables. Covariance learning is necessary to model latent interactions in high-dimensional observation spaces, e.g. when considering multiple contrasts or functional connectivity data.
The difficulties are two-fold: on the one hand, there is a shortage of data to learn a good covariance model from an individual subject, and on the other hand, subject-to-subject variability poses a serious challenge to the use of multi-subject data. While the covariance structure may vary from population to population, or depending on the input data (activation versus spontaneous activity), assuming some shared structure across problems, such as their sparsity pattern, is important in order to obtain correct estimates from noisy data. Some of the most important models are:

- **Sparse Gaussian graphical models**, as they express meaningful conditional independence relationships between regions, and do improve conditioning/avoid overfit.
- **Decomposable models**, as they enjoy good computational properties and enable intuitive interpretations of the network structure. Whether they can faithfully or not represent brain networks is still an open question.
- **PCA-based regularization of covariance** which is powerful when modes of variation are more important than conditional independence relationships.

Adequate model selection procedures are necessary to achieve the right level of sparsity or regularization in covariance estimation; the natural evaluation metric here is the out-of-sample likelihood of the associated Gaussian model. Another essential remaining issue is to develop an adequate statistical framework to test differences between covariance models in different populations. To do so, we consider different means of parametrizing covariance distributions and how these parametrizations impact the test of statistical differences across individuals.

4. Application Domains

4.1. Cognitive neuroscience

4.1.1. Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain , such as: How is the function of cortical territories constrained by anatomy (connectivity) ? How to assess the specificity of brain regions ? How can one characterize reliably inter-subject differences ?

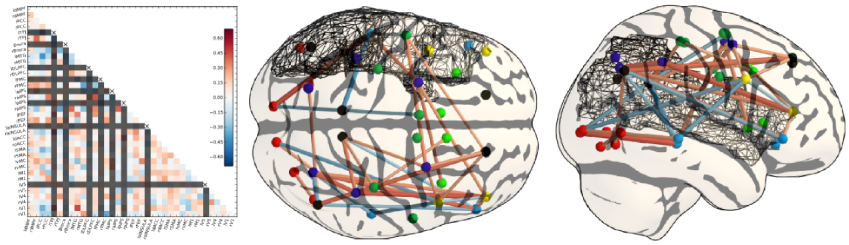


Figure 2. Example of functional connectivity analysis: The correlation matrix describing brain functional connectivity in a post-stroke patient (lesion volume outlined as a mesh) is compared to a group of control subjects. Some edges of the graphical model show a significant difference, but the statistical detection of the difference requires a sophisticated statistical framework for the comparison of graphical models.

4.1.2. Analysis of brain Connectivity

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

4.1.3. Modeling of brain processes (MEG)

While fMRI has been very useful in defining the function of regions at the mm scale, Magneto-encephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

4.1.4. Current challenges in human neuroimaging (acquisition+analysis)

Human neuroimaging targets two major goals: *i)* the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii)* the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial [1] resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.
- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements [2] and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [70], and

[1] and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

[2] The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [71]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as NeuroVault or OpenfMRI. A framework for integrating such datasets is however still missing.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- Pierre Ablin got a best student paper award at the LVA-ICA conference for his paper [34].
- First PhD prize from STIC doctoral school for Tom Dupré la Tour.

6. New Software and Platforms

6.1. Mayavi

FUNCTIONAL DESCRIPTION: Mayavi is the most used scientific 3D visualization Python software. Mayavi can be used as a visualization tool, through interactive command line or as a library. It is distributed under Linux through Ubuntu, Debian, Fedora and Mandriva, as well as in PythonXY and EPD Python scientific distributions. Mayavi is used by several software platforms, such as PDE solvers (fipy, sfepy), molecule visualization tools and brain connectivity analysis tools (connectomeViewer).

- Contact: Gaël Varoquaux
- URL: http://mayavi.sourceforge.net/

6.2. MedInria

KEYWORDS: Visualization - DWI - Health - Segmentation - Medical imaging
SCIENTIFIC DESCRIPTION: medInria aims at creating an easily extensible platform for the distribution of research algorithms developed at Inria for medical image processing. This project has been funded by the D2T (ADT MedInria-NT) in 2010, renewed in 2012. A fast-track ADT was awarded in 2017 to transition the software core to more recent dependencies and study the possibility of a consortium creation.The Visages team leads this Inria national project and participates in the development of the common core architecture and features of the software as well as in the development of specific plugins for the team's algorithm.
FUNCTIONAL DESCRIPTION: MedInria is a free software platform dedicated to medical data visualization and processing.

- Participants: Maxime Sermesant, Olivier Commowick and Théodore Papadopoulo
- Partners: HARVARD Medical School - IHU - LIRYC - NIH
- Contact: Olivier Commowick
- URL: http://med.inria.fr

6.3. Nilearn

NeuroImaging with scikit learn
KEYWORDS: Health - Neuroimaging - Medical imaging

FUNCTIONAL DESCRIPTION: NiLearn is the neuroimaging library that adapts the concepts and tools of scikit-learn to neuroimaging problems. As a pure Python library, it depends on scikit-learn and nibabel, the main Python library for neuroimaging I/O. It is an open-source project, available under BSD license. The two key components of NiLearn are i) the analysis of functional connectivity (spatial decompositions and covariance learning) and ii) the most common tools for multivariate pattern analysis. A great deal of efforts has been put on the efficiency of the procedures both in terms of memory cost and computation time.

- Participants: Alexandre Abraham, Alexandre Gramfort, Bertrand Thirion, Elvis Dohmatob, Fabian Pedregosa Izquierdo, Gaël Varoquaux, Loïc Estève, Michael Eickenberg and Virgile Fritsch
- Contact: Bertrand Thirion
- URL: http://nilearn.github.io/

6.4. PyHRF

KEYWORDS: Medical imaging - Health - Brain - IRM - Neurosciences - Statistic analysis - FMRI
SCIENTIFIC DESCRIPTION: Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that allows the non-invasive study of brain function. It is based on the hemodynamic variations induced by changes in cerebral synaptic activity following sensory or cognitive stimulation. The measured signal depends on the variation of blood oxygenation level (BOLD signal) which is related to brain activity: a decrease in deoxyhemoglobin concentration induces an increase in BOLD signal. The BOLD signal is delayed with respect to changes in synaptic activity, which can be modeled as a convolution with the Hemodynamic Response Function (HRF) whose exact form is unknown and fluctuates with various parameters such as age, brain region or physiological conditions. In this work we propose to analyze fMRI data using a Joint Detection-Estimation (JDE) approach. It jointly detects cortical activation and estimates the HRF. In contrast to existing tools, PyHRF estimates the HRF instead of considering it as a given constant in the entire brain.
FUNCTIONAL DESCRIPTION: As part of fMRI data analysis, PyHRF provides a set of tools for addressing the two main issues involved in intra-subject fMRI data analysis : (i) the localization of cerebral regions that elicit evoked activity and (ii) the estimation of the activation dynamics also referenced to as the recovery of the Hemodynamic Response Function (HRF). To tackle these two problems, PyHRF implements the Joint Detection-Estimation framework (JDE) which recovers parcel-level HRFs and embeds an adaptive spatio-temporal regularization scheme of activation maps.
NEWS OF THE YEAR: The framework to perform software tests has been further developed. Some unitary tests have been set.

- Participants: Aina Frau Pascual, Christine Bakhous, Florence Forbes, Jaime Eduardo Arias Almeida, Laurent Risser, Lotfi Chaari, Philippe Ciuciu, Solveig Badillo, Thomas Perret and Thomas Vincent
- Partners: CEA - NeuroSpin
- Contact: Florence Forbes
- Publications: Frontiers in Neuroinformatics Flexible multivariate hemodynamics fMRI data analyses and simulations with PyHRF - Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach - A Bayesian Non-Parametric Hidden Markov Random Model for Hemodynamic Brain Parcellation
- URL: http://pyhrf.org

6.5. Scikit-learn

KEYWORDS: Regession - Clustering - Learning - Classification - Medical imaging
SCIENTIFIC DESCRIPTION: Scikit-learn is a Python module integrating classic machine learning algorithms in the tightly-knit scientific Python world. It aims to provide simple and efficient solutions to learning problems, accessible to everybody and reusable in various contexts: machine-learning as a versatile tool for science and engineering.

FUNCTIONAL DESCRIPTION: Scikit-learn can be used as a middleware for prediction tasks. For example, many web startups adapt Scikitlearn to predict buying behavior of users, provide product recommendations, detect trends or abusive behavior (fraud, spam). Scikit-learn is used to extract the structure of complex data (text, images) and classify such data with techniques relevant to the state of the art.

Easy to use, efficient and accessible to non datascience experts, Scikit-learn is an increasingly popular machine learning library in Python. In a data exploration step, the user can enter a few lines on an interactive (but non-graphical) interface and immediately sees the results of his request. Scikitlearn is a prediction engine . Scikit-learn is developed in open source, and available under the BSD license.

- Participants: Alexandre Gramfort, Bertrand Thirion, Fabian Pedregosa Izquierdo, Gaël Varoquaux, Loïc Estève, Michael Eickenberg and Olivier Grisel
- Partners: CEA - Logilab - Nuxeo - Saint Gobain - Tinyclues - Telecom Paris
- Contact: Olivier Grisel
- URL: http://scikit-learn.org

6.6. MODL

Massive Online Dictionary Learning
KEYWORDS: Pattern discovery - Machine learning
FUNCTIONAL DESCRIPTION: Matrix factorization library, usable on very large datasets, with optional sparse and positive factors.

- Participants: Arthur Mensch, Gaël Varoquaux, Bertrand Thirion and Julien Mairal
- Contact: Arthur Mensch
- Publications: Subsampled online matrix factorization with convergence guarantees - Stochastic Subsampling for Factorizing Huge Matrices
- URL: http://github.com/arthurmensch/modl

6.7. MNE

MNE-Python
KEYWORDS: Neurosciences - EEG - MEG - Signal processing - Machine learning
FUNCTIONAL DESCRIPTION: Open-source Python software for exploring, visualizing, and analyzing human neurophysiological data: MEG, EEG, sEEG, ECoG, and more.

RELEASE FUNCTIONAL DESCRIPTION: http://martinos.org/mne/stable/what_new.html

- Partners: HARVARD Medical School - New York University - University of Washington - CEA - Aalto university - Telecom Paris - Boston University - UC Berkeley
- Contact: Alexandre Gramfort
- URL: http://martinos.org/mne/

7. New Results

7.1. Reducing the number of samples in spatiotemporal dMRI acquisition design

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

More information can be found in [12]

7.2. Robust EEG-based cross-site and cross-protocol classification of states of consciousness

Determining the state-of-consciousness in patients with disorders-of-consciousness (DOC) is a challenging practical and theoretical problem. Recent findings suggest that multiple markers of brain activity extracted from the electroencephalogram (EEG) may index the state of consciousness in the human brain. Furthermore, machine learning has been found to optimize their capacity to discriminate different states of consciousness in clinical practice. However, it is unknown how dependable these EEG-markers are in the face of signal variability due to different EEG-configurations, EEG-protocols and subpopulations from different centers encountered in practice. In our recent paper [11] we addressed the following questions: What is the impact of the EEG configuration (selection of sensors, duration of EEG used)? Do models based on current EEG-markers achieve prospective generalization on independent data from other EEG protocols and other hospitals? Are single markers sufficiently powerful and when does multivariate classification provide the clearest advantage? For summary of methods and approahc see Figure4. Our results highlight the effectiveness of classical well-studied EEG-signatures such as alpha [8-12Hz] and theta [5-7Hz] frequency band oscillations for detecting consciousness when combined with machine learning. While univariate predictive models achieved good performance, multivariate models showed better generalization capacity and increased robustness to different types of noise while mitigating the impact of the EEG-configuration. Our findings suggest that pooling data over multiple centers for predictive modeling of DOC is a concrete possibility and can become a promising alley for the field of cognitive neurology.

7.3. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series

Sleep stage classification constitutes an important preliminary exam in the diagnosis of sleep disorders. It is traditionally performed by a sleep expert who assigns to each 30 s of signal a sleep stage, based on the visual inspection of signals such as electroencephalograms (EEG), electrooculograms (EOG), electrocardiograms (ECG) and electromyograms (EMG). We introduce here the first deep learning approach for sleep stage classification that learns end-to-end without computing spectrograms or extracting hand-crafted features, that exploits all multivariate and multimodal Polysomnography (PSG) signals (EEG, EMG and EOG), and that can exploit the temporal context of each 30 s window of data. For each modality the first layer learns linear

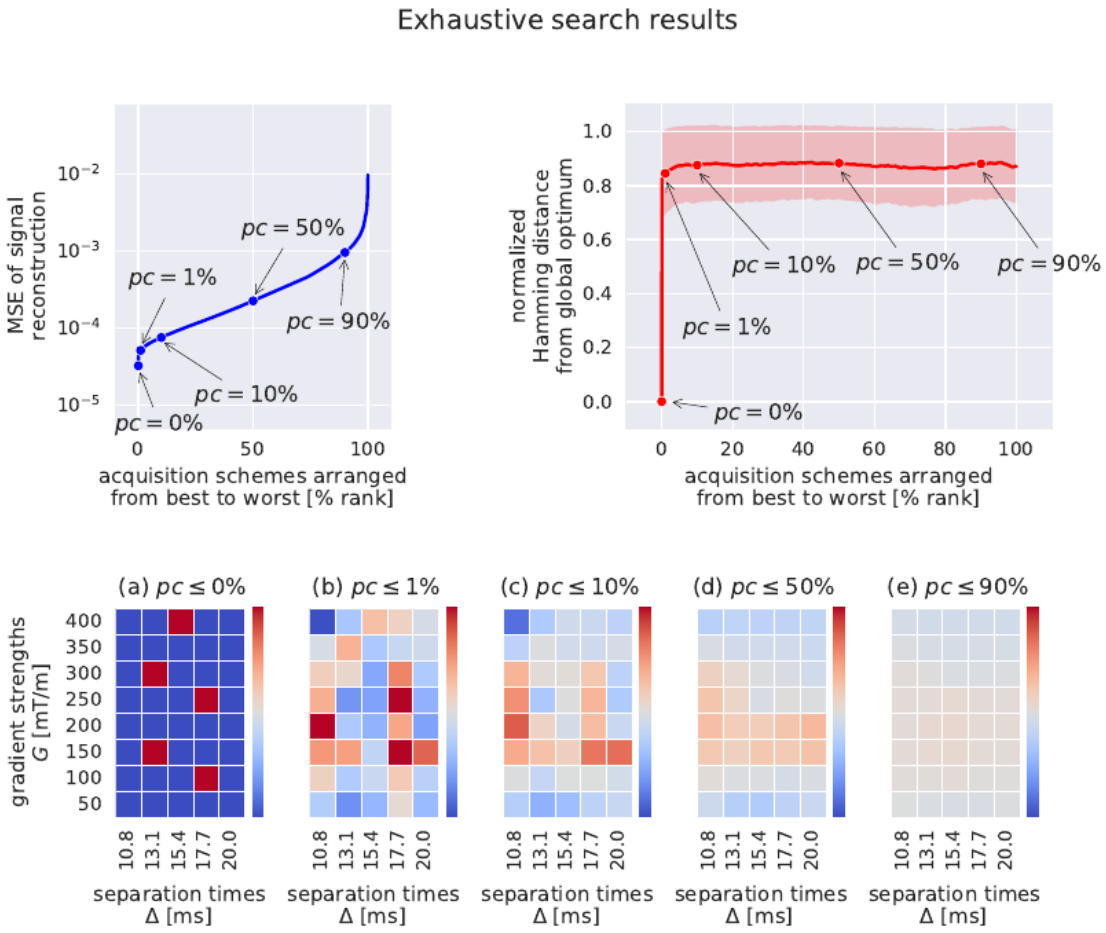


Figure 3. Exhaustive search results of the optimization by shells for the in silico experiment with nmax = 100. The plots at the top present all the 658,008 feasible acquisition schemes arranged from best to worst, illustrating the mean squared errors (MSEs) of signal reconstruction (top-left plot) and the normalized Hamming distances from the global optimum ± 1 standard deviation (top-right). In order to visualize the analyzed (G, Δ) parameter space, the percentiles pc = 0%, 1%, 10%, 50%, 90% are annotated on both plots, showing respectively the global optimum, the top 1% solutions, the top 10% solutions, etc. The corresponding cumulative averages of acquisition schemes are depicted in the heat maps at the bottom. The colors reflect the likelihood of a given (G, Δ) pair in the scheme. The heat maps for pc ≤ 0% and pc ≤ 1% represent, respectively, the global optimum and its proximity. The interval between pc = 10% and pc = 90% contains a huge spectrum of schemes with similar MSEs and almost equally large distances from the global optimum.

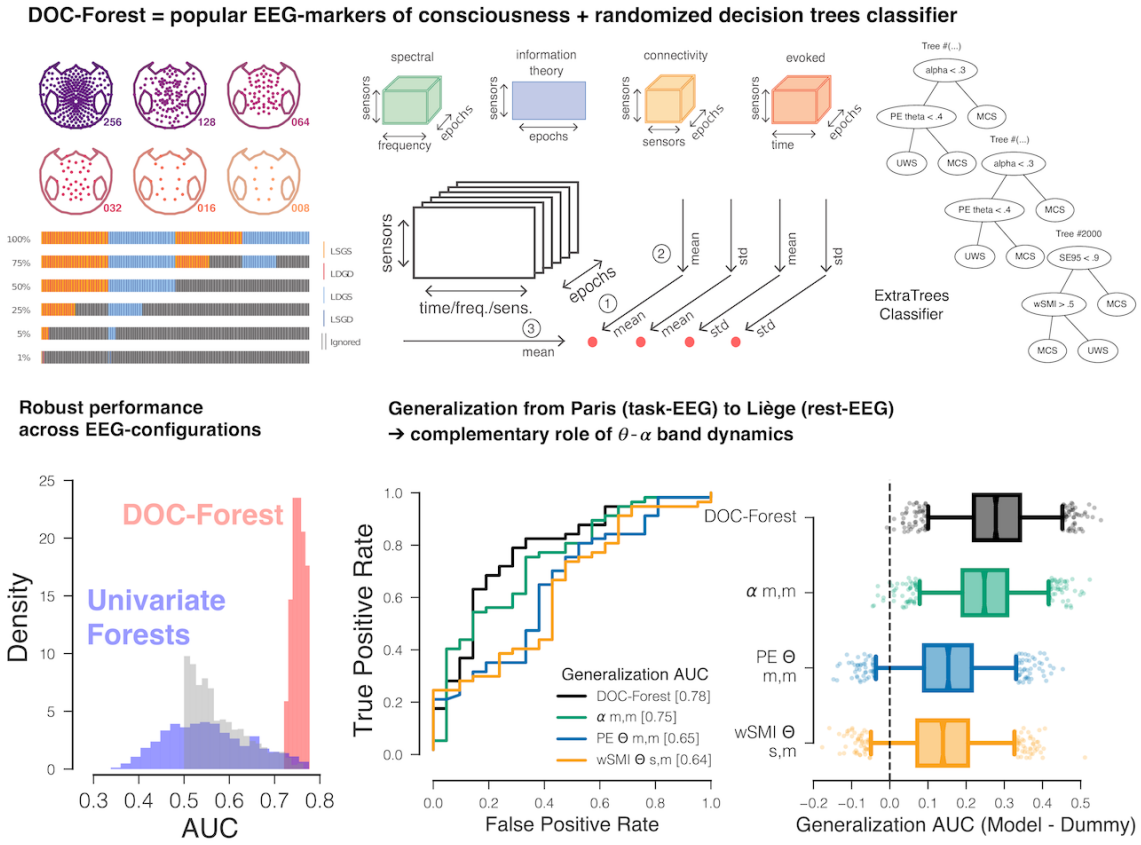


Figure 4. We probed the robustness and validity of EEG-markers of consciousness. Using the robust Extra-Trees algorithm (Geurts, Ernst, & Wehenkel, 2006) we developed a classifier trained to differentiate UWS from MCS patients. This classifier (named "DOC-forest") was trained and tested using 28 potential EEG-markers of consciousness (112 features) from 249 patients recorded at the Paris Pitié-Salpêtrière and 78 patients from the University Hospital of Liège. We used the MNE-Python software for EEG processing and the scikit-learn package for machine learning. Our results show that optimally combining multiple EEG-markers of states of consciousness using machine learning enables robust generalization across EEG-configurations, EEG-protocols and sites. Our recipe for extracting biomarkers is available on Github: https://nice-tools.github.io/nice. For a neuroscientific discussion of our work see the accompanying commentary article by Sokoliuk and Cruse (https://doi.org/10.1093/brain/awy267).

spatial filters that exploit the array of sensors to increase the signal-to-noise ratio, and the last layer feeds the learnt representation to a softmax classifier. Our model is compared to alternative automatic approaches based on convolutional networks or decisions trees. Results obtained on 61 publicly available PSG records with up to 20 EEG channels demonstrate that our network architecture yields state-of-the-art performance. Our study reveals a number of insights on the spatio-temporal distribution of the signal of interest: a good trade-off for optimal classification performance measured with balanced accuracy is to use 6 EEG with 2 EOG (left and right) and 3 EMG chin channels. Also exploiting one minute of data before and after each data segment offers the strongest improvement when a limited number of channels is available. As sleep experts, our system exploits the multivariate and multimodal nature of PSG signals in order to deliver state-of-the-art classification performance with a small computational cost.

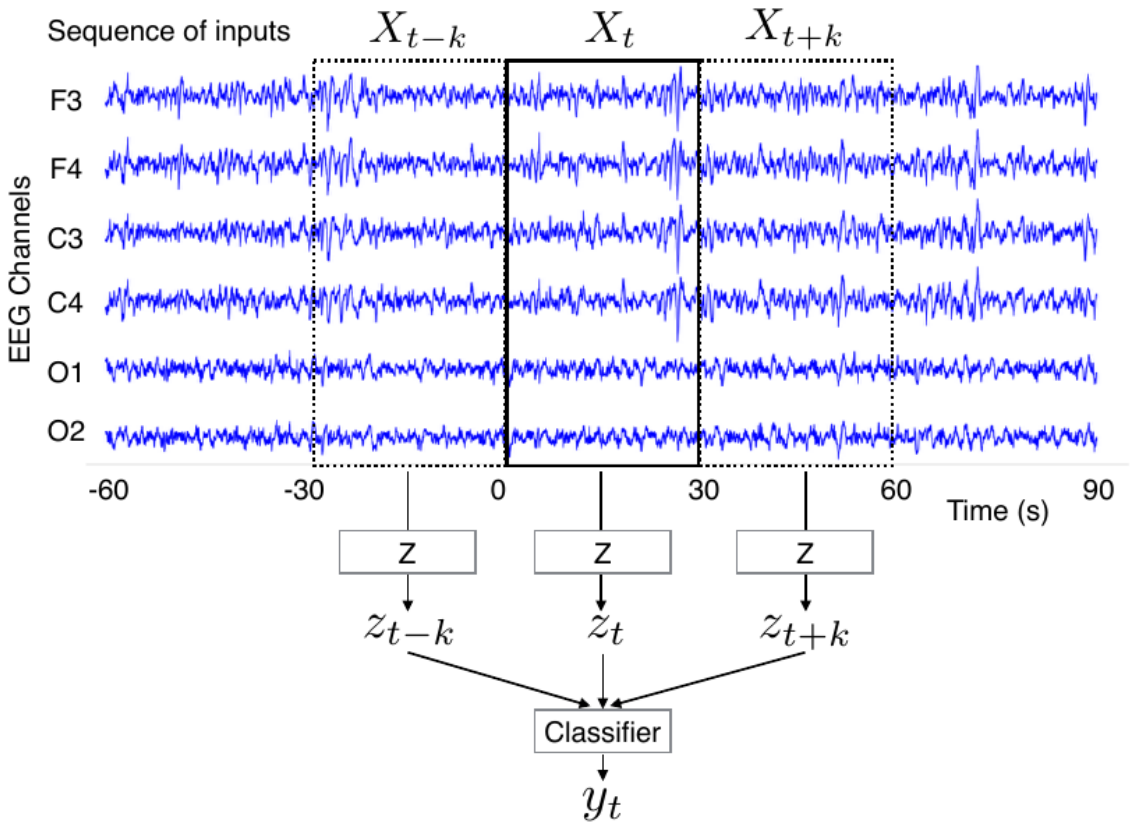


Figure 5. Time distributed architecture to process a sequence of inputs $S_t^k = \{X_{t-k}, \cdots, X_t, \cdots, X_{t+k}\}$ with k = 1. X_k stands for the multivariate input data over 30 s that is fed into the feature extractor Z. Features are extracted from consecutive 30 s samples: X_{t-k}, ..., X_t, ..., X_{t+k}. Then the obtained features are aggregated $[z_{t-k}, \cdots, z_t, \cdots, z_{t+k}]$. The resulting aggregation of features is finally fed into a classifier to predict the label y_t associated with the sample X_t.

More information can be found in [8].

7.4. Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping

Functional Magnetic Resonance Imaging (fMRI) has furthered brain mapping on perceptual, motor, as well as higher-level cognitive functions. However, to date, no data collection has systematically addressed the functional mapping of cognitive mechanisms at a fine spatial scale. The Individual Brain Charting (IBC) project stands for a high-resolution multi-task fMRI dataset that intends to provide the objective basis toward a comprehensive functional atlas of the human brain. The data refer to a cohort of 12 participants performing many different tasks. The large amount of task-fMRI data on the same subjects yields a precise mapping of the underlying functions, free from both inter-subject and inter-site variability. The present article gives a detailed description of the first release of the IBC dataset. It comprises a dozen of tasks, addressing both low- and high-level cognitive functions. This openly available dataset is thus intended to become a reference for cognitive brain mapping.

More information can be found in [25]

7.5. Atlases of cognition with large-scale brain mapping

To map the neural substrate of mental function, cognitive neuroimaging relies on controlled psychological manipulations that engage brain systems associated with specific cognitive processes. In order to build comprehensive atlases of cognitive function in the brain, it must assemble maps for many different cognitive processes, which often evoke overlapping patterns of activation. Such data aggregation faces contrasting goals: on the one hand finding correspondences across vastly different cognitive experiments, while on the other hand precisely describing the function of any given brain region. Here we introduce a new analysis framework that tackles these difficulties and thereby enables the generation of brain atlases for cognitive function. The approach leverages ontologies of cognitive concepts and multi-label brain decoding to map the neural substrate of these concepts. We demonstrate the approach by building an atlas of functional brain organization based on 30 diverse functional neuroimaging studies, totaling 196 different experimental conditions. Unlike conventional brain mapping, this functional atlas supports robust reverse inference: predicting the mental processes from brain activity in the regions delineated by the atlas. To establish that this reverse inference is indeed governed by the corresponding concepts, and not idiosyncrasies of experimental designs, we show that it can accurately decode the cognitive concepts recruited in new tasks. These results demonstrate that aggregating independent task-fMRI studies can provide a more precise global atlas of selective associations between brain and cognition.

More information can be found in [28].

7.6. Celer: a Fast Solver for the Lasso with Dual Extrapolation

Convex sparsity-inducing regularizations are ubiquitous in high-dimensional machine learning, but solving the resulting optimization problems can be slow. To accelerate solvers, state-of-the-art approaches consist in reducing the size of the optimization problem at hand. In the context of regression, this can be achieved either by discarding irrelevant features (screening techniques) or by prioritizing features likely to be included in the support of the solution (working set techniques). Convex duality comes into play at several steps in these techniques. Here, we propose an extrapolation technique starting from a sequence of iterates in the dual that leads to the construction of improved dual points. This enables a tighter control of optimality as used in stopping criterion, as well as better screening performance of Gap Safe rules. Finally, we propose a working set strategy based on an aggressive use of Gap Safe screening rules. Thanks to our new dual point construction, we show significant computational speedups on multiple real-world problems compared to alternative state-of-the-art coordinate descent solvers.

More information can be found in [54]. Code can be found at https://mathurinm.github.io/celer/.

7.7. Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals

Frequency-specific patterns of neural activity are traditionally interpreted as sustained rhythmic oscillations, and related to cognitive mechanisms such as attention, high level visual processing or motor control. While

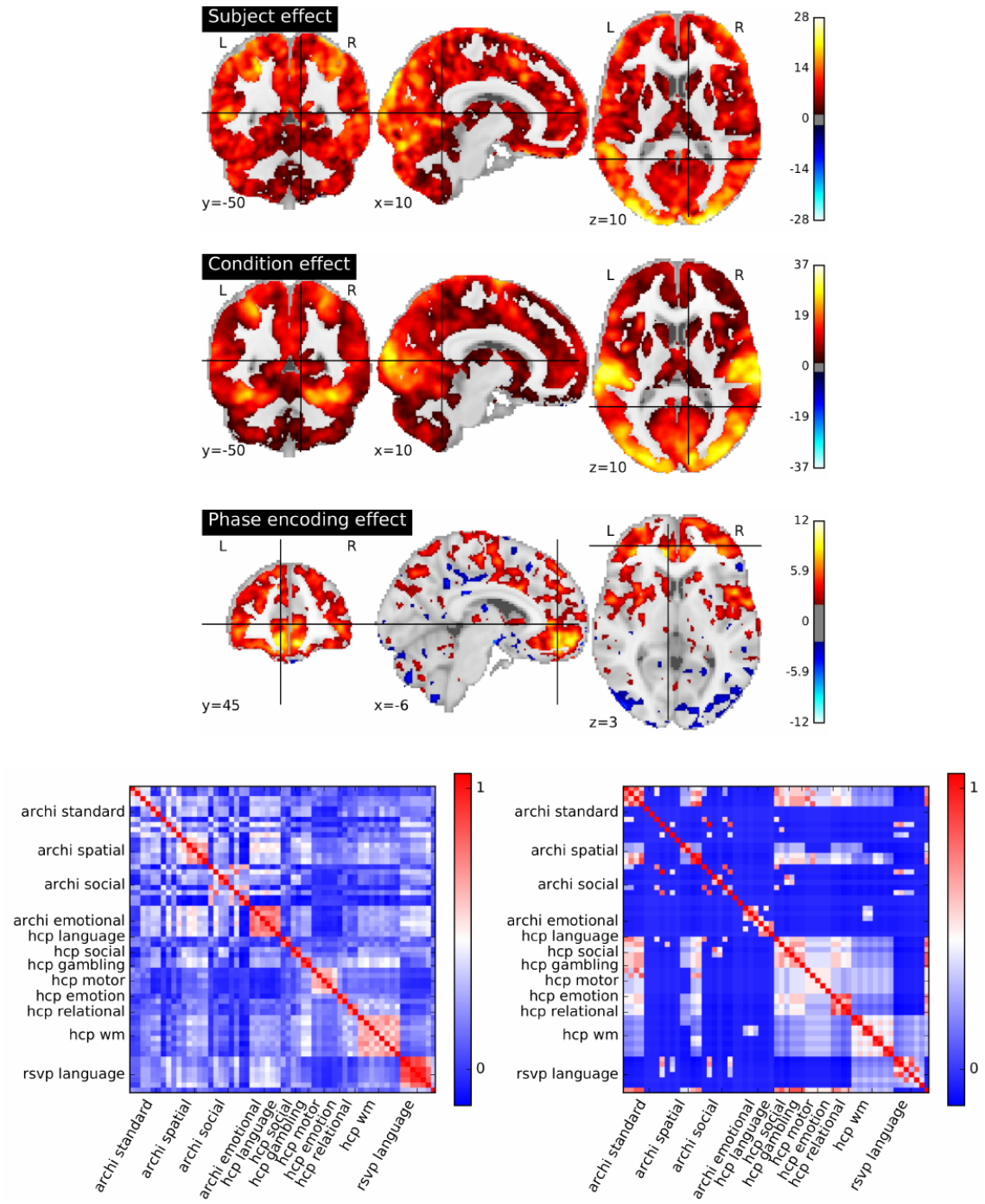


Figure 6. Overview of information conveyed by activation maps resulting from a first-level analysis. (top) Global effects of experimental subject condition, and phase-encoding direction. A per-voxel ANOVA breaks the variance of the set of brain maps into subject, experimental condition, and phase-encoding direction values. All maps are given in z-scale and thresholded at an FDR level of 0.05. (Bottom) Focusing on condition effect, the similarity between condition-related maps, averaged across subjects (left) is clearly related to the dissimilarity of the conditions, when these are characterized in terms of the Cognitive Atlas (right).

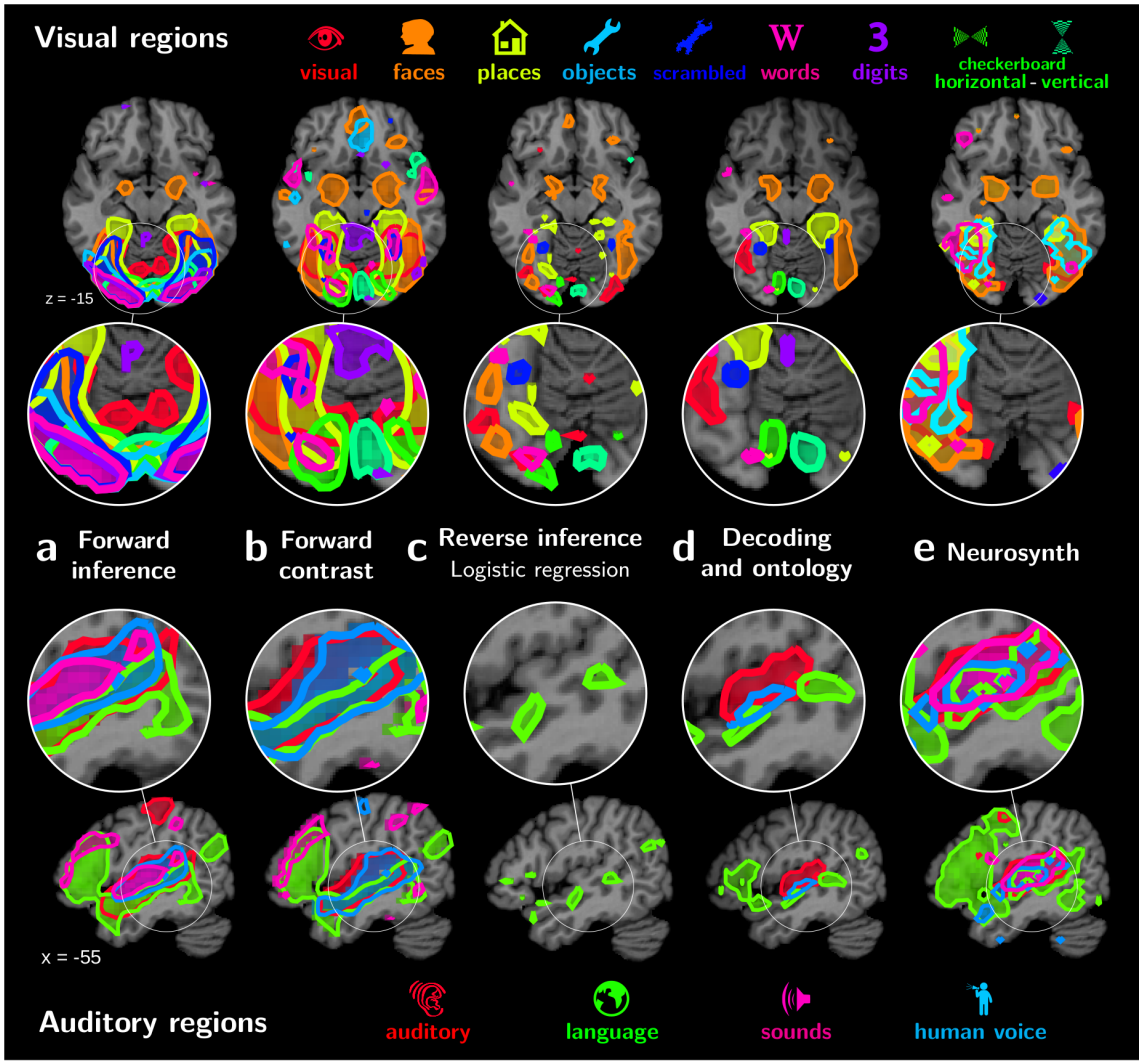


Figure 7. Different functional atlases – Regions outlined using different functional mapping approaches, from left to right: a. forward term mapping; b. forward inference with ontology contrasts (standard analysis); c. reverse inference with logistic regression; d. NeuroSynth reverse inference; and e. our approach, mapping with decoding and an ontology. The top part shows visual regions, and the lower one auditory regions in the left hemisphere. Forward term mapping outlines overlapping regions, as brain responses capture side effects such as the stimulus modality: for visual and auditory regions every cognitive term is represented in the corresponding primary cortex. Forward mapping using contrasts removes the overlap in primary regions, but a large overlap persists in mid-level regions, as control conditions are not well matched across studies. Standard reverse inference, specific to a term, creates overly sparse regions though with little overlap. Reverse inference with Neurosynth also displays large overlap in mid-level regions. Finally, ontology-based decoding maps recover known functional areas the visual and auditory cortices.

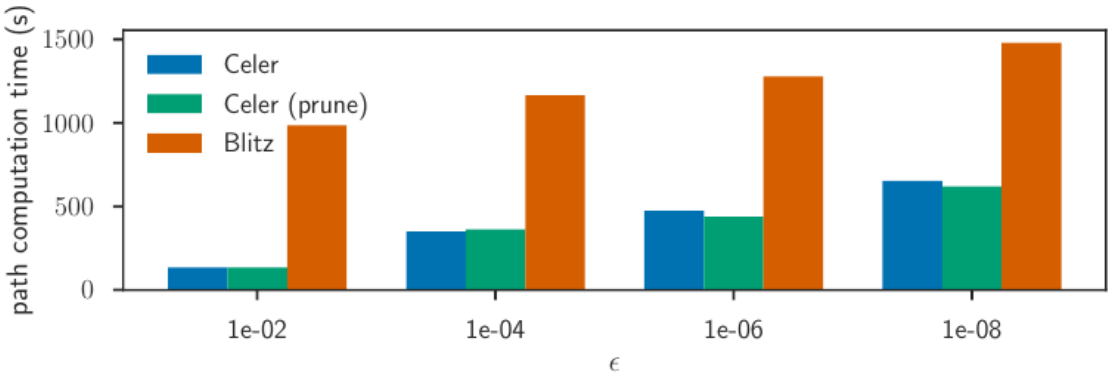


Figure 8. Times to solve the Lasso path to precision ϵ for 100 values of λ, from λ_{max} to $\lambda_{max}/100$, on the Finance data. CELER outperforms BLITZ . Both safe and prune versions behave similarly.

alpha waves (8–12 Hz) are known to closely resemble short sinusoids, and thus are revealed by Fourier analysis or wavelet transforms, there is an evolving debate that electromagnetic neural signals are composed of more complex waveforms that cannot be analyzed by linear filters and traditional signal representations. In this work, we propose to learn dedicated representations of such recordings using a multivariate convolutional sparse coding (CSC) algorithm. Applied to electroencephalography (EEG) or magnetoencephalography (MEG) data, this method is able to learn not only prototypical temporal waveforms, but also associated spatial patterns so their origin can be localized in the brain. Our algorithm is based on alternated minimization and a greedy coordinate descent solver that leads to state-of-the-art running time on long time series. To demonstrate the implications of this method, we apply it to MEG data and show that it is able to recover biological artifacts. More remarkably, our approach also reveals the presence of non-sinusoidal mu-shaped patterns, along with their topographic maps related to the somatosensory cortex.

More information can be found in [52]. Code can be found at https://alphacsc.github.io/.

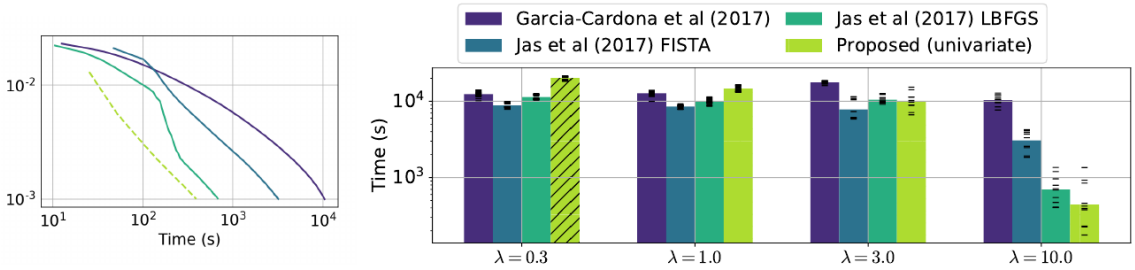
7.8. Stochastic Subsampling for Factorizing Huge Matrices

We present a matrix-factorization algorithm that scales to input matrices with both huge number of rows and columns. Learned factors may be sparse or dense and/or non-negative, which makes our algorithm suitable for dictionary learning, sparse component analysis, and non-negative matrix factorization. Our algorithm streams matrix columns while subsampling them to iteratively learn the matrix factors. At each iteration, the row dimension of a new sample is reduced by subsampling, resulting in lower time complexity compared to a simple streaming algorithm. Our method comes with convergence guarantees to reach a stationary point of the matrix-factorization problem. We demonstrate its efficiency on massive functional Magnetic Resonance Imaging data (2 TB), and on patches extracted from hyperspectral images (103 GB). For both problems, which involve different penalties on rows and columns, we obtain significant speed-ups compared to state-of-the-art algorithms.

More information can be found in [24].

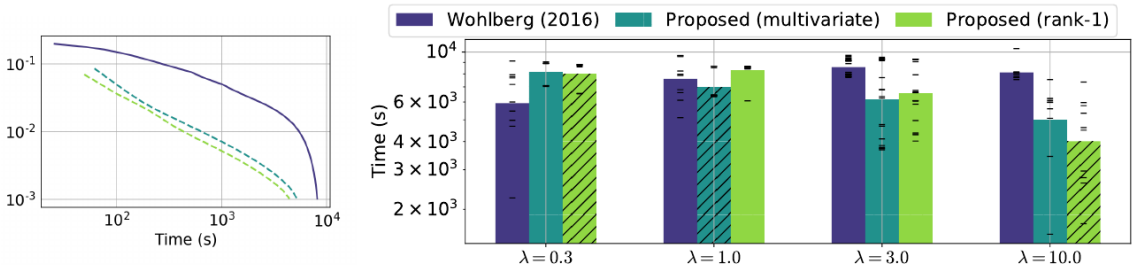
7.9. Text to brain: predicting the spatial distribution of neuroimaging observations from text reports

Despite the digital nature of magnetic resonance imaging, the resulting observations are most frequently reported and stored in text documents. There is a trove of information untapped in medical health records,



(a) $\lambda = 10$ (univariate).

(b) Time to reach a precision of 0.001 (univariate).



(c) $\lambda = 10$ (multivariate).

(d) Time to reach a precision of 0.001 (multivariate).

Figure 9. Comparison of state-of-the-art univariate (a, b) and multivariate (c, d) methods with our approach. (a) Convergence plot with the objective function relative to the obtained minimum, as a function of computational time. (b) Time taken to reach a relative precision of 10^{-3}*, for different regularization parameters* λ*. (c, d) Same as (a, b) in the multivariate setting P=5.*

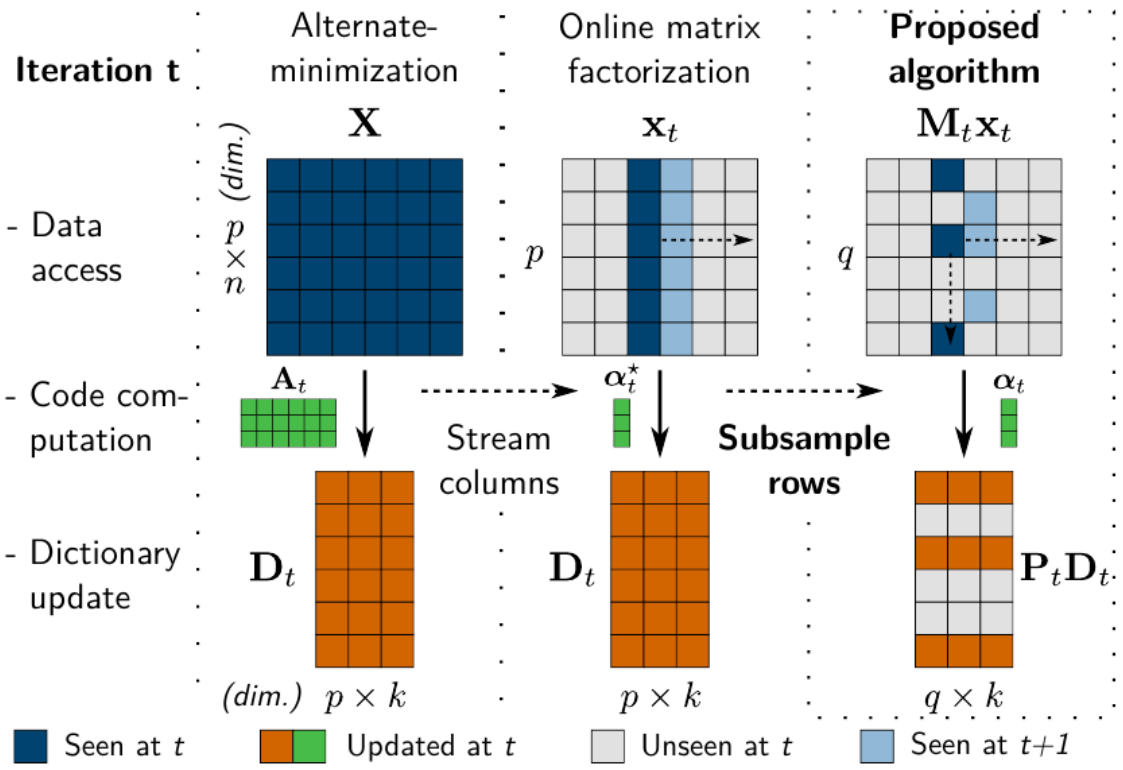


Figure 10. Stochastic subsampling further improves online matrix factorizationhandle datasets with large number of columns and rows. X is the input $p \times n$ matrix, D_t and A_t are respectively the dictionary and code at time t.

case reports, and medical publications. In this paper, we propose to mine brain medical publications to learn the spatial distribution associated with anatomical terms. The problem is formulated in terms of minimization of a risk on distributions which leads to a least-deviation cost function. An efficient algorithm in the dual then learns the mapping from documents to brain structures. Empirical results using coordinates extracted from the brain-imaging literature show that i) models must adapt to semantic variation in the terms used to describe a given anatomical structure, ii) voxel-wise parameterization leads to higher likelihood of locations reported in unseen documents, iii) least-deviation cost outperforms least-square. As a proof of concept for our method, we use our model of spatial distributions to predict the distribution of specific neurological conditions from text-only reports.

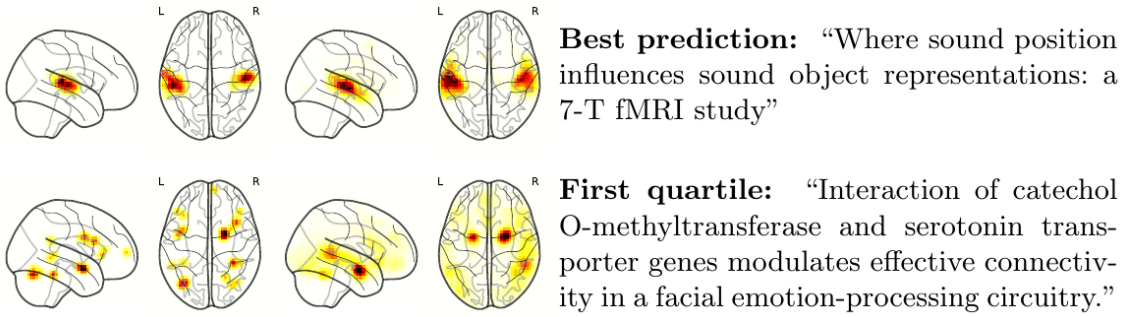


Figure 11. True probability density function (estimated with kernel density estimator) and the prediction for the articles which obtained respectively the best and the first- quartile scores.

More information can be found in [37].

7.10. Similarity encoding for learning with dirty categorical variables

For statistical learning, categorical variables in a table are usually considered as discrete entities and encoded separately to feature vectors, e.g., with one-hot encoding. "Dirty" non-curated data gives rise to categorical variables with a very high cardinality but redundancy: several categories reflect the same entity. In databases, this issue is typically solved with a deduplication step. We show that a simple approach that exposes the redundancy to the learning algorithm brings significant gains. We study a generalization of one-hot encoding, similarity encoding, that builds feature vectors from similarities across categories. We perform a thorough empirical validation on non-curated tables, a problem seldom studied in machine learning. Results on seven real-world datasets show that similarity encoding brings significant gains in prediction in comparison with known encoding methods for categories or strings, notably one-hot encoding and bag of character n-grams. We draw practical recommendations for encoding dirty categories: 3-gram similarity appears to be a good choice to capture morphological resemblance. For very high-cardinality, dimensionality reduction significantly reduces the computational cost with little loss in performance: random projections or choosing a subset of prototype categories still outperforms classic encoding approaches.

More information can be found in [7].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

In 2018, a CIFRE PhD thesis was launched with the Canadian company Interaxon https://choosemuse.com. This contract supports the PhD thesis of Hubert Banville.

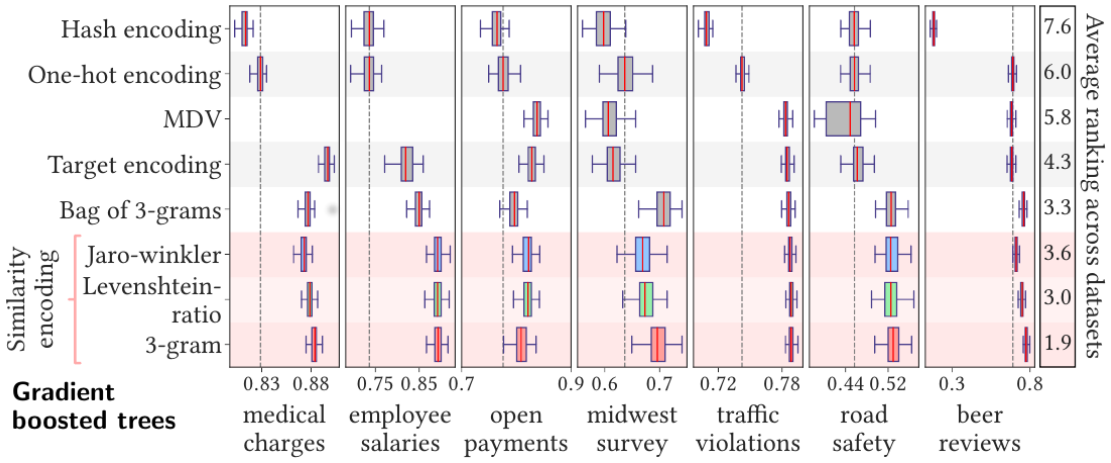


Figure 12. Performance of different encoding methods in a gradient boosting classification task. Each box-plot summarizes the prediction scores of 100 random splits (with 80% of the samples for training and 20% for testing). For all datasets, the prediction score is upper bounded by 1 (a higher score means a better prediction). The right side of the figure indicates the average ranking across datasets for each method. The vertical dashed line indicates the median value of the one-hot encoding method.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Inserm-Inria project

This project is funded by the joint Inserm and Inria program ‘médecine numérique’ and is conducted in collaborations with our clinical partners from the Lariboisière hospital, Inserm uni U942 BioCANVAS (Biomarkers in Cardio-Neuro-VAScular diseases). It supports the PhD thesis of David Sabbagh.

Participants:

- Denis Engemann [coordinator, co-advisor]
- Alexandre Gramfort [thesis director, co-advisor]
- Etienne Gayat [clinical collaborator, co-advisor]
- Fabrice Vallée [clinical collaborator]
- David Sabbagh [PhD Student]

Post-operative delirium (POD) is a potential complication of anesthesia during surgery. It is often associated with adverse outcomes and is aggravated by aging. In elderly patients, post-operative complications have been estimated to incur tens of million US dollars of costs each year in the United States by prolonging hospitalization and potentially affecting health prognosis. Recent studies suggest that POD can already be prevented by improving electrophysiological monitoring of anesthesia depth and individual dosage of anesthetic agents. Doing so probably minimizes the time patients spend in a coma-like state that manifests itself in isoelectric burst suppression, an electroencephalogram (EEG) pattern characterized by alternation between quiescence and high-amplitude bursts, and causally linked to POD. However, such an enterprise, currently, depends on the trained clinical electrophysiologist and guidance by commercially provided EEG indices of states of consciousness. One such metric is the bispectral index (BIS), which, like other related

metrics, does not explicitly take into account baseline changes related to normative aging and may therefore be biased when used naively.

While electrophysiological signatures of aging (e.g. drop in Alpha and Gamma band power), states of consciousness (e.g. drop in Theta band long-range connectivity) and drug response (e.g. anteriorization of alpha band power in propofol anesthesia) have been separately investigated in the past years, their common denominators are not known. It is therefore difficult to detect individual risk, choose the optimal dosage, and automate anesthesia monitoring readily for any patient in any hospital.

The goal of this research project is to build statistical models that enable prediction of burst suppression and subsequent POD by exploiting diverse EEG-signatures of states of consciousness in the context of aging. We approach this challenge by recasting it as a problem of learning brain-age from the point of view of electrophysiology of consciousness.

9.1.2. CoSmic project

Participants: Philippe Ciuciu [Correspondant], Carole Lazarus, Loubna El Gueddari.

This project is funded by CEA DRF-Impulsion.

This is a collaborative project with Jean-Luc Stark, (CEA) funded by the DRF-impulsion CEA program.

Compressed Sensing is a recent theory in maths that allows the perfect recovery of signals or images from compressive acquisition scenarios. This approach has been popularized in MRI over the last decade as well as in astrophysics (noticeably in radio-astronomy). So far, both of these fields have developed skills in CS separately. The aim of the COSMIC project is to foster collaborations between CEA experts in MRI (Parietal team within NeuroSpin) and in astrophysics (CosmoStat lab within the Astrophysics Department). These interactions will allow us to share different expertise in order to improve image quality, either in MRI or in radio-astronomy (thanks to the interferometry principle). In this field, given the data delivered by radio-telescopes, the goal consists in extracting high temporal resolution information in order to study fast transient events.

9.1.3. Metacog

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Jérome Dockès.

This project is funded by Digiteo.

This is a Digicosme project (2016-2019) and a collaboration with Fabian Suchanek (Telecom Paritech).

Understanding how cognition emerges from the billions of neurons that constitute the human brain is a major open problem in science that could bridge natural science –biology– to humanities –psychology. Psychology studies performed on humans with functional Magnetic Resonance Imaging (fMRI) can be used to probe the full repertoire of high-level cognitive functions. While analyzing the resulting image data for a given experiment is a relatively well-mastered process, the challenges in comparing data across multiple datasets poses serious limitation to the field. Indeed, such comparisons require to pool together brain images acquired under different settings and assess the effect of different *experimental conditions* that correspond to psychological effects studied by neuroscientists.

Such meta-analyses are now becoming possible thanks to the development of public data resources –OpenfMRI http://openfmri.org and NeuroVault http://neurovault.org. As many others, researchers of the Parietal team understand these data sources well and contribute to them. However, in such open-ended context, the description of experiments in terms of cognitive concepts is very difficult: there is no universal definition of cognitive terms that could be employed consistently by neuroscientists. Hence meta-analytic studies loose power and specificity. On the other hand, http://brainspell.org provide a set of curated annotation, albeit on much less data, that can serve as a seed or a ground truth to define a consensual ontology of cognitive concepts. Relating these terms to brain activity poses another challenge, of statistical nature, as brain patterns form high-dimensional data in perspective with the scarcity and the noise of the data.

The purpose of this project is to learn a semantic structure in cognitive terms from their occurrence in brain activation. This structure will simplify massive multi-label statistical-learning problems that arise in brain mapping by providing compact representations of cognitive concepts while capturing the imprecision on the definition these concepts.

9.1.4. HidimStat

Participants: Bertrand Thirion [Correspondant], Jerome-Alexis Chevalier, Joseph Salmon.

This project is funded by Digiteo.

This is a Digicosme project (2017-2020) and a collaboration with Joseph Salmon (Telecom Paritech).

The HiDimStat project aims at handling uncertainty in the challenging context of high dimensional regression problem. Though sparse models have been popularized in the last twenty years in contexts where many features can explain a phenomenon, it remains a burning issue to attribute confidence to the predictive models that they produce. Such a question is hard both from the statistical modeling point of view, and from a computation perspective. Indeed, in practical settings, the amount of features at stake (possibly up to several millions in high resolution brain imaging) limit the application of current methods and require new algorithms to achieve computational efficiency. We plan to leverage recent developments in sparse convex solvers as well as more efficient reformulations of testing and confidence interval estimates to provide several communities with practical software handling uncertainty quantification. Specific validation experiments will be performed in the field of brain imaging.

9.1.5. Template estimation for arbitrary alignments: application to brain imaging.

Participants: Bertrand Thirion [Correspondant], Thomas Bazeille.

This project is funded by Digiteo.

In the recent years, the nature of scientific inference has shifted quite substantially from model-based to predictive approaches, thanks to the generalization of powerful machine learning techniques. While this has certainly improved scientific standards, this has also obscured the objects and concepts on which inference is drawn. For instance, it is now possible –based on some initial data– to predict individual brain activity topographies, yet the very notion of a standard brain template has become increasingly elusive. Given the importance of establishing models for the progress of knowledge, we revisit the problem of model inference on data with high variance. Specifically, in a context where almost arbitrary transformation can successfully warp observations to each other with high accuracy, what is the common definition of a population model underlying all these observations? What is the working definition of a template ? We plan to leverage recent developments on optimal transport and multivariate analysis to build working definition of templates; we will use them in a brain imaging context to build a novel generation of brain templates.

9.1.6. AMPHI project

Participants: Bertrand Thirion [Correspondant], Joseph Salmon, Antonio Andre Monteiro Manoel.

This is a Digicosme project (2017-2020) and a collaboration with Joseph Salmon (Telecom Paritech) and Lenka Zdeborova (CEA, IPhT).

In many scientific fields, the data acquisition devices have benefited of hardware improvement to increase the resolution of the observed phenomena, leading to ever larger datasets. While the dimensionality has increased, the number of samples available is often limited, due to physical or financial limits. This is a problem when these data are processed with estimators that have a large sample complexity, such as multivariate statistical models. In that case it is very useful to rely on structured priors, so that the results reflect the state of knowledge on the phenomena of interest. The study of the human brain activity through neuroimaging belongs among these problems, with up to 10^6 features, yet a set of observations limited by cost and participant comfort. We are missing fast estimators for multivariate models with structured priors, that furthermore provide statistical control on the solution. Approximate message passing (AMP) methods are designed to work optimally with low- sample-complexity, they accommodate rather generic class of priors and come with an estimation of

statistical significance. They are therefore well suited for our purposes. We want to join forces to design a new generation of inverse problem solvers that can take into account the complex structure of brain images and provide guarantees in the low-sample-complexity regime. To this end, we will first adapt AMP to the brain mapping setting, using first standard sparsity priors (e.g. Gauss-Bernoulli) on the model. We will then consider more complex structured priors that control the variation of the learned image patterns in space. Crucial gains are expected from the use of the EM algorithm for parameter setting, that comes naturally with AMP. We will also examine the estimators provided by AMP for statistical significance. AMPHI will design a reference inference toolbox released as a generic open source library. We expect a 3- to 10-fold improvement in CPU time, that will benefit to large-scale brain mapping investigations.

9.1.7. CDS2

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Guillaume Lemaitre, Joris Van Den Bossche.

CDS2 is an "Strategic research initiative" of the Paris Saclay University Idex http://datascience-paris-saclay.fr. Although it groups together many partners of the Paris Saclay ecosystem, Parietal has been deeply involved in the project. It currently funds a post-doc for Guillaume Lemaitre and an engineer positions for Joris van den Bossche. Alexandre Boucaud was funded till December as engineer.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. Neuroref: Mathematical Models of Anatomy / Neuroanatomy / Diffusion MRI

Participants: Demian Wassermann [Correspondant], Antonia Machlouzarides Shalit, Valentin Iovene.

While mild traumatic brain injury (mTBI) has become the focus of many neuroimaging studies, the understanding of mTBI, particularly in patients who evince no radiological evidence of injury and yet experience clinical and cognitive symptoms, has remained a complex challenge. Sophisticated imaging tools are needed to delineate the kind of subtle brain injury that is extant in these patients, as existing tools are often ill-suited for the diagnosis of mTBI. For example, conventional magnetic resonance imaging (MRI) studies have focused on seeking a spatially consistent pattern of abnormal signal using statistical analyses that compare average differences between groups, i.e., separating mTBI from healthy controls. While these methods are successful in many diseases, they are not as useful in mTBI, where brain injuries are spatially heterogeneous.

The goal of this proposal is to develop a robust framework to perform subject-specific neuroimaging analyses of Diffusion MRI (dMRI), as this modality has shown excellent sensitivity to brain injuries and can locate subtle brain abnormalities that are not detected using routine clinical neuroradiological readings. New algorithms will be developed to create Individualized Brain Abnormality (IBA) maps that will have a number of clinical and research applications. In this proposal, this technology will be used to analyze a previously acquired dataset from the INTRuST Clinical Consortium, a multi-center effort to study subjects with Post-Traumatic Stress Disorder (PTSD) and mTBI. Neuroimaging abnormality measures will be linked to clinical and neuropsychological assessments. This technique will allow us to tease apart neuroimaging differences between PTSD and mTBI and to establish baseline relationships between neuroimaging markers, and clinical and cognitive measures.

9.2.1.2. DirtyData: Data integration and cleaning for statistical analysis

Participants: Gaël Varoquaux [Correspondant], Patricio Cerda Reyes, Pierre Glaser.

Machine learning has inspired new markets and applications by extracting new insights from complex and noisy data. However, to perform such analyses, the most costly step is often to prepare the data. It entails correcting errors and inconsistencies as well as transforming the data into a single matrix-shaped table that comprises all interesting descriptors for all observations to study. Indeed, the data often results from merging multiple sources of informations with different conventions. Different data tables may come without names on the columns, with missing data, or with input errors such as typos. As a result, the data cannot be automatically shaped into a matrix for statistical analysis.

This proposal aims to drastically reduce the cost of data preparation by integrating it directly into the statistical analysis. Our key insight is that machine learning itself deals well with noise and errors. Hence, we aim to develop the methodology to do statistical analysis directly on the original dirty data. For this, the operations currently done to clean data before the analysis must be adapted to a statistical framework that captures errors and inconsistencies. Our research agenda is inspired from the data-integration state of the art in database research combined with statistical modeling and regularization from machine learning.

Data integrating and cleaning is traditionally performed in databases by finding fuzzy matches or overlaps and applying transformation rules and joins. To incorporate it in the statistical analysis, an thus propagate uncertainties, we want to revisit those logical and set operations with statistical-learning tools. A challenge is to turn the entities present in the data into representations well-suited for statistical learning that are robust to potential errors but do not wash out uncertainty.

Prior art developed in databases is mostly based on first-order logic and sets. Our project strives to capture errors in the input of the entries. Hence we formulate operations in terms of similarities. We address typing entries, deduplication -finding different forms of the same entity- building joins across dirty tables, and correcting errors and missing data.

Our goal is that these steps should be generic enough to digest directly dirty data without user-defined rules. Indeed, they never try to build a fully clean view of the data, which is something very hard, but rather include in the statistical analysis errors and ambiguities in the data.

The methods developed will be empirically evaluated on a variety of dataset, including the French public-data repository, data.gouv.fr. The consortium comprises a company specialized in data integration, Data Publica, that guides business strategies by cross-analyzing public data with market-specific data.

9.2.1.3. FastBig Project

Participants: Bertrand Thirion [Correspondant], Jerome-Alexis Chevalier, Tuan Binh Nguyen.

In many scientific applications, increasingly-large datasets are being acquired to describe more accurately biological or physical phenomena. While the dimensionality of the resulting measures has increased, the number of samples available is often limited, due to physical or financial limits. This results in impressive amounts of complex data observed in small batches of samples.

A question that arises is then : what features in the data are really informative about some outcome of interest ? This amounts to inferring the relationships between these variables and the outcome, conditionally to all other variables. Providing statistical guarantees on these associations is needed in many fields of data science, where competing models require rigorous statistical assessment. Yet reaching such guarantees is very hard.

FAST-BIG aims at developing theoretical results and practical estimation procedures that render statistical inference feasible in such hard cases. We will develop the corresponding software and assess novel inference schemes on two applications : genomics and brain imaging.

9.2.1.4. MultiFracs project

Participant: Philippe Ciuciu [Correspondant].

The scale-free concept formalizes the intuition that, in many systems, the analysis of temporal dynamics cannot be grounded on specific and characteristic time scales. The scale-free paradigm has permitted the relevant analysis of numerous applications, very different in nature, ranging from natural phenomena (hydrodynamic turbulence, geophysics, body rhythms, brain activity,...) to human activities (Internet traffic, population, finance, art,...).

Yet, most successes of scale-free analysis were obtained in contexts where data are univariate, homogeneous along time (a single stationary time series), and well-characterized by simple-shape local singularities. For such situations, scale-free dynamics translate into global or local power laws, which significantly eases practical analyses. Numerous recent real-world applications (macroscopic spontaneous brain dynamics, the central application in this project, being one paradigm example), however, naturally entail large multivariate data (many signals), whose properties vary along time (non-stationarity) and across components (non-homogeneity), with potentially complex temporal dynamics, thus intricate local singular behaviors.

These three issues call into question the intuitive and founding identification of scale-free to power laws, and thus make uneasy multivariate scale-free and multifractal analyses, precluding the use of univariate methodologies. This explains why the concept of scale-free dynamics is barely used and with limited successes in such settings and highlights the overriding need for a systematic methodological study of multivariate scale-free and multifractal dynamics. The Core Theme of MULTIFRACS consists in laying the theoretical foundations of a practical robust statistical signal processing framework for multivariate non homogeneous scale-free and multifractal analyses, suited to varied types of rich singularities, as well as in performing accurate analyses of scale-free dynamics in spontaneous and task-related macroscopic brain activity, to assess their natures, functional roles and relevance, and their relations to behavioral performance in a timing estimation task using multimodal functional imaging techniques.

This overarching objective is organized into 4 Challenges:

1. Multivariate scale-free and multifractal analysis,
2. Second generation of local singularity indices,
3. Scale-free dynamics, non-stationarity and non-homogeneity,
4. Multivariate scale-free temporal dynamics analysis in macroscopic brain activity.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. Neurolang

Title: Accelerating Neuroscience Research by Unifying Knowledge Representation and Analysis Through a Domain Specific Language

Programm: ERC Starting researcher

Duration: 01/03/2018 - 28/02/2023

Coordinator: Demian Wassermann

Inria contact: Demian Wassermann

Summary:

Neuroscience is at an inflection point. The 150-year old cortical specialization paradigm, in which cortical brain areas have a distinct set of functions, is experiencing an unprecedented momentum with over 1000 articles being published every year. However, this paradigm is reaching its limits. Recent studies show that current approaches to atlas brain areas, like relative location, cellular population type, or connectivity, are not enough on their own to characterize a cortical area and its function unequivocally. This hinders the reproducibility and advancement of neuroscience.

Neuroscience is thus in dire need of a universal standard to specify neuroanatomy and function: a novel formal language allowing neuroscientists to simultaneously specify tissue characteristics, relative location, known function and connectional topology for the unequivocal identification of a given brain region.

The vision of NeuroLang is that a unified formal language for neuroanatomy will boost our understanding of the brain. By defining brain regions, networks, and cognitive tasks through a set of formal criteria, researchers will be able to synthesize and integrate data within and across diverse studies. NeuroLang will accelerate the development of neuroscience by providing a way to evaluate anatomical specificity, test current theories, and develop new hypotheses.

NeuroLang will lead to a new generation of computational tools for neuroscience research. In doing so, we will be shedding a novel light onto neurological research and possibly disease treatment and palliative care. Our project complements current developments in large multimodal studies across different databases. This project will bring the power of Domain Specific Languages to neuroscience research, driving the field towards a new paradigm articulating classical neuroanatomy with current statistical and machine learning-based approaches.

9.3.1.2. SLAB (698)

Title: Signal processing and Learning Applied to Brain data

Programm: ERC Starting researcher

Duration: 01/04/2017 - 31/08/2021

Coordinator: Alexandre Gramfort

Partner: LTCI , Telecom ParisTech (France)

Inria contact: Alexandre Gramfort

Summary:

Understanding how the brain works in healthy and pathological conditions is considered as one of the challenges for the 21st century. After the first electroencephalography (EEG) measurements in 1929, the 90's was the birth of modern functional brain imaging with the first functional MRI and full head magnetoencephalography (MEG) system. In the last twenty years, imaging has revolutionized clinical and cognitive neuroscience.

After pioneering works in physics and engineering, the field of neuroscience has to face two major challenges. The size of the datasets keeps growing. The answers to neuroscience questions are limited by the complexity of the signals observed: non-stationarity, high noise levels, heterogeneity of sensors, lack of accurate models. SLAB will provide the next generation of models and algorithms for mining electrophysiology signals which offer unique ways to image the brain at a millisecond time scale.

SLAB will develop dedicated machine learning and signal processing methods and favor the emergence of new challenges for these fields. SLAB focuses on five objectives: 1) source localization with M/EEG for brain imaging at high temporal resolution 2) representation learning to boost statistical power and reduce acquisition costs 3) fusion of heterogeneous sensors 4) modeling of non-stationary spectral interactions to identify functional coupling between neural ensembles 5) development of fast algorithms easy to use by non-experts.

SLAB aims to strengthen mathematical and computational foundations of brain data analysis. The methods developed will have applications across fields (computational biology, astronomy, econometrics). Yet, the primary impact of SLAB will be on neuroscience. The tools and high quality open software produced in SLAB will facilitate the analysis of electrophysiology data, offering new perspectives to understand how the brain works at a mesoscale, and for clinical applications (epilepsy, autism, tremor, sleep disorders).

9.3.1.3. Neuroimaging power (262)

Title: Neuroimaging power

Programm: Marie Curie Fellowhip

Duration: 01/11/2016 - 31/10/2019

Coordinator: Inria

Partner: BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY (United States)

Inria contact: Bertrand Thirion

Summary:

There is an increasing concern about statistical power in neuroscience research. Critically, an underpowered study has poor predictive power. Findings from a low-power study are unlikely to be reproducible, and thus a power analysis is a critical component of any paper. This project aims to promote and facilitate the use of power analyses.

A key component of a power analysis is the specification of an effect size. However, in neuroimaging, there is no standardised way to communicate effect sizes, which makes the choice of an appropriate

effect size a formidable task. The best way today to perform a power analysis is by collecting a pilot data set, a very expensive practice. To eliminate the need for pilot data, we will develop a standardised measure of effect size taking into account the spatial variance and the uncertainty of the measurements. Communicating effect sizes in new publications will facilitate the use of power analyses.

To further alleviate the need for pilot data, we will provide a library of effect sizes for different tasks and contrasts, using open data projects in neuroimaging. We will integrate our effect size estimator in open repositories NeuroVault and OpenfMRI. Consequently, these effect sizes can then serve as a proxy for a pilot study, and as such, a huge cost in the design of an experiment is eliminated.

A new experiment will not be identical to the open data and as such the hypothesised parameters might not be fully accurate. To address this issue, we present a flexible framework to analyse data mid-way without harming the control of the type I error rate. Such a procedure will allow re-evaluating halfway an experiment whether it is useful to continue a study, and how many more subjects are needed for statistically sound inferences. To make our methods maximally available, we will write a software suite including all these methods in different programming platforms and we will provide a GUI to further increase the use of power analyses.

9.3.1.4. HBP SGA1

Title: Human Brain Project Specific Grant Agreement 1

Programm: FET Flagship

Duration: 01/04/2016 - 31/02/2020

Coordinator: Katrin Amunts

Partners: 150 european labs, please see https://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners

Inria contact: Bertrand Thirion

Summary

Understanding the human brain is one of the greatest scientific challenges of our time. Such an understanding can provide profound insights into our humanity, leading to fundamentally new computing technologies, and transforming the diagnosis and treatment of brain disorders. Modern ICT brings this prospect within reach. The HBP Flagship Initiative (HBP) thus proposes a unique strategy that uses ICT to integrate neuroscience data from around the world, to develop a unified multi-level understanding of the brain and diseases, and ultimately to emulate its computational capabilities. The goal is to catalyze a global collaborative effort. During the HBP's first Specific Grant Agreement (SGA1), the HBP Core Project will outline the basis for building and operating a tightly integrated Research Infrastructure, providing HBP researchers and the scientific Community with unique resources and capabilities. Partnering Projects will enable independent research groups to expand the capabilities of the HBP Platforms, in order to use them to address otherwise intractable problems in neuroscience, computing and medicine in the future. In addition, collaborations with other national, European and international initiatives will create synergies, maximizing returns on research investment. SGA1 covers the detailed steps that will be taken to move the HBP closer to achieving its ambitious Flagship Objectives.

9.3.1.5. HBP SGA2

Title: Interactive Computing E-Infrastructure for the Human Brain Project

Programm: FET Flagship

Duration: 01/04/2018 - 31/03/2020

Coordinator: Katrin Amunts

Partners: see https://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners/

Inria contact: Bertrand Thirion

Summary:

The HBP Flagship was launched by the European Commission's Future and Emerging Technologies (FET) scheme in October 2013, and is scheduled to run for ten years. The Flagships, represent a new partnering model for visionary, long-term European cooperative research in the European Research Area, demonstrating the potential for common research efforts. The HBP has the following main objectives:

- Create and operate a European scientific Research Infrastructure for brain research, cognitive neuroscience, and other brain-inspired sciences
- Gather, organise and disseminate data describing the brain and its diseases
- Simulate the brain
- Build multi-scale scaffold theory and models for the brain
- Develop brain-inspired computing, data analytics and robotics
- Ensure that the HBP's work is undertaken responsibly and that it benefits society.

More information on the HBP's Flagship Objectives is available in the Framework Partnership Agreement.

The timeline of the Project is split into multiple phases, each of which will be covered by a separate funding agreement. The current phase is Specific Grant Agreement Two (SGA2), which spans the two-year period from April 2018–April 2020. The HBP is funded via several sources. Total funding is planned to be in the region of EUR 1 billion; around one half of which will be provided by the European Union, and the other by Member States and private funding sources. The European Union contributed EUR 54 million to the Project in the Ramp-Up Phase (October 2013 to March 2016), EUR 89 million for the second phase (SGA1), and EUR 88 million for the current phase (SGA2). The FET Flagships Staff Working Document provides further information on how Flagships are funded.

9.4. International Initiatives

9.4.1. Inria International Labs

Inria@SiliconValley
Associate Team involved in the International Lab:

9.4.1.1. LargeBrainNets

Title: Characterizing Large-scale Brain Networks Using Novel Computational Methods for dMRI and fMRI-based Connectivity

International Partner (Institution - Laboratory - Researcher):

Stanford Cognitive & Systems Neuroscience Lab, Stanford Medical School, USA. Contact: Vinod Menon.

Start year: 2016

See also: http://www-sop.inria.fr/members/Demian.Wassermann/large-brain-nets.html

In the past two decades, brain imaging of neurotypical individuals and clinical populations has primarily focused on localization of function and structures in the brain, revealing activation in specific brain regions during performance of cognitive tasks through modalities such as functional MRI. In parallel, technologies to identify white matter structures have been developed using diffusion MRI. More recently, interest has shifted towards developing a deeper understanding of the brain's intrinsic architecture and its influence on cognitive and affective information processing. Using for this resting state fMRI and diffusion MRI to build the functional and structural networks of the human brain.

The human brain is a complex patchwork of interconnected regions, and graph-theoretical approaches have become increasingly useful for understanding how functionally connected systems engender, and constrain, cognitive functions. The functional nodes of the human brain and their structural inter-connectivity, collectively the "connectome", are, however, poorly understood. Critically, there is a dearth of computational methods for reliably identifying functional nodes of the brain and their structural inter-connectivity in vivo, despite an abundance of high-quality data from the Human Connectome Project (HCP). Devising and validating methods for investigating the human connectome has therefore taken added significance.

The first major goal of this project is to develop and validate appropriate sophisticated computational and mathematical tools for identifying functional nodes at the whole-brain level and measuring structural and functional connectivity between them, using state-of-the-art human brain imaging techniques and open-source HCP data. To this end, we will first develop and validate novel computational tools for (1) identifying stable functional nodes of the human brain using resting-state functional MRI and (2) measuring structural connectivity between functional nodes of the brain using multi-shell high-angular diffusion MRI. Due to the complementarity of the two imaging techniques fMRI and dMRI, our novel computational methods methods, the synergy between the two laboratories of this associate team will allow us to reveal in unprecedented detail the structural and functional connectivity of the human brain.

The second major goal of this project is to use our newly developed computational tools to characterize normal structural and functional brain networks in neurotypical adults.

Inria@SiliconValley
Associate Team involved in the International Lab:

9.4.1.2. Meta&Co

Title: Meta-Analysis of Neuro-Cognitive Associations

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Psychology department. - Russel Poldrack

Start year: 2018

See also: http://team.inria.fr/parietal

Cognitive science and psychiatry describe mental operations: cognition, emotion, perception and their dysfunction. Cognitive neuroimaging bridge these mental concepts to their implementation in the brain, neural firing and wiring, by relying on functional brain imaging. Yet aggregating results from experiments probing brain activity into a consistent description faces the roadblock that cognitive concepts and brain pathologies are ill-defined. Separation between them is often blurry. In addition, these concepts and subdivisions may not correspond to actual brain structures or systems. To tackle this challenge, we propose to adapt data-mining techniques used to learn relationships in computational linguistics. Natural language processing uses distributional semantics to build semantic relationships and ontologies. New models are needed to learn relationships from heterogeneous signals: functional magnetic resonance images (fMRI), on the one hand, combined with related psychology and neuroimaging annotations or publications, on the other hand. Such a joint effort will rely on large publicly-available fMRI databases shared by Podrack Lab, as well as literature mining.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- June 2018: Prof. Lilianne Mujica-Parodi (Univ Stony-Brook, NY USA)
- April-June 2018: Dr Abderrahim Halimi (Edinburgh, UK)
- October 2018: Prof. Nikos Makris (Harvard Medical School)

- December 2018: Dr. Lang Chen (Stanford Medical University)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Bertrand Thirion co-organized the "Imagerie médicale et apprentissage automatique : vers une intelligence artificielle ?" colloque at Collège de France on May, 2nd, 2018.

10.1.1.2. Member of the Organizing Committees

Demian Wassermann: MICCAI 2018

Gaël Varoquaux: PyParis 2018

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Philippe Ciuciu: Chairman of scientific oral sessions in ISBI 2018 (Washington DC) and EUSIPCO 2018 (Roma)
- Bertrand Thirion: Member of Program Committee for the OHBM 2018 meeting (Singapore),
- Gaël Varoquaux: program committee of NIPS, ICML, ICLR.
- Alexandre Gramfort: program committee of NIPS, ICML, ICLR.
- Demian Wassermann: ISMRM 2018

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Philippe Ciuciu: Associate Editor for EUSPICO 2018, Roma
- Philippe Ciuciu: Associate Editor for ISBI 2019, Venice
- Alexandre Gramfort: Editor, NeuroImage, Journal of Machine Learning Research (JMLR), Frontiers in Brain Imaging Methods
- Bertrand Thirion: Editor, Frontiers in Brain Imaging Methods

10.1.3.2. Reviewer - Reviewing Activities

- Philippe Ciuciu is reviewer for Nature Communication, Biological Pysichiatry, Plos Computational Biology, Scientific Reports, Journal of Neuroscience, IEEE Trans Signal Processing, IEEE Signal Processing Letters, IEEE Trans Medical Imaging, Frontiers in Neuroscience, Magnetic Resonance in Medicine, SIAM Imaging Science
- Gaël Varoquaux: Nature Methods, JMLR, PLOS Bio, NeuroImage, IEEE TBME, IEEE TMI, Annals of Applied Statistics, Biological Psychiatry, MedIA, Science, GigaScience
- Alexandre Gramfort: JMLR, PLOS Computational Biology, NeuroImage, IEEE TBME, IEEE TMI, IEEE TSP, MedIA, NIPS, ICML, ICLR, ICASSP, Scientific Reports, Frontiers in Brain Imaging Methods, Journal of Neuroscience Methods
- Bertrand Thirion: Nature communications, Neuroimage, Medical Image Analysis, IEEE TMI, PNAS, PLOS Comp Bio, Brain Structure and Function, NIPS, ICML, IPMI, OHBM, PRNI, AISTATS

- Denis Engemann PLOS Biology, PLOS Computational Biology, Scientific Reports, Neuroimage, Neuroimage Clinical, Human Brain Mapping, Journal of Machine Learning Research, Brain Topography, Brain Connectivity, Journal of Alzheimer's Disease, Neuroscience of Consciousness, PLOS ONE, Frontiers in Neuroscience, Journal of Computational Neuroscience, Psychiatry and Clinical Neurosciences, Sensors
- Demian Wassermann: NeuroImage, MRM, JMRI, Brain Structure and Function, Cortex, MedIA.

10.1.4. Invited Talks

- Philippe Ciuciu, May 2018: Keynote Lecture at the 8th International Conference on New Computational Methods for Inverse Problems, "Distribution-controlled and optimally spread sampling trajectories for accelerated Magnetic Resonance Imaging" (http://complement.farman.ens-cachan.fr/documents_web_NCMIP_2018/NCMIP_2018_program.pdf)
- Philippe Ciuciu, Nov 2018: Heriott-Watt University, "SPARKLING: variable-density k-space filling curves for accelerated T2*-weighted MRI"
- Philippe Ciuciu, Dec 2018: SHFJ, CEA Orsay, "Recent Advances in Compressed Sensing MRI for Highly Accelerated T2* -weighted Imaging"
- Denis Engemann, Dec 2018, ICM, Paris, MEG user day, 'Preparing data for source reconstruction: dos and donts'
- Denis Engemann, Nov 2018, ICM, Paris, 'Machine Learning with MEG EEG in Cognitive Neurology. Challenges and Opportunities'
- Denis Engemann, Nov 2018, Paris Machine Learning Meetup, invited talk, 'Random forest methods for EEG-based diagnosis of disorders of consciousness'
- Denis Engemann, Oct 2018, CRNL, Lyon, invited talk, 'Large-Scale Analysis of MEG/EEG in Cognitive Neurology. Challenges and Opportunities'
- Denis Engemann, Oct 2018, CRNL, Lyon, 3-day MNE-Python training workshop
- Denis Engemann, Apr 2018, Hôpital Erasme, Brussels, invited talk, 'The Challenge of Large-Scale and Population Analysis using MEG/EEG' LCFC
- Alexandre Gramfort, Fév 2018, invited talk, séminaire du Centre de Mathématiques Appliquées (CMAP) de l'Ecole Polytechnique
- Alexandre Gramfort, Mars 2018, invited talk, conseil d'administration Institut National du Cancer (INCA), Paris
- Alexandre Gramfort, Mars 2018, invited talk, Center for Data Science, Grenoble
- Alexandre Gramfort, June 2018, invited talk, ICML workshop of reproducibility in machine learning, Stockholm
- Alexandre Gramfort, June 2018, oral presentations, OHBM Conference, Singapore
- Alexandre Gramfort, June 2018, invited talk, DTU, Copenhagen
- Alexandre Gramfort, Aug 2018, oral presentations, Biomag International Conference, Philadelphia
- Alexandre Gramfort, Sept 2018, invited talk, Imperial College, UK
- Alexandre Gramfort, Sept 2018, invited talk, BCG Gamma FreshFromTheLabs Conference, Paris
- Alexandre Gramfort, Oct 2018, invited talk, Universität Heidelberg, Heidelberg
- Alexandre Gramfort, Nov 2018, oral presentation, Society for Neuroscience conference, San Diego
- Alexandre Gramfort, Nov 2018, invited talk, France is AI conference, Paris
- Alexandre Gramfort, Dec 2018, invited talk, Montreal Artificial Intelligence and Neuroscience (MAIN) workshop, Montreal
- Bertrand Thirion, Jan 2018, IHES, Bures sur Yvette, 'Toward a rigorous statistical framework for brain mapping'

- Bertrand Thirion, June 2018, IPHT seminary, Gif sur Yvette, 'A rigorous causal framework for brain mapping'
- Bertrand Thirion, May 2018, High Tech Peripherique conference, 'Large-scale machine learning for medical imaging'
- Bertrand Thirion, April 2018, Tau seminary, Gif sur Yvette, 'Causal analysis for Brain Mapping'
- Bertrand Thirion, July 2018, JST workshop in Paris, 'Toward rigorous e-sciences: High-dimensional statistical inference'
- Bertrand Thirion, Dec 2018, ESSI colloquium, Evry, 'High-dimensional statistical inference for e-sciences'
- Demian Wassermann, Feb 2018, Stanford Medical School, 'Microstructure Imaging with Diffusion MRI'
- Demian Wassermann, June 2018, Harvard Medical School, 'Random Effect Models for Structure Connectivity-Based Cortical Clustering'
- Demian Wassermann, June 2018, New York University, 'Recent advances in Micro and Macro scale brain analysis with Diffusion MRI'
- Gaël Varoquaux, Sept 2018, IPAM (Institute for Pure and Applied Mathematics), UCLA, Long Program Science at Extreme Scales: Where Big Data Meets Large-Scale Computing
- Gaël Varoquaux, August 2018, keynote speaker NeuroInformatics, Montreal
- Gaël Varoquaux, August 2018, invited talk MILA (Montreal Institute for Learning Algorithms), Montreal
- Gaël Varoquaux, August 2018, invited talk MNI (Montréal Neurological Institute), Montréal
- Gaël Varoquaux, August 2018, invited talk Institut de Gériatrie, Montréal
- Gaël Varoquaux, June 2018, invited talk, BCG Gamma days, Paris
- Gaël Varoquaux, Nov 2018, invited talk, chair DAMI (Data, Analytics and Models for Insurance), Paris
- Gaël Varoquaux, Feb 2018, invited talk, Gatsby institute for theoretical neuroscience, UCL, London
- Gaël Varoquaux, Dev 2018, invited talk, Journée Nationale de la Science Ouverte, Paris
- Gaël Varoquaux, Sept 2018, invited talk, MASES International Workshop on Machine Learning and Software Engineering in Symbiosis, ASE (Automated Software Engineering), Montpellier
- Gaël Varoquaux, Sept 2018, invited talk, MICCAI workshop, Granada
- Gaël Varoquaux, June 2018, invited talk, Machine learning in the real world, Paris
- Gaël Varoquaux, July 2018, invited talk, DataIA–Japan Science and Technology agency, Paris
- Gaël Varoquaux, Feb 2018, invited talk, Pycon Belarus, Minsk
- Gaël Varoquaux, May 2018, invited talk, Brainhack Paris
- Gaël Varoquaux, Oct 2018, invited talk, Biomarker days, Toulouse

10.1.5. Leadership within the Scientific Community

- Demian Wassermann: Organising committee of the BrainHack community — Paris Chapter
- Demian Wassermann: ISMRM and MICCAI action organization towards reducing gender and minority biases.

10.1.6. Scientific Expertise

- Philippe Ciuciu has been member of the 45th ANR Scientic Evaluation Committee in charge of evaluating the projects dealing with maths, signal processing, computer science methods for medicine and biology.

- Alexandre Gramfort has been member of the 45th ANR Scientic Evaluation Committee (CE23) in charge of evaluating the projects dealing with machine learning, data knowledge, statistics, optimization.
- Demian Wassermann FET-OPEN ERCA Action
- Gaël Varoquaux was expert for the startup incubator agoranov

10.1.7. Research Administration

- In Nov 2018, Philippe Ciuciu Elected Vice-Chair of the SAT Biomedical Image & Signal Analytics (EURASIP technical committee) for 2019-2020.
- In Dec 2018, Philippe Ciuciu has been Elected IEEE Signal Processing Society Representative at the 2019 IEEE ISBI conference.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Alexandre Gramfort, Optimization for Data Science, 20h, Msc 2 Data Science Master Ecole Polytechnique, France

Master: Alexandre Gramfort, DataCamp, 20h, Msc 2 Data Science Master Ecole Polytechnique, France

Master: Alexandre Gramfort, Source Imaging with EEG and MEG, 7h, Msc 2 in Biomedical Imaging at Télécom Paristech

Doctoral School: Alexandre Gramfort, Sparse Optimization, 4h, school at iTWIST workshop, Marseille

Master: Bertrand Thirion, Functional neuroimaging and BCI, 12h, Master MVA, ENS Paris-Saclay, France

Master: Philippe, fMRI course: From acquisition to data Analysis, 6h, Msc 2 in Biomedical Imaging, Université Paris-Sud

Bachelor: Demian Wassermann, CSE201 class, 15h, C++ programming, Ecole Polytechnique

Extension: Demian Wassermann, Data Science, 20h, Ecole Polytechnique

Master: Gaël Varoquaux, Machine learning in Python, 3h, ENSAE

Master: Gaël Varoquaux, Functional brain connectivity, 7h, Msc 2 in Biomedical Imaging at Télécom Paristech

10.2.2. Supervision

The following PhD defense have taken place

HdR : Gaël Varoquaux, Estimating brain functional connectivity and its variations from fMRI, Université Paris VI, 25/03/2018

PhD : Stanislas Chambon, *Learning from electrophysiology time series: From scoring to event detection*, Université Paris-Saclay, 14/12/2018, under the direction of Alexandre Gramfort

PhD : Tom Dupré La Tour, *Non-linear models for neurophysiological time series*, Université Paris-Saclay, 26/11/2018, under the direction of Alexandre Gramfort

PhD : Yousra Bekhti, *Contributions to sparse source localization for MEG/EEG brain imaging*, Université Paris-Saclay, 22/03/2018, under the direction of Alexandre Gramfort

PhD : Mainak Jas, *Advances in automating analysis of neural time series data*, Université Paris-Saclay, 12/04/2018, under the direction of Alexandre Gramfort

PhD : Arthur Mensch, , Université Paris-Saclay, 30/09/2018, under the direction of Bertrand Thirion, Gaël Varoquaux and Julien Mairal.

PhD : Carole Lazarus, , Université Paris-Saclay, 30/09/2018, under the direction of Philippe Ciuciu.

PhD in progress : Pierre Ablin, , 01/10/2016, coadvised by Alexandre Gramfort and Jean-François Cardoso,

PhD in progress : Mathurin Massias, , 01/10/2016, coadvised by Alexandre Gramfort and Joseph Salmon,

PhD in progress : Hicham Janati, , 01/10/2017, , coadvised by Alexandre Gramfort and Marco Cuturi

PhD in progress : Quentin Bertrand, , 01/10/2018, coadvised by Alexandre Gramfort and Joseph Salmon

PhD in progress : Hubert Banville, , 01/10/2018, coadvised by Alexandre Gramfort and Denis Engemann

PhD in progress : David Sabbagh, , 01/10/2018, coadvised by Alexandre Gramfort and Denis Engemann

PhD in progress : Jérôme-Alexis Chevalier, , 01/10/2017, coadvised by Bertrand Thirion and Joseph Salmon

PhD in progress : Hugo Richard, , 01/10/2018, coadvised by Bertrand Thirion and Olivier Grisel

PhD in progress : Thomas Bazeille, , 01/10/2018, advised by Bertrand Thirion

PhD in progress : Tuan Binh Nguyen, , 01/10/2018, coadvised by Bertrand thirion and Sylvain Arlot

PhD in progress : Valentin Iovene, , 01/10/2018, advised by Demian Wassermann

PhD in progress : Antonia Machlouzarides Shalit, , 01/10/2018, coadvised by Demian Wassermann and Bertrand Thirion

PhD in progress : Loubna El Gueddari, , 01/10/2016, coadvised by Philippe Ciuciu and Alexandre Vignaud

PhD in progress : Hamza Cherkaoui, , 01/10/2017, advised by Philippe Ciuciu

PhD in progress : Patricio Cerda Reyes, , 01/10/2016, coadvised by Gaël Varoquaux and Balazs Kegl

PhD in progress : Maeliss Jallais, , 01/10/2018, advised by Demian Wassermann

PhD in progress : Jerome Dockès, , 01/10/2016, coadvised by Fabian Shuchanek and Gaël Varoquaux

10.2.3. Juries

Alexandre Gramfort has been involved in the following PhD committees:

- Marine Le Morvan (Reviewer) / Mines ParisTech
- Jérémy Guillon (Examiner) / Inria Paris
- Rémi Leblond (Examiner) / Inria Paris
- Andreas Trier (Reviewer) / DTU, Copenhague

Alexandre Gramfort was involved in a hiring committee for Telecom ParisTech and a hiring committee for Université Paris Diderot.

Bertrand Thirion has been involved in the following committees:

- PhD Sebastian Tarando / ESSI Evry
- PhD Guillermo Gallardo / Inria Sophia-Antipolis
- Habilitation Florent Meyniel / Université Paris VI

Bertrand Thirion was involved in a hiring committee for Centrale-Supelec.

Gaël Varoquaux was involved in a hiring committee for Centrale-Supelec and a hiring committee for Inria Saclay.

Demian Wassermann was involved in the following committees:

- PhD Guillermo Gallardo / Inria Sophia-Antipolis
- Habilitation Daniel Margullies / ICM – UPMC

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- Philippe Ciuciu has been member of the Inria Saclay scientific commission since 2016
- Alexandre Gramfort is member of the steering committee of the Paris-Saclay Center for Data Science
- Alexandre is a Member of the technical development committee of Inria Saclay since Dec 2018
- Bertrand Thirion is Deputy head of reasearch of Inria Saclay research center
- Bertrand Thirion, Leader of the Datasense research axis of the Digicosme Labex
- Bertrand Thirion, Member of the steering committee of the Dataia Convergence Institute
- Bertrand Thirion, Member of the steering committee of the Computer Science Department of Paris Saclay University.
- Gaël Varoquaux was Member of the technical development committee of Inria Saclay until Dec 2018
- Gaël Varoquaux is member of the doctoral monitoring committee of Inria Saclay
- Gaël Varoquaux is member of the steering committee of the Paris-Saclay Center for Data Science

10.3.2. Articles and contents

- Gaël Varoquaux was interviewed by ActuIA
- Alexandre Gramfort was interviewed by ActuIA
- Alexandre Gramfort was interviewed by Libération
- Alexandre Gramfort, Olivier Grisel Gaël Varoquaux are interviewed by Les Echos

10.3.3. Interventions

- Olivier Grisel gave an interview at the Paris Open Source Summit

10.3.4. Internal action

- In June 2018, Philippe Ciuciu did a training session on the DRF Impulsion funding mechanism at the seminar of new hired CEA DRF scientists"
- In April 2018, Denis Engemann and Alexandre Gramfort gave a 2-day educational course for cognitive neuroscientists at NeuroSpin on analysis of MEG and EEG data using Python.

10.3.5. Creation of media or tools for science outreach

Philippe Ciuciu made Two videos together with the CEA communication division on COSMIC project (funded by CEA)

- Long version: https://www.youtube.com/watch?v=gQh6D_vpkSo&t=23s&fbclid=IwAR0OcU3JEy4KQmo6DD-iN8otjlowyeXDkEF7ljniuzOkI-aWhGqyHhQzruo
- Short version: https://www.youtube.com/watch?v=p_KMEQGK-WA&fbclid=IwAR3xcf8e98M77lF9mpQSDWfKppjOt7p

Gaël Varoquaux, Olivier Grisel, Alexandre Gramfort, Guillaume Lemaître, Joris van den Bossche participated in a general-public movie about scikit-learn: https://www.youtube.com/watch?v=twqdXTCkeyk&t=9s

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] A. MENSCH. *Learning representations from functional MRI data*, Université Paris-Saclay, September 2018, https://tel.archives-ouvertes.fr/tel-01891633

Articles in International Peer-Reviewed Journals

[2] P. A. ABLIN, J.-F. CARDOSO, A. GRAMFORT. *Faster Independent Component Analysis by Preconditioning With Hessian Approximations*, in "IEEE Transactions on Signal Processing", August 2018, vol. 66, n^{o} 15, pp. 4040-4049, https://hal.inria.fr/hal-01970746

[3] Y. BEKHTI, F. LUCKA, J. SALMON, A. GRAMFORT. *A hierarchical Bayesian perspective on majorization-minimization for non-convex sparse regression: application to M/EEG source imaging*, in "Inverse Problems", August 2018, vol. 34, n^{o} 8, 085010 p. , https://hal.inria.fr/hal-01970744

[4] D. BZDOK, N. ALTMAN, M. KRZYWINSKI. *Points of Significance: Statistics versus Machine Learning*, in "Nature Methods", April 2018, pp. 1-7, https://hal.archives-ouvertes.fr/hal-01723223

[5] D. BZDOK, M. KRZYWINSKI, N. ALTMAN. *Machine learning: Supervised methods, SVM and kNN*, in "Nature Methods", January 2018, pp. 1-6, https://hal.archives-ouvertes.fr/hal-01657491

[6] D. BZDOK, A. MEYER-LINDENBERG. *Machine learning for precision psychiatry: Opportunites and challenges*, in "Biological Psychiatry: Cognitive Neuroscience and Neuroimaging", February 2018, https://hal.archives-ouvertes.fr/hal-01643933

[7] P. CERDA, G. VAROQUAUX, B. KÉGL. *Similarity encoding for learning with dirty categorical variables*, in "Machine Learning", June 2018, https://arxiv.org/abs/1806.00979 [*DOI :* 10.1007/s10994-018-5724-2], https://hal.inria.fr/hal-01806175

[8] S. CHAMBON, M. GALTIER, P. J. ARNAL, G. WAINRIB, A. GRAMFORT. *A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series*, in "IEEE Transactions on Neural Systems and Rehabilitation Engineering", March 2018, vol. 26, n^{o} 4, 17683810 p. , https://arxiv.org/abs/1707.03321 [*DOI :* 10.1109/TNSRE.2018.2813138], https://hal.archives-ouvertes.fr/hal-01810436

[9] C. CURY, J. GLAUNÈS, R. TORO, M. CHUPIN, G. SCHUMANN, V. FROUIN, J.-B. POLINE, O. COLLIOT. *Statistical Shape Analysis of Large Datasets Based on Diffeomorphic Iterative Centroids*, in "Frontiers in Neuroscience", November 2018, vol. 12 [*DOI :* 10.3389/FNINS.2018.00803], https://hal.inria.fr/hal-01920263

[10] E. DOHMATOB, G. VAROQUAUX, B. THIRION. *Inter-subject registration of functional images: do we need anatomical images ?*, in "Frontiers in Neuroscience", March 2018, https://hal.archives-ouvertes.fr/hal-01701619

[11] D. ENGEMANN, F. RAIMONDO, J.-R. KING, B. ROHAUT, G. LOUPPE, F. FAUGERAS, J. ANNEN, H. CASSOL, O. GOSSERIES, D. FERNANDEZ-SLEZAK, S. LAUREYS, L. NACCACHE, S. DEHAENE, J. SITT.

Robust EEG-based cross-site and cross-protocol classification of states of consciousness, in "Brain - A Journal of Neurology ", October 2018, vol. 141, n^{o} 11, pp. 3179–3192 [*DOI :* 10.1093/BRAIN/AWY251], https://hal.inria.fr/hal-01887793

[12] P. FILIPIAK, R. H. FICK, A. PETIET, M. SANTIN, A.-C. PHILIPPE, S. LEHÉRICY, P. CIUCIU, R. DERICHE, D. WASSERMANN. *Reducing the number of samples in spatiotemporal dMRI acquisition design*, in "Magnetic Resonance in Medicine", November 2018 [*DOI :* 10.1002/MRM.27601], https://hal.archives-ouvertes.fr/hal-01928734

[13] F. HADJ-SELEM, T. LOFSTEDT, E. DOHMATOB, V. FROUIN, M. DUBOIS, V. GUILLEMOT, E. DUCHESNAY. *Continuation of Nesterov's Smoothing for Regression with Structured Sparsity in High-Dimensional Neuroimaging*, in "IEEE Transactions on Medical Imaging", 2018, vol. 2018 [*DOI :* 10.1109/TMI.2018.2829802], https://hal-cea.archives-ouvertes.fr/cea-01883286

[14] G. HARTWIGSEN, D. BZDOK. *Multivariate single-subject analysis of short-term reorganization in the language network*, in "Cortex", July 2018, 4 p. [*DOI :* 10.1016/J.CORTEX.2018.06.013], https://hal.archives-ouvertes.fr/hal-01824229

[15] Y. HONG, L. J. O'DONNELL, P. SAVADJIEV, F. ZHANG, D. WASSERMANN, O. PASTERNAK, H. J. JOHNSON, J. PAULSEN, J.-P. VONSATTEL, N. MAKRIS, C.-F. WESTIN, Y. RATHI. *Genetic load determines atrophy in hand cortico-striatal pathways in presymptomatic Huntington's disease*, in "Human Brain Mapping", 2018 [*DOI :* 10.1002/HBM.24217], https://hal.inria.fr/hal-01787886

[16] M. JAS, E. . LARSON, D. ENGEMANN, J. LEPPÄKANGAS, S. TAULU, M. HÄMÄLÄINEN, A. GRAMFORT. *A Reproducible MEG/EEG Group Study With the MNE Software: Recommendations, Quality Assessments, and Good Practices*, in "Frontiers in Neuroscience", August 2018, vol. 12 [*DOI :* 10.3389/FNINS.2018.00530], https://hal.archives-ouvertes.fr/hal-01854552

[17] J. M. KERNBACH, T. D. SATTERTHWAITE, D. S. BASSETT, J. SMALLWOOD, D. MARGULIES, S. KRALL, P. SHAW, G. VAROQUAUX, B. THIRION, K. KONRAD, D. BZDOK. *Shared Endo-phenotypes of Default Mode Dysfunction in Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder*, in "Translational Psychiatry", July 2018, https://hal.archives-ouvertes.fr/hal-01790245

[18] J. M. KERNBACH, B. T. T. YEO, J. SMALLWOOD, D. MARGULIES, M. THIEBAUT DE SCHOTTEN, H. WALTER, M. SABUNCU, A. J. HOLMES, A. GRAMFORT, G. VAROQUAUX, B. THIRION, D. BZDOK. *Subspecialization within default mode nodes characterized in 10,000 UK Biobank participants*, in "Proceedings of the National Academy of Sciences of the United States of America ", November 2018 [*DOI :* 10.1073/PNAS.1804876115], https://hal.archives-ouvertes.fr/hal-01926796

[19] M. KOWALSKI, A. MEYNARD, H.-T. WU. *Convex Optimization approach to signals with fast varying instantaneous frequency*, in "Applied and Computational Harmonic Analysis", January 2018, vol. 44, n^{o} 1, pp. 89 - 122, https://arxiv.org/abs/1503.07591 [*DOI :* 10.1016/J.ACHA.2016.03.008], https://hal.archives-ouvertes.fr/hal-01199615

[20] C. LAZARUS, P. WEISS, A. VIGNAUD, P. CIUCIU. *An Empirical Study of the Maximum Degree of Undersampling in Compressed Sensing for T2*-weighted MRI*, in "Magnetic Resonance Imaging", 2018, pp. 1-31, https://hal.inria.fr/hal-01829323

[21] J. LEFORT-BESNARD, D. S. BASSETT, J. SMALLWOOD, D. S. MARGULIES, B. DERNTL, O. GRUBER, A. ALEMAN, R. JARDRI, G. VAROQUAUX, B. THIRION, S. B. EICKHOFF, D. BZDOK. *Different shades of default mode disturbance in schizophrenia: Subnodal covariance estimation in structure and function*, in "Human Brain Mapping", January 2018, pp. 1-52, https://hal.archives-ouvertes.fr/hal-01620441

[22] J. LEFORT-BESNARD, G. VAROQUAUX, B. DERNTL, O. GRUBER, A. ALEMAN, R. JARDRI, I. SOMMER, B. THIRION, D. BZDOK. *Patterns of Schizophrenia Symptoms: Hidden Structure in the PANSS Questionnaire*, in "Translational Psychiatry", 2018, https://hal.archives-ouvertes.fr/hal-01888918

[23] L. M. M. LÊ, B. KÉGL, A. GRAMFORT, C. MARINI, D. NGUYEN, M. CHERTI, S. TFAILI, A. TFAYLI, A. BAILLET-GUFFROY, P. PROGNON, P. CHAMINADE, E. CAUDRON. *Optimization of classification and regression analysis of four monoclonal antibodies from Raman spectra using collaborative machine learning approach*, in "Talanta", July 2018, vol. 184, pp. 260-265, https://hal.archives-ouvertes.fr/hal-01969999

[24] A. MENSCH, J. MAIRAL, B. THIRION, G. VAROQUAUX. *Stochastic Subsampling for Factorizing Huge Matrices*, in "IEEE Transactions on Signal Processing", January 2018, vol. 66, n^{o} 1, pp. 113-128, https://arxiv.org/abs/1701.05363 [*DOI :* 10.1109/TSP.2017.2752697], https://hal.archives-ouvertes.fr/hal-01431618

[25] A. L. PINHO, A. AMADON, T. RUEST, M. FABRE, E. DOHMATOB, I. DENGHIEN, C. GINISTY, S. BECUWE-DESMIDT, S. ROGER, L. LAURIER, V. JOLY-TESTAULT, G. MÉDIOUNI-CLOAREC, C. DOUBLÉ, B. MARTINS, P. PINEL, E. EGER, G. VAROQUAUX, C. PALLIER, S. DEHAENE, L. HERTZ-PANNIER, B. THIRION. *Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping*, in "Scientific Data ", June 2018, vol. 5, 180105 p. [*DOI :* 10.1038/SDATA.2018.105], https://hal.archives-ouvertes.fr/hal-01817528

[26] V. SYDNOR, A. M. RIVAS-GRAJALES, A. LYALL, F. ZHANG, S. BOUIX, S. KARMACHARYA, M. SHENTON, C.-F. WESTIN, N. MAKRIS, D. WASSERMANN, L. J. O'DONNELL, M. KUBICKI. *A comparison of three fiber tract delineation methods and their impact on white matter analysis*, in "NeuroImage", May 2018, vol. 178, pp. 318-331 [*DOI :* 10.1016/J.NEUROIMAGE.2018.05.044], https://hal.inria.fr/hal-01807178

[27] G. VAROQUAUX, R. POLDRACK. *Predictive models avoid excessive reductionism in cognitive neuroimaging*, in "Current Opinion in Neurobiology", April 2019, vol. 55 [*DOI :* 10.1016/J.CONB.2018.11.002], https://hal.archives-ouvertes.fr/hal-01856412

[28] G. VAROQUAUX, Y. SCHWARTZ, R. POLDRACK, B. GAUTHIER, D. BZDOK, J.-B. POLINE, B. THIRION. *Atlases of cognition with large-scale brain mapping*, in "PLoS Computational Biology", 2018, https://hal.inria.fr/hal-01908189

[29] L. WALLER, A. BROVKIN, L. DORFSCHMIDT, D. BZDOK, H. WALTER, J. D. KRUSCHWITZ. *GraphVar 2.0: a user-friendly toolbox for machine learning on functional connectivity measures*, in "Journal of Neuroscience Methods", January 2018, 40 p. , https://hal.archives-ouvertes.fr/hal-01828991

[30] H.-T. WANG, D. BZDOK, D. MARGULIES, C. CRADDOCK, M. MILHAM, E. JEFFERIES, J. SMALLWOOD. *Patterns of thought: population variation in the associations between large-scale network organisation and self-reported experiences at rest*, in "NeuroImage", May 2018, https://hal.archives-ouvertes.fr/hal-01782292

[31] A. DE PIERREFEU, T. FOVET, F. HADJ-SELEM, T. LOFSTEDT, P. CIUCIU, S. LEFEBVRE, P. THOMAS, R. LOPES, R. JARDRI, E. DUCHESNAY. *Prediction of activation patterns preceding hallucinations in patients with schizophrenia using machine learning with structured sparsity*, in "Human Brain Mapping", April

2018, vol. 39, n^{o} 4, pp. 1777 - 1788 [*DOI :* 10.1002/HBM.23953], https://hal-cea.archives-ouvertes.fr/cea-01883271

[32] A. DE PIERREFEU, T. LOFSTEDT, F. HADJ-SELEM, M. DUBOIS, R. JARDRI, T. FOVET, P. CIUCIU, V. FROUIN, E. DUCHESNAY. *Structured Sparse Principal Components Analysis With the TV-Elastic Net Penalty*, in "IEEE Transactions on Medical Imaging", February 2018, vol. 37, n^{o} 2, pp. 396 - 407 [*DOI :* 10.1109/TMI.2017.2749140], https://hal-cea.archives-ouvertes.fr/cea-01883278

[33] A. DE PIERREFEU, T. LÖFSTEDT, C. LAIDI, F. HADJ-SELEM, J. BOURGIN, T. HAJEK, F. SPANIEL, M. KOLENIC, P. CIUCIU, N. HAMDANI, M. LEBOYER, T. FOVET, R. JARDRI, J. HOUENOU, E. DUCHESNAY. *Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity*, in "Acta Psychiatrica Scandinavica", 2018, vol. 2018, pp. 1 - 10 [*DOI :* 10.1111/ACPS.12964], https://hal-cea.archives-ouvertes.fr/cea-01883283

International Conferences with Proceedings

[34] P. A. ABLIN, J.-F. CARDOSO, A. GRAMFORT. *Accelerating likelihood optimization for ICA on real signals*, in "LVA-ICA 2018", Guildford, United Kingdom, July 2018, https://arxiv.org/abs/1806.09390 , https://hal.inria.fr/hal-01822602

[35] A. ALIMI, R. H. FICK, D. WASSERMANN, R. DERICHE. *Dmipy, a Diffusion Microstructure Imaging toolbox in Python to improve research reproducibility*, in "MICCAI 2018 - Workshop on Computational Diffusion MRI", Granada, Spain, September 2018, https://hal.inria.fr/hal-01873353

[36] H. CHERKAOUI, L. E. GUEDDARI, C. LAZARUS, A. GRIGIS, F. POUPON, A. VIGNAUD, S. FARRENS, J.-L. STARCK, P. CIUCIU. *Analysis vs Synthesis-based Regularization for combined Compressed Sensing and Parallel MRI Reconstruction at 7 Tesla*, in "26th European Signal Processing Conference (EUSIPCO 2018)", Roma, Italy, September 2018, https://hal.inria.fr/hal-01800700

[37] J. DOCKÈS, D. WASSERMANN, R. POLDRACK, F. M. SUCHANEK, B. THIRION, G. VAROQUAUX. *Text to brain: predicting the spatial distribution of neuroimaging observations from text reports*, in "MICCAI 2018 - 21st International Conference on Medical Image Computing and Computer Assisted Intervention", Granada, Spain, September 2018, pp. 1-18, https://arxiv.org/abs/1806.01139 , https://hal.archives-ouvertes.fr/hal-01807295

[38] L. EL GUEDDARI, C. LAZARUS, H. CARRIÉ, A. VIGNAUD, P. CIUCIU. *Self-calibrating nonlinear reconstruction algorithms for variable density sampling and parallel reception MRI*, in "10th IEEE Sensor Array and Multichannel Signal Processing workshop", Sheffield, United Kingdom, July 2018, pp. 1-5, https://hal.inria.fr/hal-01782428

[39] P. FILIPIAK, R. FICK, A. PETIET, M. SANTIN, A.-C. PHILIPPE, S. LEHÉRICY, R. DERICHE, D. WASSERMANN. *Coarse-Grained Spatiotemporal Acquisition Design for Diffusion MRI*, in "ISBI 2019 - Proceedings of The IEEE International Symposium on Biomedical Imaging", Venice, Italy, April 2019, https://hal.inria.fr/hal-01973588

[40] G. GALLARDO, N. GAYRAUD, R. DERICHE, M. CLERC, S. DESLAURIERS-GAUTHIER, D. WASSERMANN. *Solving the Cross-Subject Parcel Matching Problem using Optimal Transport*, in "International Conference on Medical Image Computing and Computer-Assisted Intervention 2018", Granada, Spain, September 2018, https://hal.archives-ouvertes.fr/hal-01935684

[41] D. La Rocca, P. Ciuciu, V. van Wassenhove, H. Wendt, P. Abry, R. Leonarduzzi. *Scale-free functional connectivity analysis from source reconstructed MEG data*, in "EUSIPCO 2018 - 26th European Signal Processing Conference", Roma, Italy, September 2018, pp. 1-5, https://hal.inria.fr/hal-01800620

[42] C. Maumet, G. Flandin, M. Perez-Guevara, J.-B. Poline, J. Rajendra, R. Reynolds, B. Thirion, T. E. Nichols. *A standardised representation for non-parametric fMRI results*, in "OHBM 2018 - Annual meeting of the Organization of Human Brain Mapping", Singapore, Singapore, June 2018, pp. 1-4, http://www.hal.inserm.fr/inserm-01828914

[43] A. Mensch, M. Blondel. *Differentiable Dynamic Programming for Structured Prediction and Attention*, in "35th International Conference on Machine Learning", Stockholm, Sweden, Proceedings of the 35th International Conference on Machine Learning, July 2018, vol. 80, https://hal.archives-ouvertes.fr/hal-01809550

[44] H. Wendt, P. Abry, P. Ciuciu. *Spatially regularized wavelet leader scale-free analysis of fMRI data*, in "IEEE International Symposium on Biomedical Imaging", Washington, DC, United States, April 2018, https://hal.inria.fr/hal-01782332

Conferences without Proceedings

[45] S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, A. Gramfort. *A deep learning architecture to detect events in EEG signals during sleep*, in "IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2018)", Aalborg, Denmark, September 2018, https://hal.archives-ouvertes.fr/hal-01917529

[46] D. Chyzhyk, G. Varoquaux, B. Thirion, M. Milham. *Controlling a confound in predictive models with a test set minimizing its effect*, in "PRNI 2018 - 8th International Workshop on Pattern Recognition in Neuroimaging", Singapore, Singapore, June 2018, pp. 1-4, https://hal.archives-ouvertes.fr/hal-01831701

[47] T. Dupré la Tour, Y. Grenier, A. Gramfort. *Driver estimation in non-linear autoregressive models*, in "43nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2018)", Calgary, Canada, April 2018, https://hal.archives-ouvertes.fr/hal-01696786

[48] M. Frigo, G. Gallardo, I. Costantini, A. Daducci, D. Wassermann, R. Deriche, S. Deslauriers-Gauthier. *Reducing false positive connection in tractograms using joint structure-function filtering*, in "OHBM 2018 - Organization for Human Brain Mapping", Singapore, Singapore, June 2018, pp. 1-3, https://hal.inria.fr/hal-01737434

[49] G. Gallardo, S. Bouix, D. Wassermann. *Diffusion Driven Label Fusion for White Matter Multi-Atlas Segmentation*, in "OHBM 2018 - Organization for Human Brain Mapping", Singapore, Singapore, June 2018, pp. 1-2, https://hal.archives-ouvertes.fr/hal-01737422

[50] N. T. H. Gayraud, G. Gallardo, M. Clerc, D. Wassermann. *Solving the Cross-Subject Parcel Matching Problem: Comparing Four Methods Using Extrinsic Connectivity*, in "OHBM 2018", Singapore, Singapore, June 2018, https://hal.archives-ouvertes.fr/hal-01737366

[51] T. Kerdreux, F. Pedregosa, A. d'Aspremont. *Frank-Wolfe with Subsampling Oracle*, in "ICML 2018 - 35th International Conference on Machine Learning", Stockholm, Sweden, July 2018, https://arxiv.org/abs/1803.07348 , https://hal.archives-ouvertes.fr/hal-01927391

[52] T. D. LA TOUR, T. MOREAU, M. JAS, A. GRAMFORT. *Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals*, in "Advances in Neural Information Processing Systems (NeurIPS)", Montréal, Canada, December 2018, https://arxiv.org/abs/1805.09654 , https://hal.archives-ouvertes.fr/hal-01966685

[53] M. MASSIAS, O. FERCOQ, A. GRAMFORT, J. SALMON. *Generalized Concomitant Multi-Task Lasso for Sparse Multimodal Regression*, in "21st International Conference on Artificial Intelligence and Statistics (AISTATS 2018)", Lanzarote, Spain, April 2018, https://hal.archives-ouvertes.fr/hal-01812011

[54] M. MASSIAS, A. GRAMFORT, J. SALMON. *Celer: a Fast Solver for the Lasso with Dual Extrapolation*, in "ICML 2018 - 35th International Conference on Machine Learning", Stockholm, Sweden, PMLR, July 2018, vol. 80, pp. 3321-3330, https://hal.archives-ouvertes.fr/hal-01833398

[55] H. RICHARD, A. PINHO, B. THIRION, G. CHARPIAT. *Optimizing deep video representation to match brain activity*, in "CCN 2018 - Conference on Cognitive Computational Neuroscience", Philadelphia, United States, September 2018, https://arxiv.org/abs/1809.02440 , https://hal.archives-ouvertes.fr/hal-01868735

[56] J.-B. SCHIRATTI, J.-E. LE DOUGET, M. LE VAN QUYEN, S. ESSID, A. GRAMFORT. *An ensemble learning approach to detect epileptic seizures from long intracranial EEG recordings*, in "International Conference on Acoustics, Speech, and Signal Processing", Calgary, Canada, April 2018, https://hal.archives-ouvertes.fr/hal-01724272

[57] A. DE PIERREFEU, T. LOFSTEDT, C. LAIDI, F. HADJ-SELEM, M. LEBOYER, P. CIUCIU, J. HOUENOU, E. DUCHESNAY. *Interpretable and stable prediction of schizophrenia on a large multisite dataset using machine learning with structured sparsity*, in "2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)", Singapore, Singapore, IEEE, June 2018 [*DOI :* 10.1109/PRNI.2018.8423946], https://hal-cea.archives-ouvertes.fr/cea-01883311

Other Publications

[58] P. A. ABLIN, J.-F. CARDOSO, A. GRAMFORT. *Beyond Pham's algorithm for joint diagonalization*, November 2018, https://arxiv.org/abs/1811.11433 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01936887

[59] P. A. ABLIN, D. FAGOT, H. WENDT, A. GRAMFORT, C. FÉVOTTE. *A Quasi-Newton algorithm on the orthogonal manifold for NMF with transform learning*, November 2018, https://arxiv.org/abs/1811.02225 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01912918

[60] F. ALMAIRAC, P. FILIPIAK, L. SLABU, M. CLERC, T. PAPADOPOULO, D. FONTAINE, L. MONDOT, S. CHANELET, D. WASSERMANN, R. DERICHE. *Bridging Brain Structure and Function by Correlating Structural Connectivity and Cortico-Cortical Transmission*, June 2018, 2nd C@UCA meeting, Poster, https://hal.inria.fr/hal-01852956

[61] D. BZDOK, D. ENGEMANN, O. GRISEL, G. VAROQUAUX, B. THIRION. *Prediction and inference diverge in biomedicine: Simulations and real-world data*, April 2018, working paper or preprint [*DOI :* 10.1101/327437], https://hal.archives-ouvertes.fr/hal-01848319

[62] D. BZDOK, T. M. KARRER. *Single-Subject Prediction: A Statistical Paradigm for Precision Psychiatry*, February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01714822

[63] K. DADI, M. RAHIM, A. ABRAHAM, D. CHYZHYK, M. MILHAM, B. THIRION, G. VAROQUAUX. *Benchmarking functional connectome-based predictive models for resting-state fMRI*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01824205

[64] M. GABRIÉ, A. MANOEL, C. LUNEAU, J. BARBIER, N. MACRIS, F. KRZAKALA, L. ZDEBOROVÁ. *Entropy and mutual information in models of deep neural networks*, November 2018, working paper or preprint, https://hal-cea.archives-ouvertes.fr/cea-01930228

[65] J.-R. KING, L. . GWILLIAMS, C. . HOLDGRAF, J. . SASSENHAGEN, A. . BARACHANT, D. . ENGEMANN, E. . LARSON, A. GRAMFORT. *Encoding and Decoding Neuronal Dynamics: Methodological Framework to Uncover the Algorithms of Cognition*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01848442

[66] C. LAZARUS, P. WEISS, N. CHAUFFERT, F. MAUCONDUIT, L. EL GUEDDARI, C. DESTRIEUX, I. ZEMMOURA, A. VIGNAUD, P. CIUCIU. *Variable-density k-space filling curves for accelerated Magnetic Resonance Imaging*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01861760

[67] A. MANOEL, F. KRZAKALA, B. THIRION, G. VAROQUAUX, L. ZDEBOROVÁ. *Approximate message-passing for convex optimization with non-separable penalties*, November 2018, https://arxiv.org/abs/1809.06304 - working paper or preprint, https://hal-cea.archives-ouvertes.fr/cea-01932983

[68] A. MENSCH, J. MAIRAL, B. THIRION, G. VAROQUAUX. *Extracting Universal Representations of Cognition across Brain-Imaging Studies*, October 2018, https://arxiv.org/abs/1809.06035 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01874713

[69] D. WASSERMANN, D. V. NGUYEN, G. GALLARDO, J.-R. LI, W. CAI, V. MENON. *Sensing Von Economo Neurons in the Insula with Multi-shell Diffusion MRI*, 2018, International Society for Magnetic Resonance in Medicine, Poster, https://hal.inria.fr/hal-01807704

References in notes

[70] K. S. BUTTON, J. P. IOANNIDIS, C. MOKRYSZ, B. A. NOSEK, J. FLINT, E. S. ROBINSON, M. R. MUNAFÒ. *Power failure: why small sample size undermines the reliability of neuroscience*, in "Nature Reviews Neuroscience", 2013, vol. 14, n^{o} 5, pp. 365–376

[71] R. A. POLDRACK, C. I. BAKER, J. DURNEZ, K. J. GORGOLEWSKI, P. M. MATTHEWS, M. R. MUNAFÒ, T. E. NICHOLS, J.-B. POLINE, E. VUL, T. YARKONI. *Scanning the horizon: towards transparent and reproducible neuroimaging research*, in "Nature Reviews Neuroscience", 2017, vol. 18, n^{o} 2, pp. 115–126