



IN PARTNERSHIP WITH:
Centre CEA-Saclay

Activity Report 2017

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

1. Personnel	1
2. Overall Objectives	2
3. Research Program	3
3.1. Inverse problems in Neuroimaging	3
3.2. Multivariate decompositions	5
3.3. Covariance estimation	5
4. Application Domains	6
4.1.1. Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)	6
4.1.2. Analysis of brain Connectivity	6
4.1.3. Modeling of brain processes (MEG)	7
4.1.4. Current challenges in human neuroimaging (acquisition+analysis)	7
5. Highlights of the Year	7
6. New Software and Platforms	7
6.1. Mayavi	7
6.2. MedInria	8
6.3. Nilearn	8
6.4. PyHRF	8
6.5. Scikit-learn	9
6.6. MODL	9
6.7. MNE	9
7. New Results	10
7.1. Joint prediction of multiple scores captures better individual traits from brain images	10
7.2. Population-shrinkage of covariance to estimate better brain functional connectivity	10
7.3. Fast Regularized Ensembles of Models	11
7.4. time decoding	11
7.5. Hierarchical Region-Network Sparsity for High-Dimensional Inference in Brain Imaging	12
7.6. Cross-validation failure: small sample sizes lead to large error bars	12
7.7. Autoreject: Automated artifact rejection for MEG and EEG data	14
7.8. Learning Neural Representations of Human Cognition across Many fMRI Studies	15
7.9. SPARKLING: Novel Non-Cartesian Sampling Schemes for Accelerated 2D Anatomical Imaging at 7T Using Compressed Sensing	15
7.10. A Projection Method on Measures Sets	16
8. Bilateral Contracts and Grants with Industry	17
9. Partnerships and Cooperations	18
9.1. Regional Initiatives	18
9.1.1. CoSmic project	18
9.1.2. BrainAMP project	18
9.1.3. iConnectom project	19
9.1.4. MetaCog project	19
9.1.5. HighDimStat project	19
9.1.6. AMPHI project	20
9.1.7. CDS2	20
9.2. National Initiatives	20
9.2.1.1. MultiFrac project	20
9.2.1.2. Niconnect project	21
9.3. European Initiatives	22
9.4. International Initiatives	22
9.4.1.1. MetaMRI	22

9.4.1.2. LargeBrainNets	23
9.5. International Research Visitors	23
9.5.1. Visits of International Scientists	23
9.5.2. Visits to International Teams	24
9.5.2.1.1. Denis Engemann	24
9.5.2.1.2. Arthur Mensch	24
9.5.2.1.3. Jérôme Dockès	24
10. Dissemination	24
10.1. Promoting Scientific Activities	24
10.1.1. Scientific Events Selection	24
10.1.1.1.1. Gaël Varoquaux	24
10.1.1.1.2. Bertrand Thirion	24
10.1.2. Journal	24
10.1.2.1. Member of the Editorial Boards	24
10.1.2.2. Reviewer - Reviewing Activities	24
10.1.2.2.1. Gaël Varoquaux	24
10.1.2.2.2. Alexandre Gramfort	24
10.1.2.2.3. Philippe Ciuciu	25
10.1.2.2.4. Denis Engemann	25
10.1.2.2.5. Bertrand Thirion	25
10.1.3. Invited Talks	25
10.1.4. Scientific Expertise	25
10.1.5. Research Administration	26
10.2. Teaching - Supervision - Juries	26
10.2.1. Teaching	26
10.2.1.1. Gaël Varoquaux	26
10.2.1.2. Philippe Ciuciu	26
10.2.1.3. Alexandre Gramfort	26
10.2.1.4. Denis A. Engemann	27
10.2.1.5. Bertrand Thirion	27
10.2.2. Supervision	27
10.2.2.1. Philippe Ciuciu	27
10.2.2.2. Alexandre Gramfort	27
10.2.2.3. Denis A. Engemann	27
10.2.2.4. Bertrand Thirion	27
10.2.3. Juries	28
10.2.3.1. Philippe Ciuciu	28
10.2.3.2. Gaël Varoquaux	28
10.2.3.3. Alexandre Gramfort	28
10.2.3.4. Bertrand Thirion	28
10.3. Popularization	28
10.3.1. Gaël Varoquaux	28
10.3.2. Alexandre Gramfort	28
11. Bibliography	28

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. Personnel

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, Researcher, from Apr 2017, HDR]
- Shan Liu [Inria, Researcher, until Oct 2017]
- Gaël Varoquaux [Inria, Researcher]
- Demian Wassermann [Inria, Researcher, from Oct 2017]

Post-Doctoral Fellows

Darya Chyzyk [Inria]
Joke Durnez [Inria]
Daria La Rocca [CEA]
Antonio Andre Monteiro Manoel [CEA, until Nov 2017]
Mehdi Rahim [Inria, until Oct 2017]

PhD Students

Pierre Ablin [Inria, from Apr 2017]
Patricio Cerda Reyes [Inria]
Hamza Cherkaoui [CEA, from Nov 2017]
Jérôme-Alexis Chevalier [Inria, from Oct 2017]
Jérôme Dockès [Inria]
Elvis Dohmatob [Inria, until Oct 2017]
Loubna El Gueddari [Ministère de l'Enseignement Supérieur et de la Recherche]
Andres Hoyos Idrobo [Inria, until Apr 2017]
Hicham Janati [Inria, from Dec 2017]
Carole Lazarus [CEA]
Mathurin Massias [Inria, from Apr 2017]
Arthur Mensch [Univ Paris-Sud]

Technical staff

Kamalaker Reddy Dadi [CEA, until Sep 2017]
Loïc Esteve [Inria, until Nov 2017]
Ana Luisa Grilo Pinho [Inria]
Olivier Grisel [Inria]
Guillaume Lemaitre [Inria]
Joan Massich Vall [Inria, from Dec 2017]
Joris Van Den Bossche [Inria, from Apr 2017]

Interns

Hanae Carrie [Inria, from Jul 2017]
Arthur Imbert [Inria, from Mar 2017 until Aug 2017]
Meghana Madhyastha [Inria, from May 2017 until Aug 2017]
Hugo Richard [Ecole Normale Supérieure Lyon, from Apr 2017 until Jul 2017]

Administrative Assistants

Tiffany Caristan [Inria, until Aug 2017]
Corinne Petitot [Inria, from Oct 2017]

External Collaborators

Sergul Aydore [Amazon, from Apr 2017]
Joseph Salmon [Institut Telecom ex GET Groupe des Ecoles des Télécommunications , from May 2017]

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 $\mathbf{Y}_1, \dots, \mathbf{Y}_{n_{features}}$ of \mathbf{Y} independently, yielding $n_{features}$ problems to be solved in parallel:

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

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:

$$\hat{\mathbf{w}}_i = \operatorname{argmin}_{\mathbf{w}_i} \|\mathbf{Y}_i - \mathbf{X}\mathbf{w}_i\|^2 + \Psi(\mathbf{w}_i), \quad (1)$$

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

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

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.

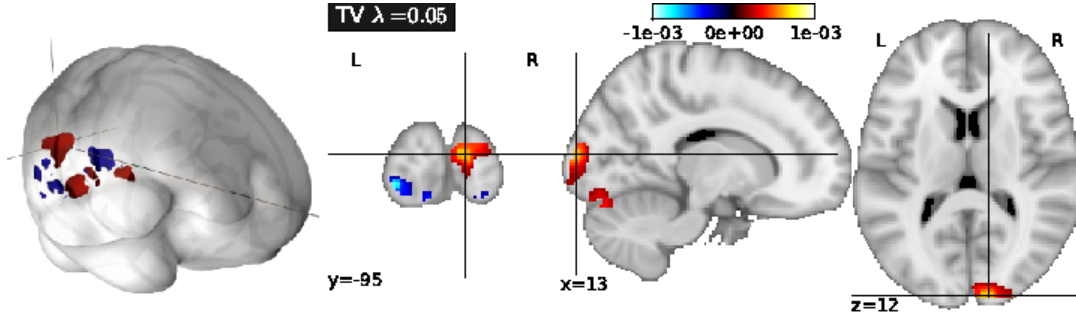


Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problem here is to predict the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the presentation of an image. Learning and test are performed across 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 $i \in \{1, \dots, n_{features}\}$. This can be computed through cross-validation, by *learning* $\hat{\mathbf{w}}_i$ on some part of the dataset, and then estimating $\|\mathbf{Y}_i - \mathbf{X}\hat{\mathbf{w}}_i\|^2$ 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 .

$$\mathbf{Y} = \mathbf{X}\mathbf{w} + \epsilon, \quad (3)$$

then

$$\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}=(\mathbf{w}_i), i=1..n_f} \sum_{i=1}^{n_f} \|\mathbf{Y}_i - \mathbf{X}\mathbf{w}_i\|^2 + \lambda \sum_{j=1}^{n_{voxels}} \sqrt{\sum_{i=1}^{n_f} \mathbf{w}_{i,j}^2} \quad (4)$$

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

$$\mathbf{X} = \mathbf{A}\mathbf{D} + \epsilon, \quad (5)$$

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 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:

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

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 [29].

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.

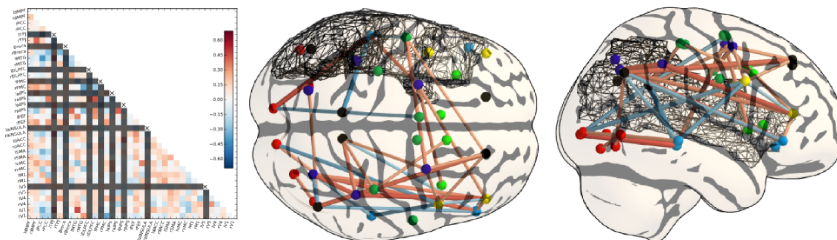


Figure 2. Example of functional connectivity analysis: The correlation matrix describing brain functional connectivity in a post-stroke patient (lesion volume outlined as a mesh) is compared to a group of control subjects. Some edges of the graphical model show a significant difference, but the statistical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

4. 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 electrophysiological 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?

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, Magnetoencephalography (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 spatial¹ 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 improvements² 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 [54], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [55]): 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. ERC

Demian Wassermann obtained an ERC starting grant, NeuroLang, *Accelerating Neuroscience Research by Unifying Knowledge Representation and Analysis Through a Domain Specific Language*.

Besides, Alexandre Gramfort joined Parietal just after the start of his ERC grant entitled SLAB, *Signal processing and Learning Applied to Brain data*.

6. New Software and Platforms

6.1. Mayavi

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

²The 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 paid by a factor of 8 in terms of SNR.

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: It 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: Health - Brain - IRM - Neurosciences - Statistic analysis - FMRI - Medical imaging

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.

- 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
- URL: <http://pyhrf.org>

6.5. Scikit-learn

KEYWORDS: Regression - 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.

- Contact: Alexandre Gramfort
- URL: <http://martinos.org/mne/>

7. New Results

7.1. Joint prediction of multiple scores captures better individual traits from brain images

To probe individual variations in brain organization, population imaging relates features of brain images to rich descriptions of the subjects such as genetic information or behavioral and clinical assessments. Capturing common trends across these measurements is important: they jointly characterize the disease status of patient groups. In particular, mapping imaging features to behavioral scores with predictive models opens the way toward more precise diagnosis. Here we propose to jointly predict all the dimensions (behavioral scores) that make up the individual profiles, using so-called multi-output models. This approach often boosts prediction accuracy by capturing latent shared information across scores. We demonstrate the efficiency of multi-output models on two independent resting-state fMRI datasets targeting different brain disorders (Alzheimer’s Disease and schizophrenia). Furthermore, the model with joint prediction generalizes much better to a new cohort: a model learned on one study is more accurately transferred to an independent one. Finally, we show how multi-output models can easily be extended to multi-modal settings, combining heterogeneous data sources for a better overall accuracy.

More information can be found in Fig. 3 in [46].

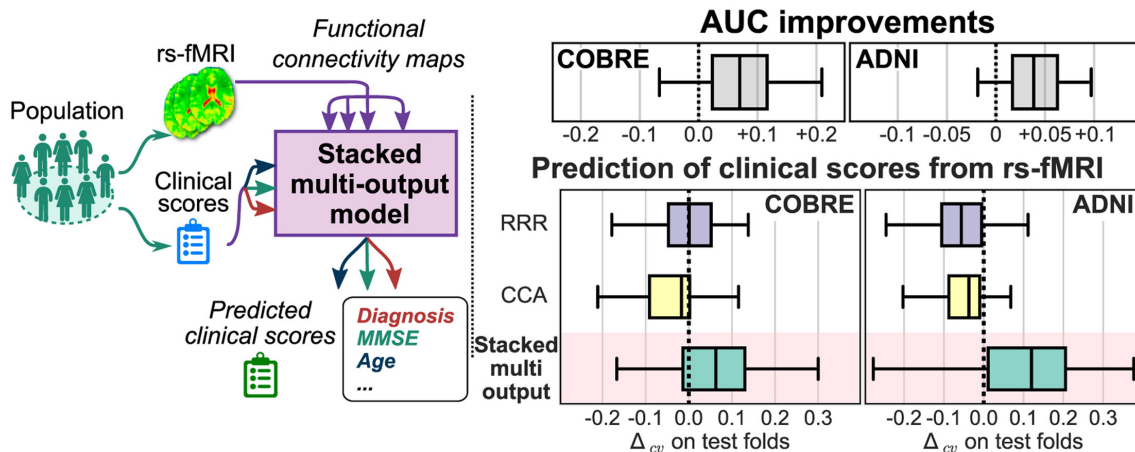


Figure 3. Joint prediction of multiple scores captures better individual traits from brain images

7.2. Population-shrinkage of covariance to estimate better brain functional connectivity

Brain functional connectivity, obtained from functional Magnetic Resonance Imaging at rest (r-fMRI), reflects inter-subject variations in behavior and characterizes neuropathologies. It is captured by the covariance matrix between time series of remote brain regions. With noisy and short time series as in r-fMRI, covariance estimation calls for penalization, and shrinkage approaches are popular. Here we introduce a new covariance estimator based on a non-isotropic shrinkage that integrates prior knowledge of the covariance distribution over a large population. The estimator performs shrinkage tailored to the Riemannian geometry of symmetric positive definite matrices, coupled with a probabilistic modeling of the subject and population covariance

distributions. Experiments on a large-scale dataset show that such estimators resolve better intra-and inter-subject functional connectivities compared existing co-variance estimates. We also demonstrate that the estimator improves the relationship across subjects between their functional-connectivity measures and their behavioral assessments. More information can be found in Fig. 4 in [47].

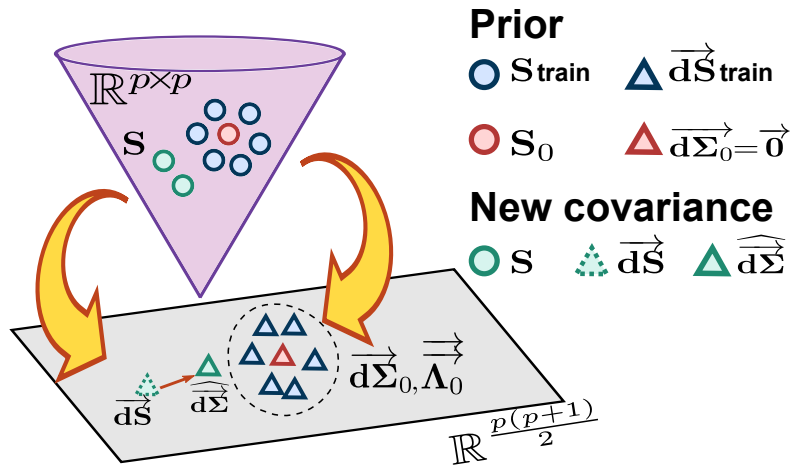


Figure 4. (a) Shrunken embedding estimation workflow: the empirical covariance is estimated from r-fMRI time-series; it is projected onto a tangent space built from a prior population; the embedding is then shrunk towards the prior $(\vec{d\Sigma}_0, \vec{\Lambda}_0)$. (b) Principle of tangent embedding shrinkage towards population distribution.

7.3. Fast Regularized Ensembles of Models

Brain decoding relates behavior to brain activity through predictive models. These are also used to identify brain regions involved in the cognitive operations related to the observed behavior. Training such multivariate models is a high-dimensional statistical problem that calls for suitable priors. State of the art priors –eg small total-variation– enforce spatial structure on the maps to stabilize them and improve prediction. However, they come with a hefty computational cost. We build upon very fast dimension reduction with spatial structure and model ensembling to achieve decoders that are fast on large datasets and increase the stability of the predictions and the maps. Our approach, fast regularized ensemble of models (FReM), includes an implicit spatial regularization by using a voxel grouping with a fast clustering algorithm. In addition, it aggregates different estimators obtained across splits of a cross-validation loop, each time keeping the best possible model. Experiments on a large number of brain imaging datasets show that our combination of voxel clustering and model ensembling improves decoding maps stability and reduces the variance of prediction accuracy. Importantly, our method requires less samples than state-of-the-art methods to achieve a given level of prediction accuracy. Finally, FReM is highly parallelizable, and has lower computation cost than other spatially-regularized methods.

More information can be found in Fig. 5 in [23].

7.4. time decoding

Most current functional Magnetic Resonance Imaging (fMRI) decoding analyses rely on statistical summaries of the data resulting from a deconvolution approach: each stimulation event is associated with a brain response. This standard approach leads to simple learning procedures, yet it is ill-suited for decoding events with short

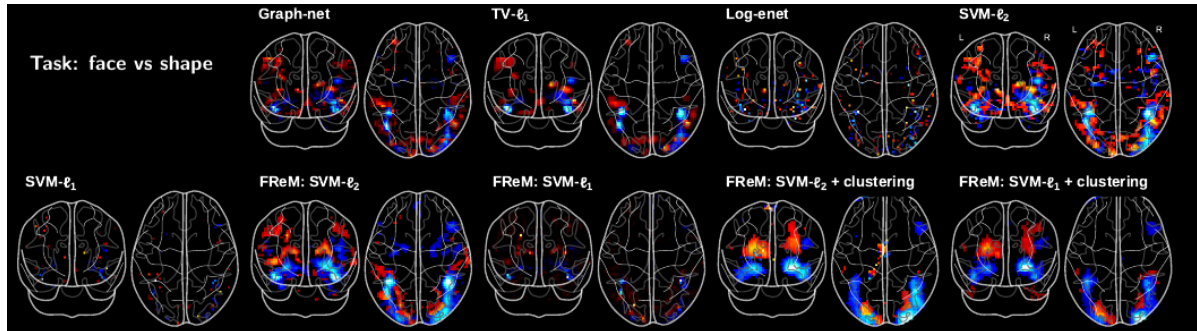


Figure 5. Qualitative comparison of decoder weight maps: Weight maps for different discriminative tasks on the HCP dataset. The maps are thresholded at the 99 percentile for visualization purposes. These correspond to a face-recognition task. The weight maps obtained with TV-L1 and FReM methods with clustering outline the organization of the functional areas of the visual mosaic, such as: primary visual areas, lateral occipital complex, the face and place specific regions in the fusiform gyrus.

inter-stimulus intervals. In order to overcome this issue, we propose a novel framework that separates the spatial and temporal components of the prediction by decoding the fMRI time-series continuously, i.e. scan-by-scan. The stimulation events can then be identified through a deconvolution of the reconstructed time series. We show that this model performs as well as or better than standard approaches across several datasets, most notably in regimes with small inter-stimuli intervals (3 to 5s), while also offering predictions that are highly interpretable in the time domain. This opens the way toward analyzing datasets not normally thought of as suitable for decoding and makes it possible to run decoding on studies with reduced scan time.

More information can be found in Fig. 6 in [28].

7.5. Hierarchical Region-Network Sparsity for High-Dimensional Inference in Brain Imaging

Structured sparsity penalization has recently improved statistical models applied to high-dimensional data in various domains. As an extension to medical imaging, the present work incorporates priors on network hierarchies of brain regions into logistic-regression to distinguish neural activity effects. These priors bridge two separately studied levels of brain architecture: functional segregation into regions and functional integration by networks. Hierarchical region-network priors are shown to better classify and recover 18 psychological tasks than other sparse estimators. Varying the relative importance of region and network structure within the hierarchical tree penalty captured complementary aspects of the neural activity patterns. Local and global priors of neurobiological knowledge are thus demonstrated to offer advantages in generalization performance, sample complexity, and domain interpretability.

More information can be found in Fig. 7 in [48].

7.6. Cross-validation failure: small sample sizes lead to large error bars

Predictive models ground many state-of-the-art developments in statistical brain image analysis: decoding, MVPA, searchlight, or extraction of biomarkers. The principled approach to establish their validity and usefulness is cross-validation, testing prediction on unseen data. Here, we raise awareness on error bars of cross-validation, which are often underestimated. Simple experiments show that sample sizes of many neuroimaging studies inherently lead to large error bars, eg $\pm 10\%$ for 100 samples. The standard error

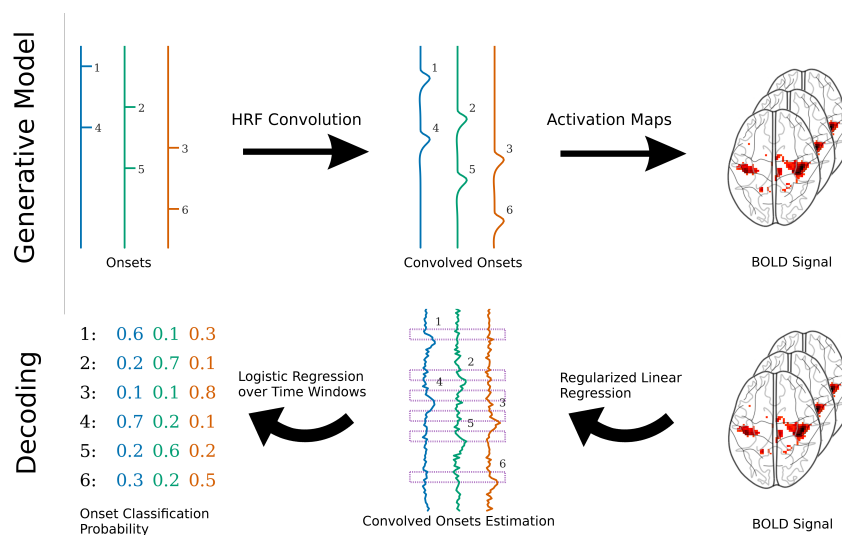


Figure 6. Schema of the time-domain decoding model. Straight arrows represent generative steps, while curved ones represent estimation steps.

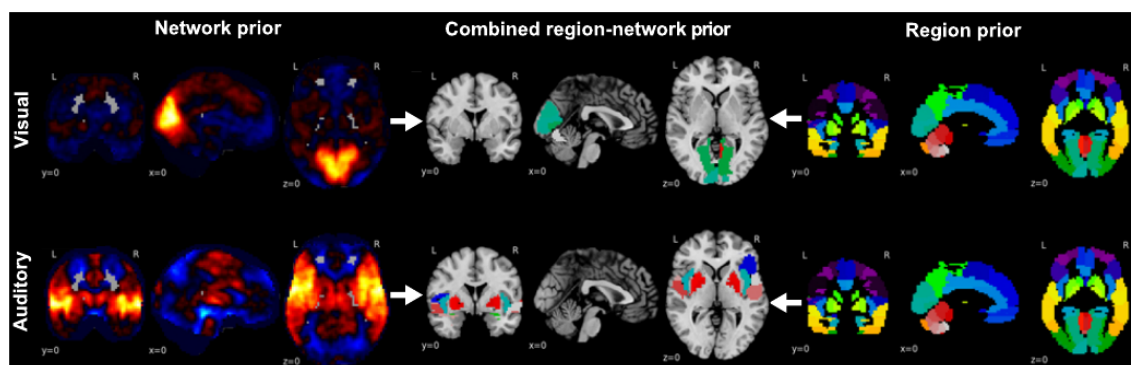


Figure 7. Building blocks of the hierarchical region-network tree. Displays the a-priori neurobiological knowledge introduced into the classification problem by hierarchical structured sparsity. Left: Continuous, partially overlapping brain network priors (hot-colored) accommodate the functional integration perspective of brain organization. Right: Discrete, non-overlapping brain region priors (single-colored) accommodate the functional segregation perspective. Middle: These two types of predefined voxel groups are incorporated into a joint hierarchical prior of parent networks with their descending region child nodes. Top to bottom: Two exemplary region-network priors are shown, including the early cortices that process visual and sound information from the environment.

across folds strongly underestimates them. These large error bars compromise the reliability of conclusions drawn with predictive models, such as biomarkers or methods developments where, unlike with cognitive neuroimaging MVPA approaches, more samples cannot be acquired by repeating the experiment across many subjects. Solutions to increase sample size must be investigated, tackling possible increases in heterogeneity of the data.

More information can be found in Fig. 8 in [33].

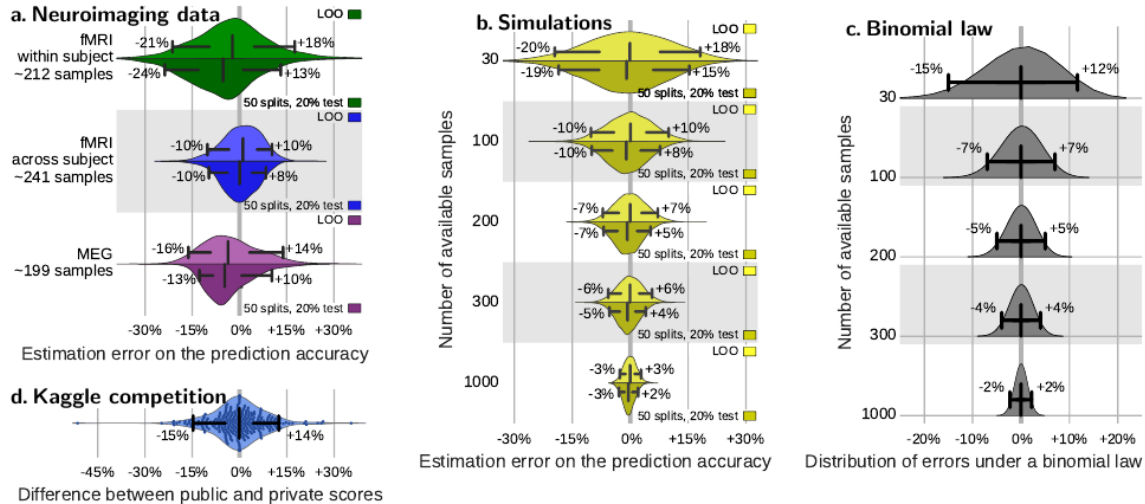


Figure 8. Cross-validation errors. *a* – Distribution of errors between the prediction accuracy as assessed via cross-validation (average across folds) and as measured on a large independent test set for different types of neuroimaging data. *b* – Distribution of errors between the prediction accuracy as assessed via cross-validation on data of various sample sizes and as measured on 10 000 new data points for simple simulations. *c* – Distribution of errors as given by a binomial law: difference between the observed prediction error and the population value of the error, $p = 75\%$, for different sample sizes. *d* – Discrepancies between private and public score. Each dot represents the difference between the accuracy of a method on the public test data and the private one. The scores are retrieved from <http://www.kaggle.com/c/mlsp-2014-mri>, in which 144 subjects were used total, 86 for training predictive model, 30 for the public test set, and 28 for the private test set. The bar and whiskers indicate the median and the 5th and 95th percentile. Measures on cross-validation (*a* and *b*) are reported for two reasonable choices of cross-validation strategy: leave one out (leave one run out or leave one subject out in data with multiple runs or subjects), or 50-times repeated splitting of 20% of the data.

7.7. Autoreject: Automated artifact rejection for MEG and EEG data

We present an automated algorithm for unified rejection and repair of bad trials in magnetoencephalography (MEG) and electroencephalography (EEG) signals. Our method capitalizes on cross-validation in conjunction with a robust evaluation metric to estimate the optimal peak-to-peak threshold – a quantity commonly used for identifying bad trials in M/EEG. This approach is then extended to a more sophisticated algorithm which estimates this threshold for each sensor yielding trial-wise bad sensors. Depending on the number of bad sensors, the trial is then repaired by interpolation or by excluding it from subsequent analysis. All steps of the algorithm are fully automated thus lending itself to the name Autoreject. In order to assess the practical significance of the algorithm, we conducted extensive validation and comparisons with state-of-the-art methods on four public datasets containing MEG and EEG recordings from more than 200

subjects. The comparisons include purely qualitative efforts as well as quantitatively benchmarking against human supervised and semi-automated preprocessing pipelines. The algorithm allowed us to automate the preprocessing of MEG data from the Human Connectome Project (HCP) going up to the computation of the evoked responses. The automated nature of our method minimizes the burden of human inspection, hence supporting scalability and reliability demanded by data analysis in modern neuroscience.

More information can be found in Fig. 9 and in [24].

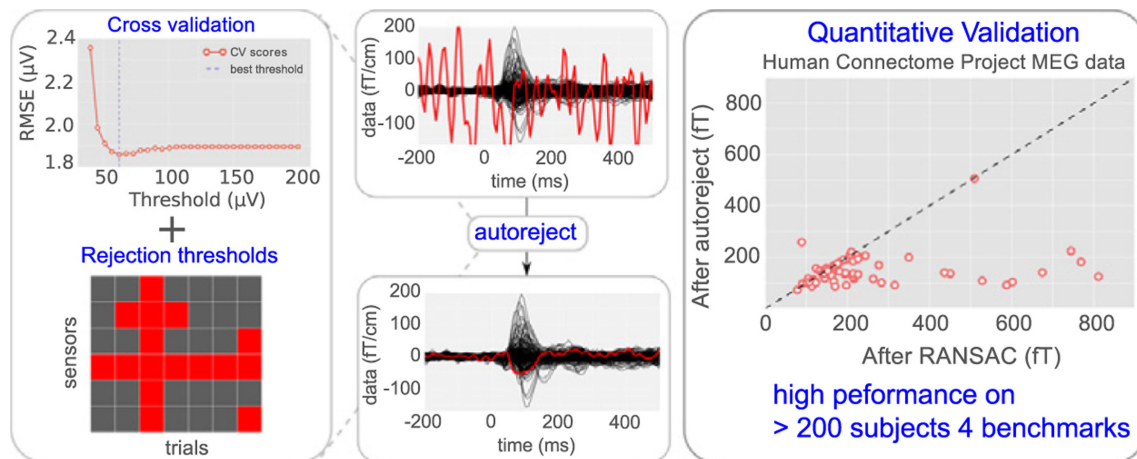


Figure 9. Autoreject: Automated artifact rejection for MEG and EEG data.

7.8. Learning Neural Representations of Human Cognition across Many fMRI Studies

Cognitive neuroscience is enjoying rapid increase in extensive public brain-imaging datasets. It opens the door to large-scale statistical models. Finding a unified perspective for all available data calls for scalable and automated solutions to an old challenge: how to aggregate heterogeneous information on brain function into a universal cognitive system that relates mental operations/cognitive processes/psychological tasks to brain networks? We cast this challenge in a machine-learning approach to predict conditions from statistical brain maps across different studies. For this, we leverage multi-task learning and multi-scale dimension reduction to learn low-dimensional representations of brain images that carry cognitive information and can be robustly associated with psychological stimuli. Our multi-dataset classification model achieves the best prediction performance on several large reference datasets, compared to models without cognitive-aware low-dimension representations; it brings a substantial performance boost to the analysis of small datasets, and can be introspected to identify universal template cognitive concepts.

More information can be found in Fig. 10 in [45].

7.9. SPARKLING: Novel Non-Cartesian Sampling Schemes for Accelerated 2D Anatomical Imaging at 7T Using Compressed Sensing

We have presented for the first time the implementation of non-Cartesian trajectories on a 7T scanner for 2D anatomical imaging. The proposed SPARKLING curves (Segmented Projection Algorithm for Random K-space samPLING) are a new type of non-Cartesian segmented sampling trajectories which allow fast and

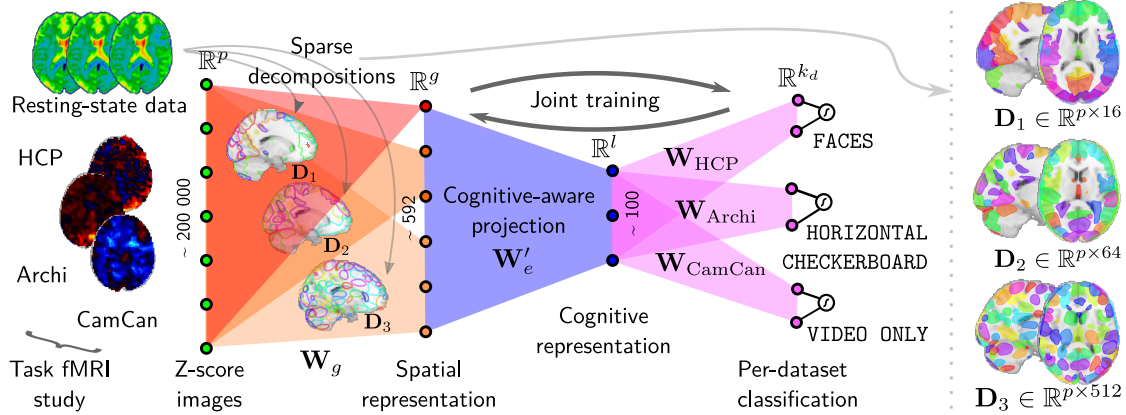


Figure 10. **Model architecture: Three-layer multi-dataset classification.** The first layer (orange) is learned from data acquired outside of cognitive experiments and captures a spatially coherent signal at multiple scales, the second layer (blue) embeds these representations in a space common to all datasets, from which the conditions are predicted (pink) from multinomial models.

efficient coverage of the k-space according to a chosen variable density. To demonstrate their potential, a high-resolution ($0.4 \times 0.4 \times 3.0\text{mm}^3$) T2*-weighted image was acquired with an 8-fold undersampled SPARKLING trajectory. Images were reconstructed using non-linear iterative reconstructions derived from the Compressed Sensing theory.

More information can be found in Fig. 11 in [43], [42].

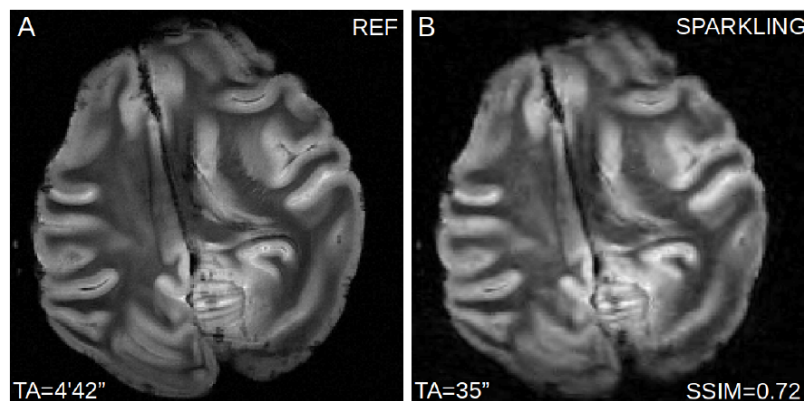


Figure 11. T2*-weighted Gradient-Echo transversal slice of the ex vivo baboon brain is displayed for A) the full Cartesian reference sampling lasting 4 minutes and 42 seconds and B) the presented 8-fold undersampled SPARKLING trajectories lasting only 35 seconds. A good visual quality of the subsampled reconstruction was observed with the preservation of major folded patterns. The structural similarity score between the SPARKLING reconstructions and the reference was also very satisfactory.

7.10. A Projection Method on Measures Sets

We consider the problem of projecting a probability measure π on a set \mathcal{M}_N of Radon measures. The projection is defined as a solution of the following variational problem:

$$\inf_{\mu \in \mathcal{M}_N} \|h * (\mu - \pi)\|_2^2,$$

where $h \in L^2(\Omega)$ is a kernel, $\Omega \subset \mathbb{R}^d$ and denotes the convolution operator. To motivate and illustrate our study, we show that this problem arises naturally in various practical image rendering problems such as stippling (representing an image with N dots) or continuous line drawing (representing an image with a continuous line). We provide a necessary and sufficient condition on the sequence $(\mathcal{M}_N)_{N \in \mathbb{N}}$ that ensures weak convergence of the projections $(\mu_N^*)_{N \in \mathbb{N}}$ to π . We then provide a numerical algorithm to solve a discretized version of the problem and show several illustrations related to computer-assisted synthesis of artistic paintings/drawings.

More information can be found in [19].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. The Wendelin FUI project

The Wendelin project has been granted on December 3rd, 2014. It has been selected at the *Programme d'Investissements d'Avenir (PIA)* that supports "cloud computing et Big Data". It gives visibility and fosters the French technological big data sector, and in particular the scikit-learn library, the NoSQL "NEO" et the decentralized "SlapOS" cloud, three open-source software supported by the Systematic *pôle de compétitivité*.

Scikit-learn is a worldwide reference library for machine learning. Gaël Varoquaux, Olivier Grisel and Alexandre Gramfort have been major players in the design of the library and Scikit-learn has then been supported by the growing scientific Python community. It is currently used by major internet companies as well as dynamic start-ups, including Google, Airbnb, Spotify, Evernote, AWeber, TinyClues; it wins more than half of the data science "Kaggle" competitions. Scikit-learn makes it possible to predict future outcomes given a training data, and thus to optimize company decisions. Almost 1 million euros will be invested to improve the algorithmic core of scikit-learn through the Wendelin project thanks to the Inria, ENS and Institut Mines Télécom teams. In particular, scikit-learn will be extended in order to ease online prediction and to include recent stochastic gradient algorithms.

NEO is the native NoSQL base of the Python language. It was initially designed by Nexedi and is currently used and embedded in the main software of company information systems. More than one million euros will be invested into NEO, so that scikit-learn can process within 10 years (out-of-core) data of 1 exabyte size.

Paris13 university and the Mines Télécom institute will extend the SlapOS distributed mesh cloud to deploy Wendelin in *Big Data as a Service (BDaaS)* mode, to achieve the interoperability between the Grid5000 and Teralab infrastructures and to extend the cloud toward smart sensor systems.

The combination of scikit-learn, NEO and SlapOS will improve the predictive maintenance of industrial plants with two major use cases: connected windmills (GDF SUEZ, Woelfel) and customer satisfaction in car sale systems (MMC Rus). In both cases it is about non-personal, yet profitable big data. The Wendelin project actually demonstrates that Big data can improve infrastructure and everyday-life equipment without intrusive data collection. For more information, please see <http://www.wendelin.io>.

The project partners are:

- Nexedi (leader)
- GDF SUEZ
- Abilian
- 2ndQuadrant
- Institut Mines Télécom
- Inria
- Université Paris 13

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. *CoSmic project*

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

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.

9.1.2. *BrainAMP project*

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Antonio Andre Monteiro Manoel.

This is a collaborative project with Lenka Zdeborová, Theoretical Physics Institute (CEA) funded by the DRF-impulsion CEA program.

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 high-field MRI 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 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. BrainAMP 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.

9.1.3. *iConnectom project*

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Elvis Dohmatob.

This is a Digiteo project (2014-2017).

Mapping brain functional connectivity from functional Magnetic Resonance Imaging (MRI) data has become a very active field of research. However, analysis tools are limited and many important tasks, such as the empirical definition of brain networks, remain difficult due to the lack of a good framework for the statistical modeling of these networks. We propose to develop population models of anatomical and functional connectivity data to improve the alignment of subjects brain structures of interest while inferring an average template of these structures. Based on this essential contribution, we will design new statistical inference procedures to compare the functional connections between conditions or populations and improve the sensitivity of connectivity analysis performed on noisy data. Finally, we will test and validate the methods on multiple datasets and distribute them to the brain imaging community.

9.1.4. *MetaCog project*

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

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.

9.1.5. *HighDimStat project*

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

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.

9.1.6. 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 position for Joris van den Bossche.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. MultiFracs project

Participants: Philippe Ciuciu [Correspondant], Daria La Rocca.

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.

9.2.1.2. Niconnect project

Participants: Bertrand Thirion, Gaël Varoquaux [Correspondant], Kamalaker Reddy Dadi, Darya Chyzyk, Mehdi Rahim.

- **Context:** The NiConnect project (2012-2017) arises from an increasing need of medical imaging tools to diagnose efficiently brain pathologies, such as neuro-degenerative and psychiatric diseases or lesions related to stroke. Brain imaging provides a non-invasive and widespread probe of various features of brain organization, that are then used to make an accurate diagnosis, assess brain rehabilitation, or make a prognostic on the chance of recovery of a patient. Among different measures extracted from brain imaging, functional connectivity is particularly attractive, as it readily probes the integrity of brain networks, considered as providing the most complete view on brain functional organization.
- **Challenges:** To turn methods research into popular tool widely usable by non specialists, the NiConnect project puts specific emphasis on producing high-quality open-source software. NiConnect addresses the many data analysis tasks that extract relevant information from resting-state fMRI datasets. Specifically, the scientific difficulties are *i)* conducting proper validation of the models and tools, and *ii)* providing statistically controlled information to neuroscientists or medical doctors. More importantly, these procedures should be robust enough to perform analysis on limited quality data, as acquiring data on diseased populations is challenging and artifacts can hardly be controlled in clinical settings.
- **Outcome of the project:** In the scope of computer science and statistics, NiConnect pushes forward algorithms and statistical models for brain functional connectivity. In particular, we are investigating structured and multi-task graphical models to learn high-dimensional multi-subject brain connectivity models, as well as spatially-informed sparse decompositions for segmenting structures from brain imaging. With regards to neuroimaging methods development, NiConnect provides systematic comparisons and evaluations of connectivity biomarkers and a software library embedding best-performing state-of-the-art approaches. Finally, with regards to medical applications, the NiConnect project also plays a support role in on going medical studies and clinical trials on neurodegenerative diseases.
- **Consortium**
 - Parietal Inria research team: applied mathematics and computer science to model the brain from MRI
 - LIF INSERM research team: medical image data analysis and modeling for clinical applications

- CATI center: medical image processing center for large scale brain imaging studies
- Henri-Mondor hospital neurosurgery and neuroradiology: clinical teams conducting research on treatments for neurodegenerative diseases, in particular Huntington and Parkinson diseases
- Logilab: consulting in scientific computing

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. HBP

Title: The Human Brain Project

Programm: FP7

Duration: October 2013 - September 2016

Coordinator: EPFL

Partners:

- Inria contact: Olivier Faugeras
- Understanding the human brain is one of the greatest challenges facing 21st century science. If we can rise to the challenge, we can gain profound insights into what makes us human, develop new treatments for brain diseases and build revolutionary new computing technologies. Today, for the first time, modern ICT has brought these goals within sight. The goal of the Human Brain Project, part of the FET Flagship Programme, is to translate this vision into reality, using ICT as a catalyst for a global collaborative effort to understand the human brain and its diseases and ultimately to emulate its computational capabilities. The Human Brain Project will last ten years and will consist of a ramp-up phase (from month 1 to month 36) and subsequent operational phases. This Grant Agreement covers the ramp-up phase. During this phase the strategic goals of the project will be to design, develop and deploy the first versions of six ICT platforms dedicated to Neuroinformatics, Brain Simulation, High Performance Computing, Medical Informatics, Neuromorphic Computing and Neurorobotics, and create a user community of research groups from within and outside the HBP, set up a European Institute for Theoretical Neuroscience, complete a set of pilot projects providing a first demonstration of the scientific value of the platforms and the Institute, develop the scientific and technological capabilities required by future versions of the platforms, implement a policy of Responsible Innovation, and a programme of transdisciplinary education, and develop a framework for collaboration that links the partners under strong scientific leadership and professional project management, providing a coherent European approach and ensuring effective alignment of regional, national and European research and programmes. The project work plan is organized in the form of thirteen subprojects, each dedicated to a specific area of activity. A significant part of the budget will be used for competitive calls to complement the collective skills of the Consortium with additional expertise.

9.4. International Initiatives

9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

9.4.1.1. MetaMRI

Title: Machine learning for meta-analysis of functional neuroimaging data

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Department of Psychology - Russ Poldrack

Start year: 2015

See also: <https://team.inria.fr/metamri>

Neuroimaging produces huge amounts of complex data that are used to better understand the relations between brain structure and function. Observing that the neuroimaging community is still largely missing appropriate tools to store and organize the knowledge related to the data, Parietal team and Poldrack's lab, have decided to join forces to set up a framework for functional brain image meta-analysis, i.e. a framework in which several datasets can be jointly analyzed in order to accumulate information on the functional specialization of brain regions. MetaMRI will build upon Poldrack's lab expertise in handling, sharing and analyzing multi-protocol data and Parietal's recent developments of machine learning libraries to develop a new generation of meta-analytic tools.

9.4.1.2. *LargeBrainNets*

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

International Partner (Institution - Laboratory - Researcher):

Stanford (United States) - Stanford Cognitive and Systems Neuroscience Laboratory -
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.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Parietal has welcome François Meyer, Univ Colorado at Boulder, for a six months visit (Jan-June 2017), funded by a D'Alembert fellowship of Paris Saclay University. The project of François is to assess novel statistical models of functional connectivity based on the generalized resistivity model he has developed within a graph theoretical framework.

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

9.5.2.1.1. Denis Engemann

has spent two months in Boston (April-May) with the MEG Core lab, Athinoula A. Martinos Center (MGH/Harvard-MIT) working on functional connectivity methods and population analysis for MEG.

9.5.2.1.2. Arthur Mensch

has spent 3 months in Japan (Sept-Dec) with NTT, working on dynamic time warping problems with Mathieu Blondel.

9.5.2.1.3. Jérôme Dockès

has spent two months with Poldracklab at Stanford, as part of the MetaMRI associated team. He has worked on the statistical relationships between neuroscientific concepts (whether anatomical or cognitive) and brain activation loci.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

10.1.1.1. Member of the Conference Program Committees

10.1.1.1.1. Gaël Varoquaux

MICCAI, NIPS, IPMI, ICML, CCN

10.1.1.1.2. Bertrand Thirion

NIPS, IPMI, OHBM, PRNI

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Gaël Varoquaux, Editor, Neuroimage
- Alexandre Gramfort, Editor, Frontiers in Brain Imaging Methods, Journal of Machine Learning Research (JMLR)
- Bertrand Thirion, Editor, Frontiers in Brain Imaging Methods

10.1.2.2. Reviewer - Reviewing Activities

10.1.2.2.1. Gaël Varoquaux

Nature Methods, JMLR, PLOS Comp Bio, NeuroImage, IEEE TBME, Annals of Applied Statistics, Biological Psychiatry, MedIA

10.1.2.2.2. Alexandre Gramfort

JMLR, PLOS Comp Bio, NeuroImage, IEEE TBME, IEEE TMI, IEEE TSP, MedIA, Brain Topography, NIPS, ICML, ICLR.

10.1.2.2.3. Philippe Ciuciu

PLOS One, NeuroImage, HBM, IEEE TMI, IEEE TSP, IEEE SP letters, MedIA, J of Neuroscience, JCBFM, ISBI, ICASSP, EUSIPCO.

10.1.2.2.4. Denis Engemann

PLOS Biology, PLOS Computational Biology, Nature Scientific Reports, Neuroimage, Neuroimage Clinical, Human Brain Mapping, Journal of Machine Learning Research, Brain Topography, Brain Connectivity, Journal of Alzheimer's Disease, PLOS ONE, Frontiers in Neuroscience, Psychiatry and Clinical Neurosciences, Sensors

10.1.2.2.5. Bertrand Thirion

Nature communications, Neuroimage, Medical Image Analysis, IEEE TMI, PNAS, PLOS Comp Bio.

10.1.3. Invited Talks

- Gaël Varoquaux, Keynote, BDEC Workshop 2017, Wuxi, China
- Gaël Varoquaux, Keynote, scipy 2017, Austin TX
- Gaël Varoquaux, Keynote, Swiss Python summit 2017, Zurich, Switzerland
- Gaël Varoquaux, Keynote, NIPS learning with limited labels workshop, Long Beach, CA
- Philippe Ciuciu, Seminar in the neuroscience Department, NYU, School of Medicine, May 2017, NYC.
- Philippe Ciuciu, Laufer Center Seminar Series, May 2017, Stony Brook Univ, NY.
- Philippe Ciuciu, Dedale workshop on Dictionary learning on manifolds, Sep. 2017, Nice, France
- Philippe Ciuciu, Colloquium at Perform centre, Dec 2017, Montreal, Canada
- Philippe Ciuciu, Seminar at Ecole Polytechnique Montreal, Dec 2017, Montreal, Canada
- Alexandre Gramfort, 2017 EU-US Frontiers of Engineering Symposium, UC Davis, CA
- Alexandre Gramfort, Computational Challenges in Image Processing, Turing Institute, Cambridge Univ., UK
- Alexandre Gramfort, Machine learning for functional brain imaging Symposium and Workshop, Karolinska Institute, Sweden
- Denis A. Engemann, Lessons learned from high-dimensional statistics and M/EEG: From automated preprocessing to detection of consciousness, University of Naples, Italy, March 2017
- Denis A. Engemann, Learning from Oscillations – New Vistas for Translational Neuroimaging, MIT, Cambridge MA, April 2017
- Denis A. Engemann, Building better biomarkers using M/EEG and statistical learning, Eespo University, Finland, June 2017
- Bertrand Thirion, ARSEP, February 2017, Paris
- Bertrand Thirion, AI for Medical Imaging symposium, June 2017, Paris
- Bertrand Thirion, Causality and big data workshop, October 2017, Saclay
- Bertrand Thirion, Machine learning workshop, Telcom Paris Sud, October 2017, Evry
- Bertrand Thirion, Icube seminary, Strasburg, July 2017
- Bertrand Thirion, Laboratoire de Mathématiques d'Orsay, December 2017

10.1.4. Scientific Expertise

- Philippe Ciuciu, Member of the "ANR evaluation committee (CES 45) in Mathématique, informatique, automatique, traitement du signal"
- Philippe Ciuciu, Member of the IEEE BISP committee for the international symposium in biomedical imaging

- Philippe Ciuciu, Member of the Biomedical Imaging & Signal Analytics special area team (SAT) in Eurasip
- Bertrand Thirion, Member of the prospective group of ITMO Neurosciences

10.1.5. Research Administration

- Gaël Varoquaux, Member of the Inria Saclay CDT, "Comission de développement technologique"
- Gaël Varoquaux, Member of the Inria Saclay CSD, "Comission de suivi doctoral"
- Gaël Varoquaux, Member of the Inria Saclay cluster committee
- Alexandre Gramfort, Member of the "comité de pilotage" DataIA
- Alexandre Gramfort, Member of the "comité de pilotage" DIM RFSI
- Philippe Ciuciu, Member of the "Inria Saclay scientific committee"
- Philippe Ciuciu, Member of the CEA/DRF Impulsion evaluation committee
- Philippe Ciuciu, Member of the "jury du prix de thèse de la société EEA et du GDR ISIS"
- Bertrand Thirion, Deputy scientific director 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.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

10.2.1.1. Gaël Varoquaux

Master : Machine learning with scikit-learn, 2 heures équivalent TD, M1, ENSAE, France
 Master : Brain functional connectivity, 8 heures équivalent TD, M2, Télécom Paris Tech, France
 Doctorat : Machine learning for brain imaging with nilearn, 16 heures équivalent TD, Université de Montréal, Canada
 Doctorat : Ecole d'hiver, Computational Brain Connectivity Mapping, Juan-Les-Pins, 1h30
 Doctorat : Ecole d'hiver, Computational Brain Connectivity Mapping, 1h30, Juan-Les-Pins
 Doctorat : Machine learning for brain imaging, 1h00, IPAM, UCLA, Los Angeles, USA
 Doctorat : Scipy: numerical algorithms in Python, 1h30, Euroscipy 2017, Erlangen, Germany
 Doctorat : Estimation of brain functional connectomes, 30mn, OHBM 2017, Vancouver, Canada
 Doctorat : Software engineering for reproducible science, Open science summer school 2017, EPFL, Lausanne, Switzerland

10.2.1.2. Philippe Ciuciu

Master 2 : "Functional MRI: From data acquisition to analysis", 3h, Univ. Paris V René Descartes & Télécom-Paristech, Master of Biomedical Engineering
 Master 2 : "fMRI data analysis", 3h, Univ. Paris-Saclay, Master of medical Physics

10.2.1.3. Alexandre Gramfort

Master 2 : "Functional Brain Imaging with EEG and MEG", 6h, Univ. Paris V René Descartes & Télécom-Paristech, Master of Biomedical Engineering
 Master 2 : "Optimization for Data Science", 20h, Univ. Paris-Saclay, Master of Mathematics / Data Science
 Master 2 : "Data Camp", 18h, Univ. Paris-Saclay, Master of Mathematics / Data Science
 Ecole d'hiver, Computational Brain Connectivity Mapping, Juan-Les-Pins, 1h30

Scikit-learn tutorial, Scipy Conf., Austin, USA, 6h

MNE tutorial, Univ. Libre de Bruxelles, Belgique, 6h

10.2.1.4. Denis A. Engemann

MNE-Python workshop, University of Bari, June 2017, 14h

10.2.1.5. Bertrand Thirion

Master : Functional Neuroimaging and brain-computer interfaces, 12h, MVA, ENS Paris-Saclay, France

Doctorat : OHBM course 2017, Pattern Recognition for brain Imaging, Vancouver, Canada

Master: Resting-State Functional Magnetic Resonance Imaging, 3h, Paris V, France.

10.2.2. Supervision

PhD in progress : Patricio Cerda, Statistical methods for analysis across datasets, started 01/09/2016, supervisors: Balazs Kegl (CNRS, LAL), Gaël Varoquaux

PhD in progress : Jérôme Dockès, Mining text and brain activity to learning the semantics of cognitive science, started 01/09/2016, supervisors: Fabian Suchanek (Télécom Paristech), Gaël Varoquaux

PhD in progress : Arthur Mensch, Learning representations of fMRI, started 01/09/2015, supervisors: Bertrand Thirion, Gaël Varoquaux, Julien Mairal (Inria Grenoble)

PhD: Elvis Dohmatob, Amélioration de connectivité fonctionnelle par utilisation de modèles déformables dans l'estimation de décompositions spatiales des images de cerveau, defended sept 2017, supervisors Bertrand Thirion, Gaël Varoquaux

10.2.2.1. Philippe Ciuciu

PhD in progress: Carole Lazarus (3rd year, Director)

PhD in progress: Loubna El Gueddari (2nd year, Director)

PhD in progress: Hamza Cherkaoui (1st year, Director)

PhD in progress: Sylvain Lannuzel (1st year, co-director)

10.2.2.2. Alexandre Gramfort

PhD defended: Albert Thomas

PhD defended: Romain Laby

PhD in progress: Yousra Bekhti

PhD in progress: Mainak Jas

PhD in progress: Tom Dupré La Tour

PhD in progress: Stanislas Chambon

PhD in progress: Pierre Ablin

PhD in progress: Mathurin Massias

PhD in progress: Hicham Janati

10.2.2.3. Denis A. Engemann

PhD in progress: Mainak Jas (co-supervision on 1 main project)

PhD in progress: Sami about (co-supervision on 1 main project)

PhD in progress: Kamalakar Reddy (co-supervision)

10.2.2.4. Bertrand Thirion

PhD defended: Elvis Dohmatob (co-supervision with G.Varoquaux)

PhD in progress: Arthur Mensch (co-supervision with G. Varoquaux and J. Mairal)

PhD in progress: Jérôme Dockès (co-supervision with G. Varoquaux and F. Suchanek)

PhD in progress: Kamalakar Reddy (co-supervision with G.Varoquaux and D. Engemann)

PhD in progress: Jérôme-Alexis Chevalier (co-supervision with J.Salmon)

10.2.3. *Juries*

10.2.3.1. *Philippe Ciuciu*

12/15: Reviewer of Lucie Thiebaut-Lonjaret PhD Thesis, Univ. Aix-Marseille, France.

04/23: Reviewer of Yosra Marnissi PhD Thesis, Univ. Paris-Est, France.

10.2.3.2. *Gaël Varoquaux*

external reviewer, PhD defence of Christian Dansereau, Université de Montréal.

10.2.3.3. *Alexandre Gramfort*

12/19: Examiner of Thomas Moreau, Univ. Paris Saclay / ENS Paris-Saclay

12/13: Reviewer of Etienne Combrisson PhD Thesis, Univ. Lyon 1, France.

11/10: Reviewer of Lankinen Kaisu PhD Thesis, Univ. Aalto, Finland.

04/26: Reviewer of Seyed Mostafa Kia PhD Thesis, Trento FBK Univ., Italy.

10.2.3.4. *Bertrand Thirion*

18/01: Reviewer of Jonathan Vacher PhD thesis, Univ. Paris VI

30/05: Examiner of Brahim Bougacha PhD thesis, University Nice

10.3. Popularization

10.3.1. *Gaël Varoquaux*

Présentation "Démocratisation de l'IA", Station F, (Dec 2017).

Présentation sur les logiciels libres de machine learning aux journées annuelles du groupe de travail sur les logiciels libres (Oct 2017)

Présentation sur scikit-learn au "Paris ML meetup" (Dec 2017)

Présentation sur scikit-learn au "Paris Pydata meetup" (Mars 2017)

10.3.2. *Alexandre Gramfort*

Présentation sur "L'intelligence artificielle au service de la société" à l'UNESCO dans le cadre des journées des ingénieurs et scientifiques de France (IESF) (Oct 2017). ³.

Présentation sur l'Intelligence Artificielle au conseil d'administration de l'Institut National du Cancer (INCA) (Nov 2017).

11. Bibliography

Major publications by the team in recent years

- [1] D. BZDOK, M. EICKENBERG, O. GRISEL, B. THIRION, G. VAROQUAUX. *Semi-Supervised Factored Logistic Regression for High-Dimensional Neuroimaging Data*, in "Neural Information Processing Systems", December 2015, <https://hal.archives-ouvertes.fr/hal-01211248>
- [2] N. CHAUFFERT, P. CIUCIU, J. KAHN, P. WEISS. *A Projection Method on Measures Sets*, in "Constructive Approximation", February 2017, vol. 45, n^o 1, pp. 83 - 111 [DOI : 10.1007/s00365-016-9346-2], <https://hal.inria.fr/hal-01432720>

³http://home.iesf.fr/752_p_49913/l-intelligence-artificielle.html

- [3] M. EICKENBERG, E. DOHMATOB, B. THIRION, G. VAROQUAUX. *Total Variation meets Sparsity: statistical learning with segmenting penalties*, in "Medical Image Computing and Computer Aided Intervention (MICCAI)", München, Germany, Proceedings of MICCAI 2015, October 2015, <https://hal.inria.fr/hal-01170619>
- [4] A. A. HOYOS-IDROBO, G. VAROQUAUX, Y. SCHWARTZ, B. THIRION. *FReM – scalable and stable decoding with fast regularized ensemble of models*, in "NeuroImage", 2017, pp. 1-16 [DOI : 10.1016/J.NEUROIMAGE.2017.10.005], <https://hal.archives-ouvertes.fr/hal-01615015>
- [5] M. JAS, D. A. ENGEMANN, Y. BEKHTI, F. A. RAIMONDO, A. GRAMFORT. *Autoreject: Automated artifact rejection for MEG and EEG data*, in "NeuroImage", 2017 [DOI : 10.1016/J.NEUROIMAGE.2017.06.030], <https://hal.inria.fr/hal-01562403>
- [6] A. MENSCH, J. MAIRAL, D. BZDOK, B. THIRION, G. VAROQUAUX. *Learning Neural Representations of Human Cognition across Many fMRI Studies*, in "Neural Information Processing Systems", Long Beach, United States, December 2017, <https://hal.archives-ouvertes.fr/hal-01626823>
- [7] 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 [DOI : 10.1109/TSP.2017.2752697], <https://hal.archives-ouvertes.fr/hal-01431618>
- [8] B. NG, G. VAROQUAUX, J.-B. POLINE, M. D. GREICIUS, B. THIRION. *Transport on Riemannian Manifold for Connectivity-based Brain Decoding*, in "IEEE Transactions on Medical Imaging", 2015, vol. PP, n^o 99, 9 p. [DOI : 10.1109/TMI.2015.2463723], <https://hal.inria.fr/hal-01185200>
- [9] F. PEDREGOSA, M. EICKENBERG, P. CIUCIU, B. THIRION, A. GRAMFORT. *Data-driven HRF estimation for encoding and decoding models*, in "NeuroImage", November 2014, pp. 209–220, <https://hal.inria.fr/hal-00952554>
- [10] Y. SCHWARTZ, B. THIRION, G. VAROQUAUX. *Mapping cognitive ontologies to and from the brain*, in "NIPS (Neural Information Processing Systems)", United States, November 2013, <http://hal.inria.fr/hal-00904763>

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [11] E. DOHMATOB. *Enhancement of functional brain connectome analysis by the use of deformable models in the estimation of spatial decompositions of the brain images*, Université Paris-Saclay, September 2017, <https://tel.archives-ouvertes.fr/tel-01630295>

Articles in International Peer-Reviewed Journals

- [12] M. ALBUGHDADI, L. CHAARI, J.-Y. TOURNERET, F. FORBES, P. CIUCIU. *A Bayesian Non-Parametric Hidden Markov Random Model for Hemodynamic Brain Parcellation*, in "Signal Processing", June 2017, vol. 135, pp. 132–146 [DOI : 10.1016/J.SIGPRO.2017.01.005], <https://hal.archives-ouvertes.fr/hal-01426385>
- [13] D. T. ALCALÁ-LÓPEZ, J. SMALLWOOD, E. JEFFERIES, F. B. VAN OVERWALLE, K. VOGLEY, R. B. MARS, B. B. TURETSKY, A. R. LAIRD, P. T. FOX, S. B. EICKHOFF, D. BZDOK. *Computing the Social Brain Connectome Across Systems and States*, in "Cerebral Cortex", 2017, pp. 1 - 26 [DOI : 10.1093/CERCOR/BHX121], <https://hal.archives-ouvertes.fr/hal-01519450>

- [14] D. BZDOK. *Classical Statistics and Statistical Learning in Imaging Neuroscience: Two Statistical Cultures in Neuroimaging*, in "Frontiers in Human Neuroscience", November 2017, <https://hal.archives-ouvertes.fr/hal-01583175>
- [15] D. BZDOK, M. KRZYWINSKI, N. ALTMAN. *Machine learning: A primer*, in "Nature Methods", November 2017, <https://hal.archives-ouvertes.fr/hal-01598285>
- [16] 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>
- [17] D. BZDOK, A. MEYER-LINDENBERG. *Machine learning for precision psychiatry: Opportunities and challenges*, in "Biological Psychiatry: Cognitive Neuroscience and Neuroimaging", February 2018, <https://hal.archives-ouvertes.fr/hal-01643933>
- [18] D. BZDOK, B. T. T. YEO. *Inference in the age of big data: Future perspectives on neuroscience*, in "NeuroImage", April 2017 [DOI : 10.1016/J.NEUROIMAGE.2017.04.061], <https://hal.archives-ouvertes.fr/hal-01516891>
- [19] N. CHAUFFERT, P. CIUCIU, J. KAHN, P. WEISS. *A Projection Method on Measures Sets*, in "Constructive Approximation", February 2017, vol. 45, n^o 1, pp. 83 - 111 [DOI : 10.1007/s00365-016-9346-2], <https://hal.inria.fr/hal-01432720>
- [20] M. DUBOL, C. TRICHARD, A. LARISA SANDU, C. LEROY, M. RAHIM, J.-L. MARTINOT, E. ARTIGES. *286. Dopamine Transporter and Reward Anticipation in Psychiatric Patients: A Positron Emission Tomography and Functional Magnetic Resonance Imaging Study*, in "Biological Psychiatry", May 2017, vol. 81, n^o 10, pp. S117 - S118 [DOI : 10.1016/J.BIOPSYCH.2017.02.300], <https://hal.inria.fr/hal-01593040>
- [21] M. DUBOL, C. TRICHARD, C. LEROY, A.-L. SANDU, M. RAHIM, B. GRANGER, E. T. TZAVARA, L. KARILA, J.-L. MARTINOT, E. ARTIGES. *Dopamine Transporter and Reward Anticipation in a Dimensional Perspective: A Multimodal Brain Imaging Study*, in "Neuropsychopharmacology", August 2017 [DOI : 10.1038/NPP.2017.183], <https://hal.inria.fr/hal-01576551>
- [22] T. DUPRÉ LA TOUR, L. TALLOT, L. GRABOT, V. DOYÈRE, V. VAN WASSENHOVE, Y. GRENIER, A. GRAMFORT. *Non-linear auto-regressive models for cross-frequency coupling in neural time series*, in "PLoS Computational Biology", December 2017, vol. 13, n^o 12, e1005893 [DOI : 10.1371/JOURNAL.PCBI.1005893], <https://hal.archives-ouvertes.fr/hal-01679078>
- [23] A. A. HOYOS-IDROBO, G. VAROQUAUX, Y. SCHWARTZ, B. THIRION. *FReM – scalable and stable decoding with fast regularized ensemble of models*, in "NeuroImage", 2017, pp. 1-16 [DOI : 10.1016/J.NEUROIMAGE.2017.10.005], <https://hal.archives-ouvertes.fr/hal-01615015>
- [24] M. JAS, D. A. ENGEMANN, Y. BEKHTI, F. A. RAIMONDO, A. GRAMFORT. *Autoreject: Automated artifact rejection for MEG and EEG data*, in "NeuroImage", 2017, <https://arxiv.org/abs/1612.08194> [DOI : 10.1016/J.NEUROIMAGE.2017.06.030], <https://hal.inria.fr/hal-01562403>
- [25] Y. LE GUEN, G. AUZIAS, F. LEROY, M. NOULHIANE, G. DEHAENE-LAMBERTZ, E. DUCHESNAY, J.-F. MANGIN, O. COULON, V. FROUIN. *Genetic Influence on the Sulcal Pits: On the Origin of the First Cortical Folds*, in "Cerebral Cortex", April 2017, pp. 1 - 12 [DOI : 10.1093/CERCOR/BHX098], <https://hal.amu.archives-ouvertes.fr/hal-01527005>

- [26] 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>
- [27] G. LEMAITRE, F. NOGUEIRA, C. K. ARIDAS. *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*, in "Journal of Machine Learning Research", 2017, vol. 18, pp. 1-5, <https://hal.inria.fr/hal-01516244>
- [28] J. LOULA, G. VAROQUAUX, B. THIRION. *Decoding fMRI activity in the time domain improves classification performance*, in "NeuroImage", August 2017 [DOI : 10.1016/J.NEUROIMAGE.2017.08.018], <https://hal.inria.fr/hal-01576641>
- [29] 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>
- [30] B. NG, G. VAROQUAUX, J. BAPTISTE POLINE, B. THIRION, M. D. GREICIUS, K. L. POSTON. *Distinct alterations in Parkinson's Medication-state and Disease-state Connectivity Running Title: PD altered connectivity*, in "Neuroimage-Clinical", 2017, <https://hal.archives-ouvertes.fr/hal-01614971>
- [31] T. E. NICHOLS, S. DAS, S. B. EICKHOFF 3, A. C. EVANS, T. GLATARD, M. HANKE, N. KRIEGESKORTE, M. P. MILHAM, R. A. POLDRACK, J.-B. P. POLINE, E. P. PROAL, B. THIRION, D. VAN ESSEN, T. WHITE, B. T. YEO. *Best Practices in Data Analysis and Sharing in Neuroimaging using MRI*, in "Nature Neuroscience", February 2017, vol. 20, pp. 299-303 [DOI : 10.1038/NN.4500], <https://hal.inria.fr/hal-01577319>
- [32] M. RAHIM, B. THIRION, D. BZDOK, I. BUVAT, G. VAROQUAUX. *Joint prediction of multiple scores captures better individual traits from brain images*, in "NeuroImage", June 2017 [DOI : 10.1016/J.NEUROIMAGE.2017.06.072], <https://hal.inria.fr/hal-01547524>
- [33] G. VAROQUAUX. *Cross-validation failure: small sample sizes lead to large error bars*, in "NeuroImage", June 2017, <https://arxiv.org/abs/1706.07581> [DOI : 10.1016/J.NEUROIMAGE.2017.06.061], <https://hal.inria.fr/hal-01545002>
- [34] D. VATANSEVER, D. BZDOK, H.-T. WANG, G. MOLLO, M. SORMAZ, C. MURPHY, T. KARAPANAGIOTIDIS, J. SMALLWOOD, E. JEFFERIES. *Varieties of semantic cognition revealed through simultaneous decomposition of intrinsic brain connectivity and behaviour*, in "NeuroImage", July 2017, 34 p. [DOI : 10.1016/J.NEUROIMAGE.2017.06.067], <https://hal.archives-ouvertes.fr/hal-01546394>
- [35] H.-T. WANG, P. GUILIA, C. MURPHY, D. BZDOK, E. JEFFERIES, J. SMALLWOOD. *Dimensions of Experience: Exploring the Heterogeneity of the Wandering Mind*, in "Psychological Science", August 2017, <https://hal.archives-ouvertes.fr/hal-01562681>

International Conferences with Proceedings

- [36] J. ARIAS, P. CIUCIU, M. DOJAT, F. FORBES, A. FRAU-PASCUAL, T. PERRET, J. M. WARNKING. *PyHRF: A Python Library for the Analysis of fMRI Data Based on Local Estimation of the Hemodynamic Response*

- Function*, in "16th Python in Science Conference (SciPy 2017)", Austin, TX, United States, July 2017 [DOI : 10.25080/SHINMA-7F4C6E7-006], <https://hal.archives-ouvertes.fr/hal-01566457>
- [37] E. B. BELILOVSKY, K. KASTNER, G. VAROQUAUX, M. B. BLASCHKO. *Learning to Discover Sparse Graphical Models*, in "International Conference on Machine Learning", Sydney, Australia, August 2017, <https://arxiv.org/abs/1605.06359> , <https://hal.inria.fr/hal-01306491>
- [38] P. CIUCIU, H. WENDT, S. COMBEXELLE, P. ABRY. *Spatially regularized multifractal analysis for fMRI Data*, in "EMBC'17 - 39th International Conference of the IEEE Engineering in Medicine and Biology Society", Jeju, South Korea, Kwang Suk Park, July 2017, 4 p. , <https://hal.inria.fr/hal-01574216>
- [39] M. CLERC, A. GRAMFORT, E. OLIVI, T. PAPADOPOULOU. *OpenMEEG software for forward problems handling non-nested geometries*, in "BaCI Conference 2017: International Conference on Basic and Clinical Multimodal Imaging", Bern, Switzerland, August 2017, <https://hal.inria.fr/hal-01581710>
- [40] A. HOYOS-IDROBO, G. VAROQUAUX, B. THIRION. *Towards a Faster Randomized Parcellation Based Inference*, in "PRNI 2017 - 7th International Workshop on Pattern Recognition in NeuroImaging", Toronto, Canada, June 2017, <https://hal.inria.fr/hal-01552237>
- [41] M. JAS, T. DUPRÉ LA TOUR, U. ŞİMŞEKLI, A. GRAMFORT. *Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding*, in "Advances in neural information processing systems", Long Beach, United States, December 2017, <https://arxiv.org/abs/1705.08006> , <https://hal.archives-ouvertes.fr/hal-01590988>
- [42] C. LAZARUS, P. WEISS, N. CHAUFFERT, F. MAUCONDUIT, M. BOTTLAENDER, A. VIGNAUD, P. CIUCIU. *SPARKLING: Novel Non-Cartesian Sampling Schemes for Accelerated 2D Anatomical Imaging at 7T Using Compressed Sensing*, in "25th annual meeting of the International Society for Magnetic Resonance Imaging", Honolulu, United States, April 2017, <https://hal.inria.fr/hal-01577200>
- [43] C. LAZARUS, P. WEISS, N. CHAUFFERT, A. VIGNAUD, P. CIUCIU. *SPARKLING: nouveaux schémas d'échantillonnage compressif prospectif pour l'IRM haute résolution*, in "GRETSI", Juan les Pins, France, September 2017, <https://hal.inria.fr/hal-01577207>
- [44] G. LEMAITRE, R. MARTÍ, M. RASTGOO, F. MÉRIAUDEAU. *Computer-Aided Detection for Prostate Cancer Detection based on Multi-Parametric Magnetic Resonance Imaging*, in "EMBC 2017 : 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society", Jeju Island, South Korea, July 2017, <https://hal.inria.fr/hal-01516245>
- [45] A. MENSCH, J. MAIRAL, D. BZDOK, B. THIRION, G. VAROQUAUX. *Learning Neural Representations of Human Cognition across Many fMRI Studies*, in "Neural Information Processing Systems", Long Beach, United States, December 2017, <https://arxiv.org/abs/1710.11438> , <https://hal.archives-ouvertes.fr/hal-01626823>
- [46] M. RAHIM, B. THIRION, G. VAROQUAUX. *Multi-output predictions from neuroimaging: assessing reduced-rank linear models*, in "PRNI 2017 - The 7th International Workshop on Pattern Recognition in Neuroimaging", Toronto, Canada, June 2017, pp. 1 - 4 [DOI : 10.1109/PRNI.2017.7981504], <https://hal.inria.fr/hal-01547572>

- [47] M. RAHIM, B. THIRION, G. VAROQUAUX. *Population-shrinkage of covariance to estimate better brain functional connectivity*, in "Medical Image Computing and Computer Assisted Interventions", Quebec city, Canada, September 2017, <https://hal.inria.fr/hal-01547612>

Conferences without Proceedings

- [48] D. BZDOK, M. EICKENBERG, G. VAROQUAUX, B. THIRION. *Hierarchical Region-Network Sparsity for High-Dimensional Inference in Brain Imaging*, in "International conference on Information Processing in Medical Imaging (IPMI) 2017", Boone, North Carolina, USA, June 2017, <https://hal.archives-ouvertes.fr/hal-01480885>
- [49] C. LAROCHE, H. PAPADOPOULOS, M. KOWALSKI, G. RICHARD. *Drum extraction in single channel audio signals using multi-layer non negative matrix factor deconvolution*, in "ICASSP", Nouvelle Orleans, United States, March 2017, <https://hal.archives-ouvertes.fr/hal-01438851>
- [50] M.-A. SCHULZ, G. VAROQUAUX, A. GRAMFORT, B. THIRION, D. BZDOK. *Label scarcity in biomedicine: Data-rich latent factor discovery enhances phenotype prediction*, in "Neural Information Processing Systems, Machine Learning in Health Workshop", Long Beach, United States, December 2017, <https://hal.archives-ouvertes.fr/hal-01633096>

Other Publications

- [51] P. ABLIN, J.-F. CARDOSO, A. GRAMFORT. *Faster ICA under orthogonal constraint*, November 2017, working paper or preprint, <https://hal.inria.fr/hal-01651842>
- [52] P. ABLIN, J.-F. CARDOSO, A. GRAMFORT. *Faster independent component analysis by preconditioning with Hessian approximations*, June 2017, <https://arxiv.org/abs/1706.08171> - working paper or preprint, <https://hal.inria.fr/hal-01552340>
- [53] G. JULIA GUIOMAR NISO, K. J. GORGOLEWSKI, E. BOCK, T. BROOKS, G. FLANDIN, A. GRAMFORT, R. N. HENSON, M. JAS, V. LITVAK, J. MOREAU, R. OOSTENVELD, J.-M. SCHOFFELEN, F. TADEL, J. WEXLER, S. BAILLET. *MEG-BIDS: an extension to the Brain Imaging Data Structure for magnetoencephalography*, September 2017, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01591080>

References in notes

- [54] 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
- [55] 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