Activity Report 2014

Project-Team PARIETAL

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

RESEARCH CENTER
Saclay - Île-de-France

THEME
Computational Neuroscience and Medecine

Table of contents

Project-Team PARIETAL

Keywords: Medical Images, Image Processing, Biological Images, Brain Computer Interface, Machine Learning

Creation of the Project-Team: 2009 July 01.

1. Members

Research Scientists
- Bertrand Thirion [Team leader, Inria, HdR]
- Philippe Ciuciu [CEA, Researcher, HdR]
- Gaël Varoquaux [Inria, Researcher]

Faculty Member
- Matthieu Kowalski [Univ. Paris XI, from Sep 2014]

Engineers
- Loïc Estève [Inria, from Sep 2014]
- Olivier Grisel [Inria]
- Salma Torkhani [Inria]
- Virgile Fritsch [EDF]

PhD Students
- Alexandre Abraham [Inria]
- Nicolas Chauffert [Univ. Paris XI]
- Elvis Dohmatob [Inria]
- Michael Eickenberg [Univ. Paris XI]
- Aina Frau Pascual [Inria]
- Andrés Hoyos Idrobo [Inria]
- Fabian Pedregosa Izquierdo [Inria]
- Viviana Siless [Inria, until Aug 2014]
- Hao Xu [Ecole Polytechnique, until Feb 2014]

Post-Doctoral Fellows
- Mehdi Rahim [CEA, from Nov 2014]
- Torsten Ruest [Inria, from Mar 2014 until Nov 2014]
- Ronald Phlypo [Inria, until Aug 2014]

Visiting Scientist
- Danilo Bzdok [Inria, from Sept 2014]

Administrative Assistant
- Régine Bricquet [Inria]

Others
- Kyle Kastner [Inria, intern, from Jun 2014 until Aug 2014]
- Cristóbal Pizarro Venegas [intern, until Mar 2014]
- Alexandre Gramfort [Institut Telecom]
- Yannick Schwartz [Inria]

2. Overall Objectives

2.1. Overall Objectives

The Parietal team focuses on mathematical methods for statistical inference based on neuroimaging data, with a particular reliance on machine learning techniques and applications of human functional imaging. This general theme splits into four research axes: Mathematical methods for multi-modal brain atlases, Statistical methods for high-dimensional data, Modeling brain function through neuroimaging and Parallel reconstruction and high-resolution 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, the neuroimaging *inverse problem* consists in predicting individual information (behavior, phenotype) from neuroimaging data, while the *forward problem* consists in fitting neuroimaging data with high-dimensional (e.g. genetic) 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 reasonably well ?), and a regularization schemes that represents a prior on the expected solution of the problem. In particular some priors enforce some properties of the solutions, such as sparsity, smoothness or being piece-wise constant.
Let us detail the model used in the inverse problem: Let $\mathbf{X}$ be a neuroimaging dataset as an (n_{subj}, n_{voxels}) matrix, where n_{subj} and n_{voxels} are the number of subjects under study, and the image size respectively, $\mathbf{Y}$ an array of values that represent characteristics of interest in the observed population, written as (n_{subj}, n_f) matrix, where n_f is the number of characteristics that are tested, and β an array of shape (n_{voxels}, n_f) that represents a set of pattern-specific maps. In the first place, we may consider the columns $\mathbf{Y}_1, .., \mathbf{Y}_{n_f}$ of Y independently, yielding n_f problems to be solved in parallel:

$$\mathbf{Y}_i = \mathbf{X}\beta_i + \epsilon_i, \forall i \in \{1, .., n_f\},$$

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

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

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

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

with λ_1, λ_2, η_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 β 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.

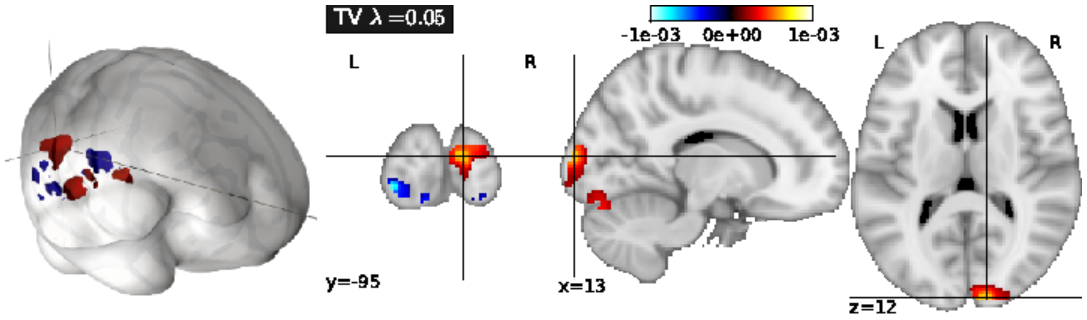

Figure 1. Example of the regularization of a brain map with total variation in an inverse problem. The problem here consists in predicting the spatial scale of an object presented as a stimulus, given functional neuroimaging data acquired during the observation of an image. Learning and test are performed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

The performance of the predictive model can simply be evaluated as the amount of variance in $\mathbf{Y}_i$ fitted by the model, for each $i \in \{1, .., n_f\}$. This can be computed through cross-validation, by *learning* $\widehat{\beta}_i$ on some part of the dataset, and then estimating $(Y_i - X\widehat{\beta}_i)$ 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 is particularly important to include external anatomical priors on the relevant 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 regression*, where a logistic 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* is an important question, as it makes little sense that a learned model depends dramatically on the particular observations used for learning. This is an important issue, as this kind of robustness is somewhat opposite to sparsity requirements.
- *Multi-task learning*: if several target variables are thought to be related, it might be useful to constrain the estimated parameter vector β 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 β_i and thus, improve the prediction of $\mathbf{Y}_i$. Yet this does not impose constraints on the non-zero parameters of the parameters β_i.

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

then

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

3.2. Multivariate decompositions

Multivariate decompositions are an important tool 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 depicting similar activities 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_{subj}, n_{voxels}) matrix, after proper centering; the model reads

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

where $\mathbf{D}$ represents a set of n_{comp} spatial maps, hence a matrix of shape (n_{comp}, n_{voxels}), and $\mathbf{A}$ the associated subject-wise loadings. While traditional PCA and independent components analysis are limited to reconstruct 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:

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

where $(\mathbf{A}_i),\ i \in \{1..n_f\}$ 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).

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 assess whether an observation is aberrant or not or in classification problems. 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 an important question that needs to be addressed.
- **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-samples 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.

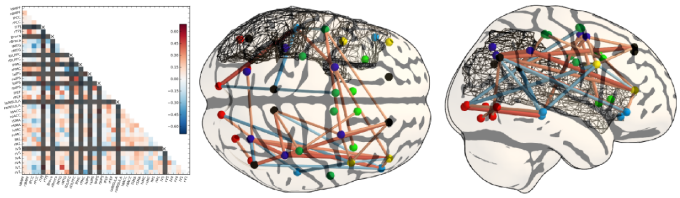

Figure 2. Example of functional connectivity analysis: The correlation matrix describing brain functional connectivity in a post-stroke patient (lesion volume outlined as a mesh) is compared to a group of control subjects. Some edges of the graphical model show a significant difference, but the statistical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

4. Application Domains

4.1. Human neuroimaging data and their use

Human neuroimaging consists in acquiring non-invasively image data from normal and diseased human populations. Magnetic Resonance Imaging (MRI) can be used to acquire information on brain structure and function at high spatial resolution.

- T1-weighted MRI is used to obtain a segmentation of the brain into different different tissues, such as gray matter, white matter, deep nuclei, cerebro-spinal fluid, at the millimeter or sub-millimeter resolution. This can then be used to derive geometric and anatomical information on the brain, e.g. cortical thickness.
- Diffusion-weighted MRI measures the local diffusion of water molecules in the brain at the resolution of 1 to 2mm, in a set of directions (60 typically). Local anisotropy, observed in white matter, yields a local model of fiber orientation that can be integrated into a geometric model of fiber tracts along which water diffusion occurs, and thus provides information on the connectivity structure of the brain.
- Functional MRI measures the blood-oxygen-level-dependent (BOLD) contrast that reflects neural activity in the brain, at a spatial resolution of 1.5 to 3mm, and a temporal resolution of about 2s. This yields a spatially resolved image of brain functional networks that can be modulated either by specific cognitive tasks or exhibit spontaneous co-activations.
- Electro- and Magneto-encephalography (MEEG) are two additional modalities that complement functional MRI, as they directly measure the electric and magnetic signals elicited by neural activity, at the millisecond scale. These modalities rely on surface measurements and do not localize brain activity very accurately in the spatial domain.

4.2. High-field MRI

High field MRI as performed at NeuroSpin (7T on humans, 11.7T in 2017, 17.6T on rats) brings an improvement over traditional MRI acquisitions at 1.5T or 3T, related to to a higher signal-to-noise ratio in the data. Depending on the data and applicative context, this gain in SNR can be traded against spatial resolution improvements, thus helping in getting more detailed views of brain structure and function. This comes at the risk of higher susceptibility distortions of the MRI scans and signal inhomogeneities, that need to be corrected for. Improvements at the acquisition level may come from the use of new coils (such as the 32 channels coil on the 7T at NeuroSpin), as well as the use of multi-band sequences [44].

4.3. Technical challenges for the analysis of neuroimaging data

The first limitation of Neuroimaging-based brain analysis is the limited Signal-to-Noise Ratio of the data. A particularly striking case if functional MRI, where only a fraction of the data is actually understood, and from which it is impossible to observe by eye the effect of neural activation on the raw data. Moreover, far from traditional i.i.d. Gaussian models, the noise in MRI typically exhibits local and long-distance correlations (e.g. motion-related signal) and has potentially large amplitude, which can make it hard to distinguish from true signal on a purely statistical basis. A related difficulty is the *lack of salient structure* in the data: it is hard to infer meaningful patterns (either through segmentation or factorization procedures) based on the data only. A typical case is the inference of brain networks from resting-state functional connectivity data.

Regarding statistical methodology, neuroimaging problems also suffer from the relative paucity of the data, i.e. the relatively small number of images available to learn brain features or models, e.g. with respect to the size of the images or the number of potential structures of interest. This leads to several kinds of difficulties, known either as *multiple comparison problems* or *curse of dimensionality*. One possibility to overcome this challenge is to increase the amount of data by using images from multiple acquisition centers, at the risk of introducing scanner-related variability, thus challenging the homogeneity of the data. This becomes an important concern with the advent of cross-modal neuroimaging-genetics studies.

5. New Software and Platforms

5.1. Scikit learn

Participants: Olivier Grisel [correspondant], Gaël Varoquaux, Bertrand Thirion, Michael Eickenberg, Loïc Estève, Alexandre Gramfort, Fabian Pedregosa Izquierdo.

Scikit-learn is an open-source machine learning toolkit written in Python/C that provides generic tools to learn information for the classification of various kinds of data, such as images or texts. It is tightly associated to the scientific Python software suite (Numpy/Scipy) for which it aims at providing a complementary toolkit for machine learning (classification, clustering, dimension reduction, regression). There is an important focus on code quality (API consistency, code readability, tests, documentation and examples), and on efficiency, as the scikit-learn compares favorably to state-of-the-art modules developed in R in terms of computation time or memory requirements. Scikit-learn is currently developed by more than 60 contributors, but the core developer team has been with the Parietal Inria team at Saclay-Île-de-France since January 2010. The scikit-learn has recently become the reference machine learning library in Python.

- Version: 0.15.2
- Programming language: Python, C/Cython

5.2. Nilearn

Participants: Gaël Varoquaux [correspondant], Bertrand Thirion, Loïc Estève, Alexandre Abraham, Michael Eickenberg, Alexandre Gramfort, Fabian Pedregosa Izquierdo, Elvis Dohmatob, Virgile Fritsch.

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. NiLearn is maintained both through the help of Inria: a developer funded by Saclay CRI in 2012-2013, a 2013-2014 ADT and through the NiConnect project.

- Version: 0.1
- Programming language: Python

5.3. Mayavi

Participant: Gaël Varoquaux [Correspondant].

Mayavi is the most used scientific 3D visualization Python software (http://mayavi.sourceforge.net/). It has been developed by Prabhu Ramachandran (IIT Bombay) and Gaël Varoquaux (PARIETAL, Inria Saclay). 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 (http://pyrx.scripps.edu) and brain connectivity analysis tools (connectomeViewer).

See also the web page http://mayavi.sourceforge.net/ and the following paper http://hal.inria.fr/inria-00528985/en.

- Version: 3.4.0

5.4. Nipy

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

Nipy is an open-source Python library for neuroimaging data analysis, developed mainly at Berkeley, Stanford, MIT and Neurospin. It is open to any contributors and aims at developing code and tools sharing. Some parts of the library are completely developed by Parietal. It is devoted to algorithmic solutions for various issues in neuroimaging data analysis. The Nipy project is available, under BSD license, and within NeuroDebian.

See also the web page http://nipy.org.

- Version: 0.3

5.5. PyHRF

Participants: Philippe Ciuciu [correspondant], Aina Frau Pascual, Salma Torkhani.

PyHRF is a set of tools for within-subject fMRI data analysis, focused on the characterization of the hemodynamics. Within the chain of fMRI data processing, these tools provide alternatives to the classical within-subject GLM estimation step. The inputs are preprocessed within-subject data and the outputs are statistical maps and/or fitted HRFs. The package is mainly written in Python and provides the implementation of the two following methods:

- The joint-detection estimation (JDE) approach, that divides the brain into functionally homogeneous regions and provides one HRF estimate per region as well as response levels specific to each voxel and each experimental condition. This method embeds a temporal regularization on the estimated HRFs and an adaptive spatial regularization on the response levels.
- The Regularized Finite Impulse Response (RFIR) approach, that provides HRF estimates for each voxel and experimental conditions. This method embeds a temporal regularization on the HRF shapes, but proceeds independently across voxels (no spatial model).

The development of PyHRF is now funded by an Inria ADT, in collaboration with MISTIS.

- Version: 0.1
- Keywords: Hemodynamic response function; estimation; detection; fMRI
- License: BSD 4
- Multiplatform: Windows - Linux - MacOSX
- Programming language: Python

6. New Results

6.1. Highlights of the Year

- Congratulations also to Alex and Daniel Strohmeier for their best paper award at the PRNI 2014 conference: “Improved MEG/EEG source localization with reweighted mixed-norms“.
- Elvis Dohmatob got a honorable mention for the student paper award at PRNI 2014 for the work “Benchmarking solvers for TV-l1 least-squares and logistic regression in brain imaging”

6.2. Which fMRI clustering gives good brain parcellations?

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

Analysis and interpretation of neuroimaging data often require one to divide the brain into a number of regions, or parcels, with homogeneous characteristics, be these regions defined in the brain volume or on on the cortical surface. While predefined brain atlases do not adapt to the signal in the individual subjects images, parcellation approaches use brain activity (e.g. found in some functional contrasts of interest) and clustering techniques to define regions with some degree of signal homogeneity. In this work, we address the question of which clustering technique is appropriate and how to optimize the corresponding model. We use two principled criteria: goodness of fit (accuracy), and reproducibility of the parcellation across bootstrap samples. We study these criteria on both simulated and two task-based functional Magnetic Resonance Imaging datasets for the Ward, spectral and K-means clustering algorithms. We show that in general Ward’s clustering performs better than alternative methods with regard to reproducibility and accuracy and that the two criteria diverge regarding the preferred models (reproducibility leading to more conservative solutions), thus deferring the practical decision to a higher level alternative, namely the choice of a trade-off between accuracy and stability.

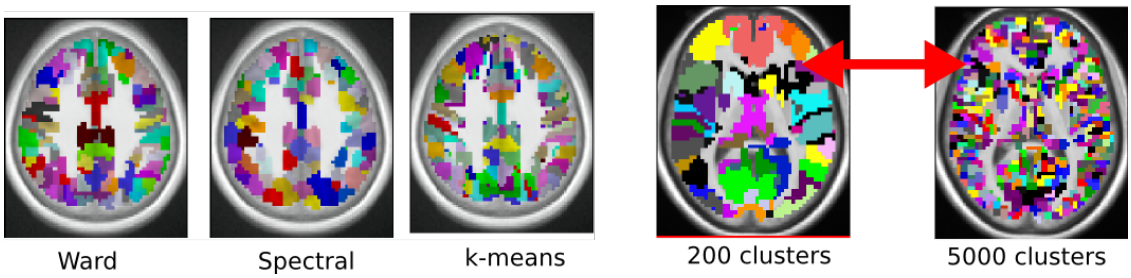

Figure 3. Practitioner have to decide which clustering method to use and how to select the number of clusters. In [21], we provide empirical guidelines and criteria to guide that choice in the context of functional brain imaging.

More details can be found in [21].

6.3. Principal Component Regression predicts functional responses across individuals

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Olivier Grisel.

Inter-subject variability is a major hurdle for neuroimaging group-level inference, as it creates complex image patterns that are not captured by standard analysis models and jeopardizes the sensitivity of statistical procedures. A solution to this problem is to model random subjects effects by using the redundant information conveyed by multiple imaging contrasts. In this paper, we introduce a novel analysis framework, where we estimate the amount of variance that is fit by a random effects subspace learned on other images; we show that a principal component regression estimator outperforms other regression models and that it fits a significant proportion (10% to 25%) of the between-subject variability. This proves for the first time that the accumulation of contrasts in each individual can provide the basis for more sensitive neuroimaging group analyzes.

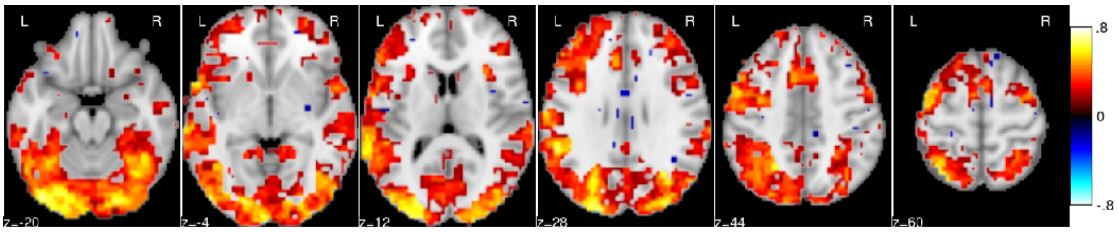

Figure 4. In most brain regions, knowing the amount of activation related to a set of reference contrasts yields an accurate prediction of the activation for a target contrast.

More details can be found in [36].

6.4. Deriving a multi-subject functional-connectivity atlas to inform connectome estimation

Participants: Ronald Phlypo [Correspondant], Bertrand Thirion, Gaël Varoquaux.

The estimation of functional connectivity structure from functional neuroimaging data is an important step toward understanding the mechanisms of various brain diseases and building relevant biomarkers. Yet, such inferences have to deal with the low signal-to-noise ratio and the paucity of the data. With at our disposal a steadily growing volume of publicly available neuroimaging data, it is however possible to improve the estimation procedures involved in connectome mapping. In this work, we propose a novel learning scheme for functional connectivity based on sparse Gaussian graphical models that aims at minimizing the bias induced by the regularization used in the estimation, by carefully separating the estimation of the model support from the coefficients. Moreover, our strategy makes it possible to include new data with a limited computational cost. We illustrate the physiological relevance of the learned prior, that can be identified as a functional connectivity atlas, based on an experiment on 46 subjects of the Human Connectome Dataset.

More details can be found in [35].

6.5. Machine Learning Patterns for Neuroimaging-Genetic Studies in the Cloud

Participants: Virgile Fritsch, Bertrand Thirion, Gaël Varoquaux.

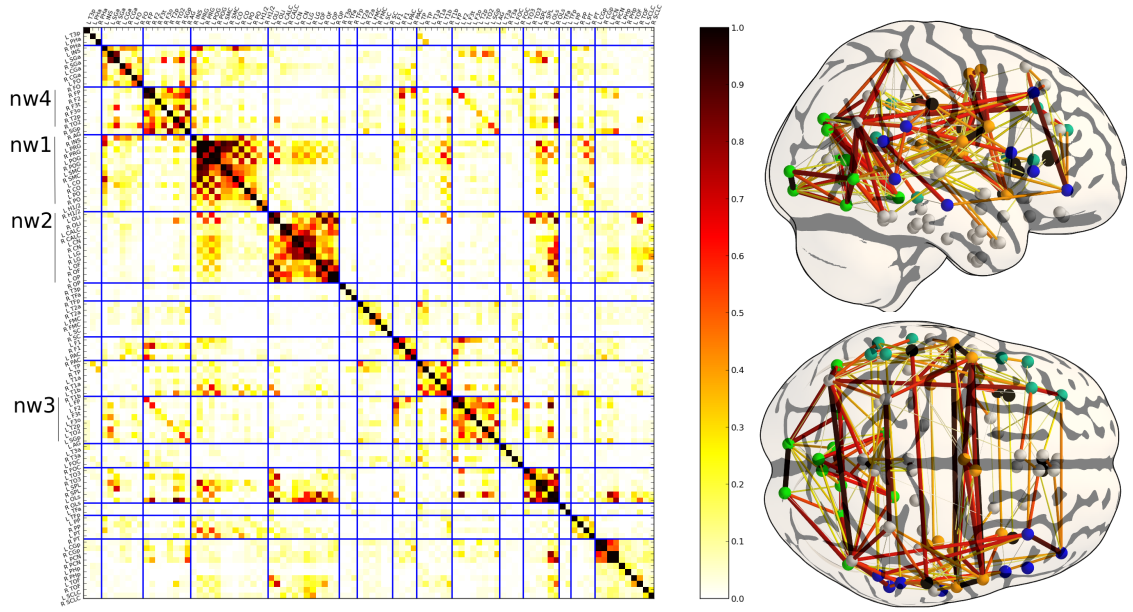

Figure 5. Prior on the functional connectivity: the coefficient of the matrix represent the frequency of an edge at each position. This model can be interpreted as a data-driven atlas of brain functional connections. In the current framework, it can easily be updated to take into account more data.

Brain imaging is a natural intermediate phenotype to understand the link between genetic information and behavior or brain pathologies risk factors. Massive efforts have been made in the last few years to acquire high-dimensional neuroimaging and genetic data on large cohorts of subjects. The statistical analysis of such data is carried out with increasingly sophisticated techniques and represents a great computational challenge. Fortunately, increasing computational power in distributed architectures can be harnessed, if new neuroinformatics infrastructures are designed and training to use these new tools is provided. Combining a MapReduce framework (TomusBLOB) with machine learning algorithms (Scikit-learn library), we design a scalable analysis tool that can deal with non-parametric statistics on high-dimensional data. End-users describe the statistical procedure to perform and can then test the model on their own computers before running the very same code in the cloud at a larger scale. We illustrate the potential of our approach on real data with an experiment showing how the functional signal in subcortical brain regions can be significantly fit with genome-wide genotypes. This experiment demonstrates the scalability and the reliability of our framework in the cloud with a two weeks deployment on hundreds of virtual machines.

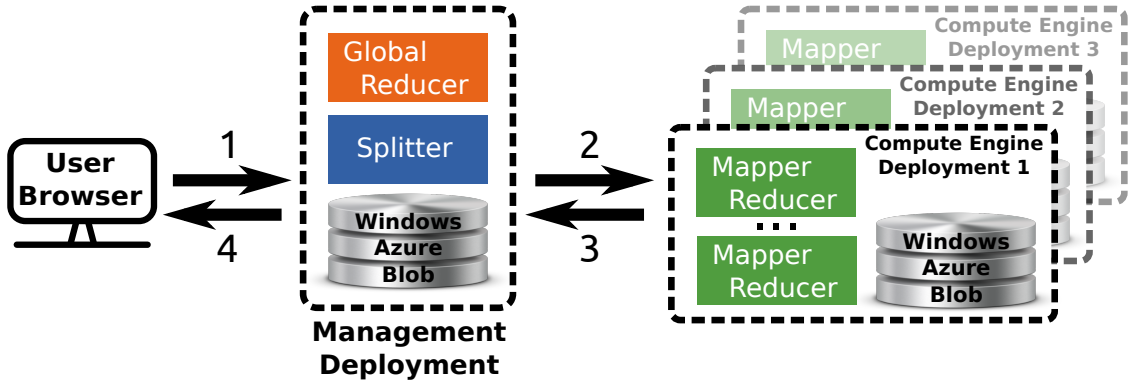

Figure 6. Overview of the multi site deployment of a hierarchical Tomus-MapReduce compute engine. 1) The end-user uploads the data and configures the statistical inference procedure on a webpage. 2) The Splitter partitions the data and manages the workload. The compute engines retrieves job information trough the Windows Azure Queues. 3) Compute engines perform the map and reduce jobs. The management deployment is informed of the progression via the Windows Azure Queues system and thus can manage the execution of the global reducer. 4) The user downloads the results of the computation on the webpage of the experiment.

More details can be found in [17].

6.6. Data-driven HRF estimation for encoding and decoding models

Participants: Fabian Pedregosa Izquierdo [correspondant], Michael Eickenberg, Alexandre Gramfort, Philippe Ciuciu, Bertrand Thirion, Gaël Varoquaux.

Despite the common usage of a canonical, data-independent, hemodynamic response function (HRF), it is known that the shape of the HRF varies across brain regions and subjects. This suggests that a data-driven estimation of this function could lead to more statistical power when modeling BOLD fMRI data. However, unconstrained estimation of the HRF can yield highly unstable results when the number of free parameters is large. We develop a method for the joint estimation of activation and HRF using a rank constraint causing the estimated HRF to be equal across events/conditions, yet permitting it to be different across voxels. Model estimation leads to an optimization problem that we propose to solve with an efficient quasi-Newton method exploiting fast gradient computations. This model, called GLM with Rank-1 constraint (R1-GLM), can be extended to the setting of GLM with separate designs which has been shown to improve decoding

accuracy in brain activity decoding experiments. We compare 10 different HRF modeling methods in terms of encoding and decoding score in two different datasets. Our results show that the R1-GLM model significantly outperforms competing methods in both encoding and decoding settings, positioning it as an attractive method both from the points of view of accuracy and computational efficiency.

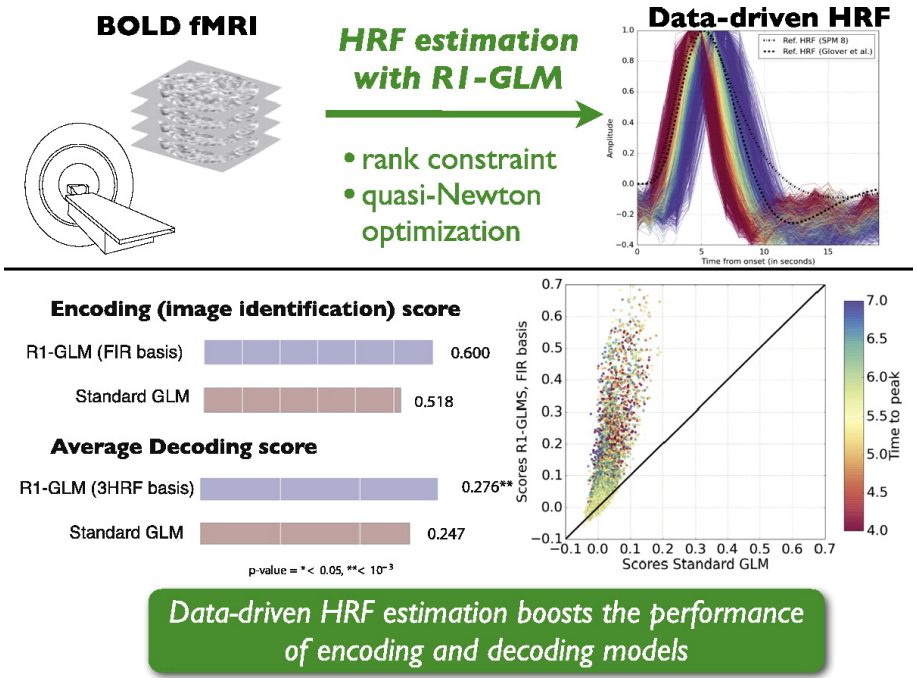

Figure 7. Illustration of the hemodynamic response function estimation framework introduced in [19].

More details can be found in [19].

6.7. Benchmarking solvers for TV-l1 least-squares and logistic regression in brain imaging

Participants: Elvis Dohmatob [correspondant], Michael Eickenberg, Gaël Varoquaux, Bertrand Thirion.

Learning predictive models from brain imaging data, as in decoding cognitive states from fMRI (functional Magnetic Resonance Imaging), is typically an ill-posed problem as it entails estimating many more parameters than available sample points. This estimation problem thus requires regularization. Total variation regularization, combined with sparse models, has been shown to yield good predictive performance, as well as stable and interpretable maps. However, the corresponding optimization problem is very challenging: it is non-smooth, non-separable and heavily ill-conditioned. For the penalty to fully exercise its structuring effect on the maps, this optimization problem must be solved to a good tolerance, resulting in a computational challenge. In this work, we explore a wide variety of solvers and exhibit their convergence properties on fMRI data. We introduce a variant of smooth solvers and show that it is a promising approach in these settings. Our findings show that care must be taken in solving TV-l1 estimation in brain imaging and highlight the successful strategies.

More details can be found in [30]

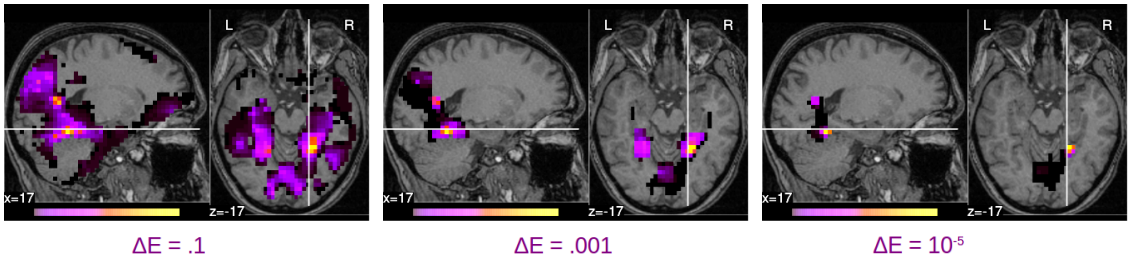

Figure 8. TV-l1 maps for a face-house discrimination task taken from a visual recognition dataset, with regularization parameters chosen by cross-validation, for different stopping criteria. Note that the stopping criterion is defined as a threshold on the energy decrease per one iteration of the algorithm. This figure shows the importance of convergence of the multivariate estimator, and motivates the need for a fast solver.

6.8. Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks

Participant: Philippe Ciuciu [correspondant].

Studies employing functional connectivity-type analyses have established that spontaneous fluctuations in functional magnetic resonance imaging (fMRI) signals are organized within large-scale brain networks. Meanwhile, fMRI signals have been shown to exhibit 1/f-type power spectra – a hallmark of scale-free dynamics. We studied the interplay between functional connectivity and scale-free dynamics in fMRI signals, utilizing the fractal connectivity framework – a multivariate extension of the univariate fractional Gaussian noise model, which relies on a wavelet formulation for robust parameter estimation. We applied this framework to fMRI data acquired from healthy young adults at rest and performing a visual detection task. First, we found that scale-invariance existed beyond univariate dynamics, being present also in bivariate cross-temporal dynamics. Second, we observed that frequencies within the scale-free range do not contribute evenly to inter-regional connectivity, with a systematically stronger contribution of the lowest frequencies, both at rest and during task. Third, in addition to a decrease of the Hurst exponent and inter-regional correlations, task performance modified cross-temporal dynamics, inducing a larger contribution of the highest frequencies within the scale-free range to global correlation.

More details can be found in [16].

6.9. Supramodal processing optimizes visual perceptual learning and plasticity

Participants: Philippe Ciuciu [correspondant], Alexandre Gramfort.

Multisensory interactions are ubiquitous in cortex and it has been suggested that sensory cortices may be supramodal i.e. capable of functional selectivity irrespective of the sensory modality of inputs. Here, we asked whether learning to discriminate visual coherence could benefit from supramodal processing. To this end, three groups of participants were briefly trained to discriminate which of a red or green intermixed population of random-dot-kinematograms (RDKs) was most coherent in a visual display while being recorded with magnetoencephalography (MEG). During training, participants heard no sound (V), congruent acoustic textures (AV) or auditory noise (AVn); importantly, congruent acoustic textures shared the temporal statistics – i.e. coherence – of visual RDKs. After training, the AV group significantly outperformed participants trained in V and AVn although they were not aware of their progress. In pre- and post-training blocks, all participants were tested without sound and with the same set of RDKs. When contrasting MEG data collected in these experimental blocks, selective differences were observed in the dynamic pattern and the cortical loci

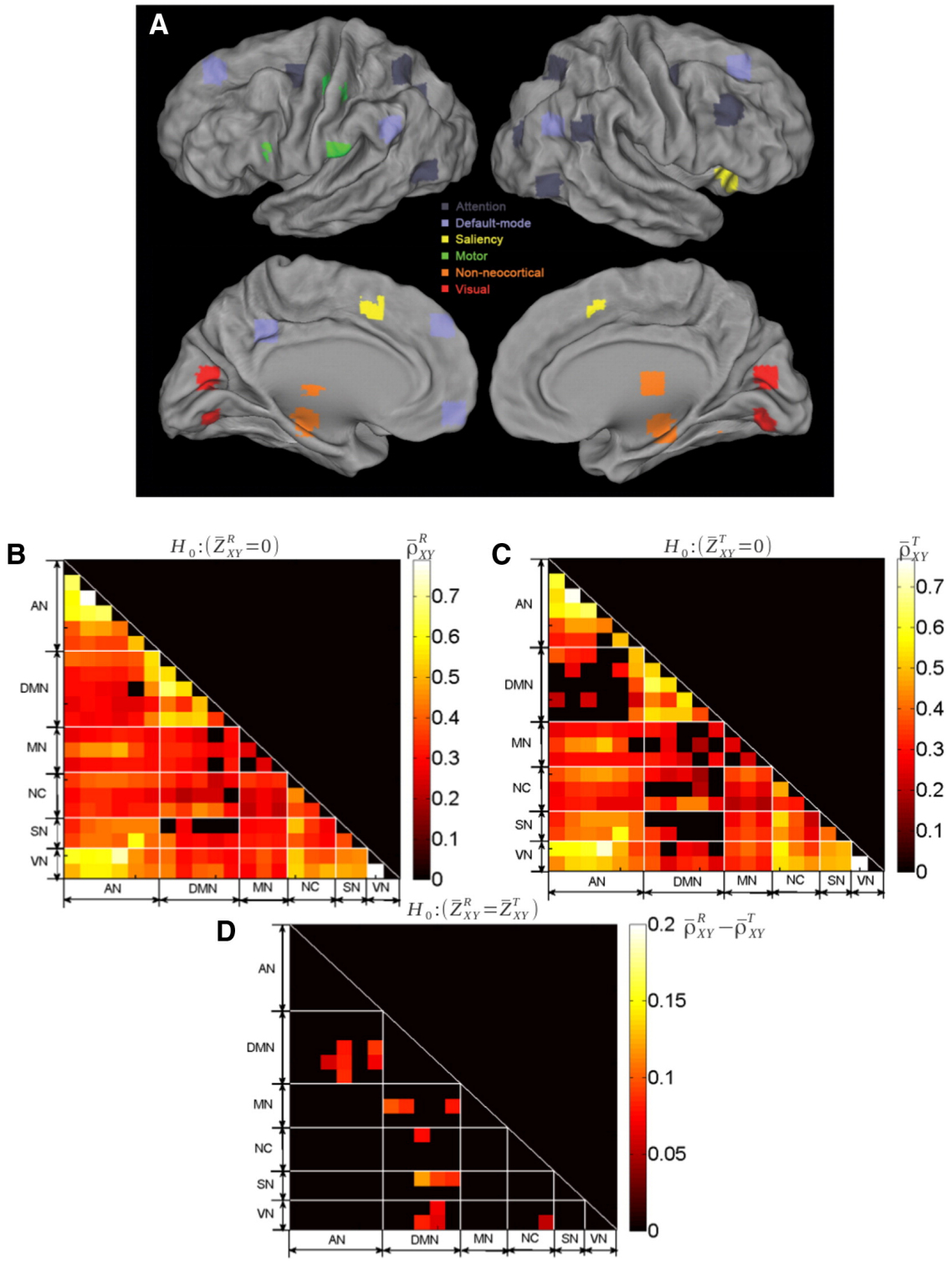

Figure 9. Networks definition and correlation structure. Top (A): ROIs mapped onto the cortical surface, with each color denoting a different network. Middle (B): Group-averaged inter-regional correlation matrix at rest ($p < 0.05$, Bonferroni corrected). Regions are grouped by network to ease visualization. Middle (C): Group-averaged inter-regional correlation matrix during the visual detection task ($p < 0.05$, Bonferroni corrected). Bottom (D): Difference of the correlation coefficients between rest and task (thresholded at $p < 0.01$, uncorrected, two-sample t-test for rest vs. task). The ROIs are grouped by networks; these networks correspond to the diagonal triangles surrounded by white dashed lines.

responsive to visual RDKs. First and common to all three groups, vlPFC showed selectivity to the learned coherence levels whereas selectivity in visual motion area hMT+ was only seen for the AV group. Second and solely for the AV group, activity in multisensory cortices (mSTS, pSTS) correlated with post-training performances; additionally, the latencies of these effects suggested feedback from vlPFC to hMT+ possibly mediated by temporal cortices in AV and AVn groups. Altogether, we interpret our results in the context of the Reverse Hierarchy Theory of learning in which supramodal processing optimizes visual perceptual learning by capitalizing on sensory-invariant representations - here, global coherence levels across sensory modalities.

More details can be found in [25].

6.10. Variable density sampling with continuous trajectories. Application to MRI.

Participants: Nicolas Chauffert, Philippe Ciuciu [correspondant].

Reducing acquisition time is a crucial challenge for many imaging techniques. Compressed Sensing (CS) theory offers an appealing framework to address this issue since it provides theoretical guarantees on the reconstruction of sparse signals by projection on a low dimensional linear subspace. In this paper, we focus on a setting where the imaging device allows to sense a fixed set of measurements. We first discuss the choice of an optimal sampling subspace (smallest subset) allowing perfect reconstruction of sparse signals. Its standard design relies on the random drawing of independent measurements. We discuss how to select the drawing distribution and show that a mixed strategy involving partial deterministic sampling and independent drawings can help breaking the so-called "coherence barrier". Unfortunately, independent random sampling is irrelevant for many acquisition devices owing to acquisition constraints. To overcome this limitation, the notion of Variable Density Samplers (VDS) is introduced and defined as a stochastic process with a prescribed limit empirical measure. It encompasses samplers based on independent measurements or continuous curves. The latter are crucial to extend CS results to actual applications. Our main contribution lies in two original continuous VDS. The first one relies on random walks over the acquisition space whereas the second one is heuristically driven and rests on the approximate solution of a Traveling Salesman Problem. Theoretical analysis and retrospective CS simulations in magnetic resonance imaging highlight that the TSP-based solution provides improved reconstructed images in terms of signal-to-noise ratio compared to standard sampling schemes (spiral, radial, 3D iid...).

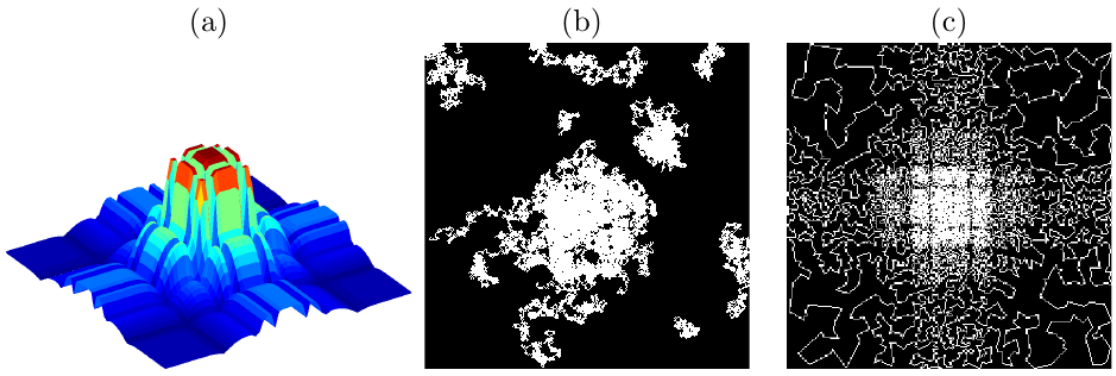

Figure 10. (a): Target distribution π to be approximated. Continuous random trajectories reaching distribution π based on Markov chains (b) and on a TSP solution (c). The latter is much more accurate.

More details can be found in [15].

7. Bilateral Contracts and Grants with Industry

7.1. The LearnClues Labcomm

The LearnClues LabComm has been granted on Oct 2.

Statistical learning is a field of mathematics and computer science that enables the extraction of predictive models from data with weak signal to noise ratio. These techniques are behind the successes of Google or the progresses of automatic medical diagnostic. Combined with a knowledge of the field of application, they open the door to optimal decisions. Tinyclues is a start-up that applies statistical learning to e-commerce, adapting the marketing practice from customer databases. Parietal is an Inria research group that develops statistical learning for neurosciences and is the driving force behind the software tool "scikit-learn", that is a standard in statistical learning.

The goal of this proposed common lab is to transfer the expertise of Parietal on big data and to improve statistical learning techniques and implementation on distributed systems to open the door to faster analysis of very large datasets. Indeed, processing more data implies detecting smaller effects in the signals. Tinyclues already uses the tools developed par Parietal on the "cloud", and thus in distributed computing environments. The practical experience of Parietal enables us to plan substantial improvements to computational performance as well as to the amount of information extracted from big data.

From a strategical standpoint for Tinyclues, such progress are important to vary the number of domain scenarios that it can address, by analyzing jointly more data of a wider type, and to render fully automatic the data analysis platform that it is offering to its customers, replacing challenging tasks currently performed by experts. These developments are particularly important given that Tinyclues is developing at a very fast rate and is processing bigger and bigger datasets and an increasing number of different problems.

The project partners are:

- Parietal, Inria
- Tiny Clues

7.2. 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. Gael 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 www.wendelin.io.

The project partners are:

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

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. 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.

8.1.2. SUBSAMPLE Digiteo chair

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

Parietal is associated with this Digiteo Chair by Dimitris Samaras, in which we will address the probabilistic structure learning of salient brain states (PhD of Alexandre Abraham, 2012-2015).

Cognitive tasks systematically involve several brain regions, and exploratory approaches are generally necessary given the lack of knowledge of the complex mechanisms that are observed. The goal of the project is to understand the neurobiological mechanisms that are involved in complex neuro-psychological disorders. A crucial and poorly understood component in this regard refers to the interaction patterns between different regions in the brain. In this project we will develop machine learning methods to capture and study complex functional network characteristics. We hypothesize that these characteristics not only offer insights into brain function but also can be used as concise features that can be used instead of the full dataset for tasks like classification of healthy versus diseased populations or for clustering subjects that might exhibit similarities in brain function. In general, the amount of correlation between distant brain regions may be a more reliable feature than the region-based signals to discriminate between two populations e.g. in schizophrenia. For such exploratory methods to be successful, close interaction with neuroscientists is necessary, as the salience of the features depends on the population and the observed effects of psychopathology. For this aim we propose to develop a number of important methodological advances in the context of prediction of treatment outcomes for drug addicted populations, e.g. for relapse prediction.

8.1.3. Medilearn/braincodes Inria-MSR project

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Andrés Hoyos Idrobo.

Neuroimaging is accumulating large functional MRI datasets that display –among artefacts and noise– brain activation patterns giving access to a meaningful representation of brain spatial organization. This ongoing accumulation is intensified via new large-scale international initiatives such as the *Human Connectome Project* (www.humanconnectomeproject.org). but also to existing open repositories of functional neuroimaging datasets (https://openfmri.org/) or http://www.fmridc.org/. These datasets represent a very significant resource for the community, but require new analytic approaches in order to be fully exploited. The MediLearn/BrainCodes project strives to provide a synthetic picture of the brain substrate of human cognition and its pathologies. In practice, this can be achieved by learning from large-scale datasets a brain atlas that summarizes adequately these functional activation maps drawing from a large number of protocols and subjects. Once learned, such an atlas is extremely useful to understand the large-scale functional organization of the brain: it is a tool for understanding *brain segregation*, the different encoding of many cognitive parameters into different brain regions, as well as *brain integration*, *i.e.* how remote brain regions co-activate across subjects and experiments.

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. BrainPedia project

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Yannick Schwartz, Virgile Fritsch.

BrainPedia is an ANR JCJC (2011-2015) which addresses the following question: Neuroimaging produces huge amounts of complex data that are used to better understand the relations between brain structure and function. While the acquisition and analysis of this data is getting standardized in some aspects, the neuroimaging community is still largely missing appropriate tools to store and organize the knowledge related to the data. Taking advantage of common coordinate systems to represent the results of group studies, coordinate-based meta-analysis approaches associated with repositories of neuroimaging publications provide a crude solution to this problem, that does not yield reliable outputs and looses most of the data-related information. In this project, we propose to tackle the problem in a statistically rigorous framework, thus providing usable information to drive neuroscientific knowledge and questions.

8.2.1.2. IRMgroup project

Participants: Bertrand Thirion [Correspondant], Alexandre Gramfort, Michael Eickenberg.

This is a joint project with Polytechnique/CMAP http://www.cmap.polytechnique.fr/: Stéphanie Allassonnière and Stéphane Mallat (2010-2014).

Much of the visual cortex is organized into visual field maps, which means that nearby neurons have receptive fields at nearby locations in the image. The introduction of functional magnetic resonance imaging (fMRI) has made it possible to identify visual field maps in human cortex, the most important one being the medial occipital cortex (V1,V2,V3). It is also possible to relate directly the activity of simple cells to an fMRI activation pattern and Parietal developed some of the most effective methods. However, the simple cell model is not sufficient to account for high-level information on visual scenes, which requires the introduction of specific semantic features. While the brain regions related to semantic information processing are now well understood, little is known on the flow of visual information processing between the primary visual cortex and the specialized regions in the infero-temporal cortex. A central issue is to better understand the behavior of intermediate cortex layers.

Our proposition is to use our mathematical approach to formulate explicitly some generative model of information processing, such as those that characterize complex cells in the visual cortex, and then to identify the brain substrate of the corresponding processing units from fMRI data. While fMRI resolution is still too coarse for a very detailed mapping of detailed cortical functional organization, we conjecture that some of the functional mechanisms that characterize biological vision processes can be captured through fMRI; in parallel we will push the fMRI resolution to increase our chance to obtain a detailed mapping of visual cortical regions.

8.2.1.3. Niconnect project

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

- **Context:** The NiConnect project (2012-2016) 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

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. HBP

Type: FP7

Defi: Future and Emerging Technologies

Instrument: Collaborative Project with Coordination and Support Action

Objectif: FET Flagships

Duration: October 2013 - March 2016

Coordinator: Henry Markram (EPFL, Switzerland)

Partners: 86 partners, https://www.humanbrainproject.eu/fr/discover/the-community/partners

Inria contact: Olivier Faugeras

Abstract:

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 disease and build revolutionary new computing technologies. Today, for the first time, modern ICT has brought these goals within sight.

Convergence of ICT and Biology The convergence between biology and ICT has reached a point at which it can turn the goal of understanding the human brain into a reality. This realization motivates the Human Brain Project – an EU Flagship initiative in which over 80 partners will work together to realize a new "ICT-accelerated" vision for brain research and its applications.

One of the major obstacles to understanding the human brain is the fragmentation of brain research and the data it produces. Our most urgent need is thus a concerted international effort that uses emerging emerging ICT technologies to integrate this data in a unified picture of the brain as a single multi-level system.

Research Areas The HBP will make fundamental contributions to neuroscience, to medicine and to future computing technology.

In *neuroscience*, the project will use neuroinformatics and brain simulation to collect and integrate experimental data, identifying and filling gaps in our knowledge, and prioritizing future experiments.

In *medicine*, the HBP will use medical informatics to identify biological signatures of brain disease, allowing diagnosis at an early stage, before the disease has done irreversible damage, and enabling personalized treatment, adapted to the needs of individual patients. Better diagnosis, combined with disease and drug simulation, will accelerate the discovery of new treatments, drastically lowering the cost of drug discovery.

In *computing*, new techniques of interactive supercomputing, driven by the needs of brain simulation, will impact a vast range of industries. Devices and systems, modeled after the brain, will overcome fundamental limits on the energy-efficiency, reliability and programmability of current technologies, clearing the road for systems with brain-like intelligence.

The Future of Brain Research

Applying ICT to brain research and its applications promises huge economic and social benefits. But to realize these benefits, the technology needs to be made accessible to scientists – in the form of research platforms they can use for basic and clinical research, drug discovery and technology development. As a foundation for this effort, the HBP will build an integrated system of ICT-based research platforms, building and operating the platforms will require a clear vision, strong, flexible leadership, long-term investment in research and engineering, and a strategy that leverages the diversity and strength of European research. It will also require continuous dialogue with civil society, creating consensus and ensuring the project has a strong grounding in ethical standards.

The Human Brain Project will last ten years and will consist of a ramp-up phase (2013-2016) followed by an operational phase (2016-2023). Bertrand Thirion is responsible for the 2.1.1 task, *Anatomo-functional mapping of the human brain*.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

8.4.1.1. Internships

Gaspar Pizarro made a three months internship (January-March 2014), funded by Inria Chile and Conycit. His research topic was *Improving the fit of functional MRI data through the use of sparse linear models.*.

8.4.1.2. Other visitors

Danilo Bzdok (Forschungszentrum Jülich, institue of neuroscience and medicine) visited Parietal several months in 2014 (February-March, then September-), to develop collaborations on the use of machine learning techniques to model behavioral variables and find data-driven characterization of brain diseases.

8.4.2. Visits to International Teams

8.4.2.1. Research stays abroad

As part of the SubSample Digiteo chair, Alexandre Abraham spent six months in the USA at Stony Brook University and Nathan Klein Institute.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. Member of the organizing committee

- Gaël Varoquaux: PRNI, Euroscipy
- Philippe Ciuciu: ISBI

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committee

P.Ciuciu: ICASSP 2014, Associate Editor of the BISP (Bio Imaging Signal Processing) section

9.1.2.2. Reviewer

- Bertrand Thirion: IEEE ISBI, IPMI, MICCAI, IEEE PRNI, MLINI, NIPS
- Gaël Varoquaux: IPMI, MICCAI, IEEE PRNI, MLINI
- Philippe Ciuciu: MICCAI, IEEE ICASSP, IEEE ISBI, EUSIPCO, IEEE EMBC, IEEE PRNI

9.1.3. Journal

9.1.3.1. Member of the editorial board

- Bertrand Thirion: Frontiers in Neuroscience, Brain Imaging Methods
- Gaël Varoquaux: NeuroImage, Frontiers in NeuroInformatics and Frontiers in Brain Imaging methods
- Philippe Ciuciu: Frontiers in Neuroscience, Brain Imaging Methods

9.1.3.2. Reviewer

- Bertrand Thirion: Medical Image Analysis, IEEE Transactions on Medical Imaging, NeuroImage, Human Brain Mapping, PNAS, Nature Neuroscience.
- Gaël Varoquaux: HBM, MedIA, TMI, Frontiers in NeuroInformatics, Frontiers in Brain Imaging methods, Trends in cognitive science
- Philippe Ciuciu: SIAM Journal on Imaging Science, IEEE Trans Image Processing, IEEE Trans Medical Imaging, Proceedings of the IEEE, Signal Processing, NeuroImage, Journal of Neuroscience Methods, Plos One, MAGMA, Human Brain Mapping, Journal of Neural Systems, Journal of Neuroscience.

9.1.4. Scientific workshop

Bertrand Thirion took part to the CCC Brain workshop, Washington DC, Dec. 3-5.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Gael Varoquaux

- Stat Course cogmaster (3 × 3H)
- Python course Inria Rocquencourt et Rennes: 8Hrs each time
- Optimization tutoral at Euroscipy: 2H
- Scikit-learn tutorial at Scipy: 4H
- Functional connectivity course at OHBM: 30mn
- MSR TechDays 2014 : Scikit-Learn: Machine Learning en Python (2H)

Bertrand Thirion

- Master MVA, Imagerie fonctionnelle cérébrale et interface cerveau machine, 12h + 3h, M2, ENS Cachan, France.
- Functional connectivity course at ISMRM: 30mn
- Machine learning course, cortical mapping course at OHBM: 2 × 30 mn.

Philippe Ciuciu

- Master 2 Biomedical Engineering Université Paris V & Télécom Paris-Tech
- Master 2 Imagerie Médicale Université Paris-Sud

9.2.2. Supervision

PhD defense:

- Viviana Siless: July 8th, *Multi-modal registration of T1 brain image and geometric descriptors of white matter tracts*.
- Nicolas Zilber: March 10th, *ERF and scale-free analyses of source-reconstructed MEG brain signals during a multisensory learning paradigm*.
- Hao Xu: March 31st, *Probabilistic atlas statistical estimation with multimodal datasets and its application to atlas based segmentation*.

9.2.3. Juries

Bertrand Thirion was reviewer in the following PhD thesis committees:

- Mathieu Ruiz
- Ben Cassidy, university of South Wales, Australia
- Kasper Winther Andersen, April 22nd, DTU, Denmark

He was Examiner in the following PhD thesis committee:

- Hugo Raguet, Sept 22nd, Université Paris Dauphine.

9.3. Popularization

- Bertrand Thirion has taken part to the *Autour de la question* broadcast program on Feb. 24th.
- Bertrand Thirion co-authored the popularization paper : "Le décodage cérébral : exemple de la vision" in Clefs CEA, issue 62.
- Gaël Varoquaux gave a tutorial on scikit learn at the Microsoft Tech Days (February).

10. Bibliography

Major publications by the team in recent years

[1] N. CHAUFFERT, P. CIUCIU, J. KAHN, P. WEISS. *Variable density sampling with continuous trajectories. Application to MRI*, in "SIAM Journal of Imaging Sciences", 2014, vol. 7, n^{o} 4, pp. 1962–1992, https://hal.inria.fr/hal-00908486

[2] B. DA MOTA, V. FRITSCH, G. VAROQUAUX, T. BANASCHEWSKI, G. J. BARKER, A. L. W. BOKDE, U. BROMBERG, P. J. CONROD, J. GALLINAT, H. GARAVAN, J.-L. MARTINOT, F. NEES, T. PAUS, Z. PAUSOVA, M. RIETSCHEL, M. N. SMOLKA, A. STRÖHLE, V. FROUIN, J.-B. POLINE, B. THIRION. *Randomized parcellation based inference*, in "NeuroImage", November 2013, pp. 203 - 215, Digitéo (HiDiNim project and ICoGeN project) [*DOI :* 10.1016/J.NEUROIMAGE.2013.11.012], http://hal.inria.fr/hal-00915243

[3] A. KNOPS, B. THIRION, E. HUBBARD, V. MICHEL, S. DEHAENE. *Recruitment of an area involved in eye movements during mental arithmetic*, in "Science", Jun 2009, vol. 324, n^{o} 5934, pp. 1583–1585

[4] V. MICHEL, A. GRAMFORT, G. VAROQUAUX, E. EGER, B. THIRION. *Total variation regularization for fMRI-based prediction of behaviour*, in "IEEE Transactions on Medical Imaging", February 2011, vol. 30, n^{o} 7, pp. 1328 - 1340 [*DOI :* 10.1109/TMI.2011.2113378], http://hal.inria.fr/inria-00563468/en

[5] 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

[6] 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

[7] G. VAROQUAUX, A. GRAMFORT, F. PEDREGOSA, V. MICHEL, B. THIRION. *Multi-subject dictionary learning to segment an atlas of brain spontaneous activity*, in "Information Processing in Medical Imaging", Kaufbeuren, Germany, Lecture Notes in Computer Science, Springer, July 2011, vol. 6801, pp. 562-573 [*DOI :* 10.1007/978-3-642-22092-0_46], http://hal.inria.fr/inria-00588898/en

[8] G. VAROQUAUX, A. GRAMFORT, J.-B. POLINE, B. THIRION. *Brain covariance selection: better individual functional connectivity models using population prior*, in "Advances in Neural Information Processing Systems", Canada Vancouver, John Lafferty, Dec 2010

[9] G. VAROQUAUX, A. GRAMFORT, B. THIRION. *Small-sample brain mapping: sparse recovery on spatially correlated designs with randomization and clustering*, in "International Conference on Machine Learning", Edimbourg, United Kingdom, L. JOHN, P. JOELLE (editors), Andrew McCallum, June 2012, http://hal.inria.fr/hal-00705192

[10] G. VAROQUAUX, Y. SCHWARTZ, P. PINEL, B. THIRION. *Cohort-level brain mapping: learning cognitive atoms to single out specialized regions*, in "IPMI - Information Processing in Medical Imaging - 2013", Asilomar, United States, W. M. WELLS, S. JOSHI, K. M. POHL (editors), Springer, July 2013, vol. 7917, pp. 438-449 [*DOI :* 10.1007/978-3-642-38868-2_37], http://hal.inria.fr/hal-00841502

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] F. PEDREGOSA. *Feature extraction and supervised learning on fMRI: from practice to theory*, Université Pierre et Marie Curie, February 2015, https://tel.archives-ouvertes.fr/tel-01100921

[12] N. ZILBER. *ERF and scale-free analyses of source-reconstructed MEG brain signals during a multisensory learning paradigm*, Université Paris Sud - Paris XI, March 2014, https://tel.archives-ouvertes.fr/tel-00984990

Articles in International Peer-Reviewed Journals

[13] K. AMUNTS, M. HAWRYLYCZ, C. VAN ESSEN, J. VAN HORN., N. HAREL, J.-B. POLINE, F. DE MARTINO, J. G. BJAALIE, G. DEHAENE-LAMBERTZ, S. DEHAENE, P. VALDES-SOSA, B. THIRION, K. ZILLES, S. HILL, M. ABRAMS, P. TASS, W. VANDUFFEL, A. EVANS, S. EICKHOFF. *Interoperable atlases of the human brain*, in "NeuroImage", June 2014, vol. 99, n^{o} 1, 8 p. [*DOI :* 10.1016/J.NEUROIMAGE.2014.06.010], https://hal.inria.fr/hal-01094749

[14] L. CHAARI, P. CIUCIU, S. MÉRIAUX, J.-C. PESQUET. *Spatio-temporal wavelet regularization for parallel MRI reconstruction: application to functional MRI*, in "Magnetic Resonance Materials in Physics, Biology and Medicine", March 2014, 41 p. [*DOI :* 10.1007/S10334-014-0436-5], https://hal.inria.fr/hal-01084324

[15] N. CHAUFFERT, P. CIUCIU, J. KAHN, P. WEISS. *Variable density sampling with continuous trajectories. Application to MRI*, in "SIAM Journal of Imaging Sciences", 2014, vol. 7, n^{o} 4, pp. 1962–1992, https://hal.inria.fr/hal-00908486

[16] P. CIUCIU, P. ABRY, J. HE. *Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks*, in "NeuroImage", July 2014, vol. 95, pp. 248 - 263 [*DOI :* 10.1016/J.NEUROIMAGE.2014.03.047], https://hal.inria.fr/hal-01084242

[17] B. DA MOTA, R. TUDORAN, A. COSTAN, G. VAROQUAUX, G. BRASCHE, P. J. CONROD, H. LEMAITRE, T. PAUS, M. RIETSCHEL, V. FROUIN, J.-B. POLINE, G. ANTONIU, B. THIRION. *Machine Learning Patterns for Neuroimaging-Genetic Studies in the Cloud*, in "Frontiers in Neuroinformatics", April 2014, vol. 8 [*DOI :* 10.3389/FNINF.2014.00031], https://hal.inria.fr/hal-01057325

[18] A.-C. MILAZZO, B. NG, H. JIANG, W. SHIRER, G. VAROQUAUX, J.-B. POLINE, B. THIRION, M. D. GREICIUS. *Identification of Mood-Relevant Brain Connections Using a Continuous, Subject-Driven Rumination Paradigm*, in "Cerebral Cortex", October 2014, 12 p. , https://hal.inria.fr/hal-01094759

[19] F. PEDREGOSA, M. EICKENBERG, P. CIUCIU, B. THIRION, A. GRAMFORT. *Data-driven HRF estimation for encoding and decoding models*, in "NeuroImage", January 2015, pp. 209–220, https://hal.inria.fr/hal-00952554

[20] S. TAKERKART, G. AUZIAS, B. THIRION, L. RALAIVOLA. *Graph-based inter-subject pattern analysis of fMRI data*, in "PLoS ONE", July 2014 [*DOI :* 10.1371/JOURNAL.PONE.0104586], https://hal.archives-ouvertes.fr/hal-01027769

[21] B. THIRION, G. VAROQUAUX, E. DOHMATOB, J.-B. POLINE. *Which fMRI clustering gives good brain parcellations?*, in "Frontiers in neuroscience", May 2014, vol. 8, n^{o} 167, 13 p. [*DOI :* 10.3389/FNINS.2014.00167], https://hal.inria.fr/hal-01015172

[22] G. VAROQUAUX, B. THIRION. *How machine learning is shaping cognitive neuroimaging*, in "GigaScience", 2014, vol. 3, 28 p. [*DOI :* 10.1007/S12021-008-9041-Y], https://hal.inria.fr/hal-01094737

[23] T. VINCENT, S. BADILLO, L. RISSER, L. CHAARI, C. BAKHOUS, F. FORBES, P. CIUCIU. *Frontiers in Neuroinformatics Flexible multivariate hemodynamics fMRI data analyses and simulations with PyHRF*, in "Frontiers in Neuroscience", April 2014, vol. 8, n^{o} Article 67, 23 p. [*DOI :* 10.3389/FNINS.2014.00067], https://hal.inria.fr/hal-01084249

[24] H. XU, B. THIRION, S. ALLASSONNIÈRE. *Probabilistic Atlas and Geometric Variability Estimation to Drive Tissue Segmentation*, in "Statistics in Medicine", April 2014, vol. 33, n° 20, 24 p. [*DOI :* 10.1002/SIM.6156], https://hal.inria.fr/hal-01094739

[25] N. ZILBER, P. CIUCIU, A. GRAMFORT, L. AZIZI, V. VAN WASSENHOVE. *Supramodal processing optimizes visual perceptual learning and plasticity*, in "NeuroImage", June 2014, vol. 93, pp. 32 - 46 [*DOI :* 10.1016/J.NEUROIMAGE.2014.02.017], https://hal.inria.fr/hal-01084251

Invited Conferences

[26] F. FORBES, A. FRAU-PASCUAL, T. VINCENT, J. SLOBODA, P. CIUCIU. *Physiologically informed Bayesian analysis of ASL fMRI data*, in "Statistical Challenges in Neuroscience workshop", Warwick, United Kingdom, September 2014, https://hal.inria.fr/hal-01107613

International Conferences with Proceedings

[27] M. ALBUGHDADI, L. CHAARI, F. FORBES, J.-Y. TOURNERET, P. CIUCIU. *Model Selection for Hemodynamic Brain Parcellation in fMRI*, in "22nd European Signal Processing Conference", Lisbon, Portugal, September 2014, https://hal.inria.fr/hal-01107475

[28] Y. BEKHTI, N. ZILBER, F. PEDREGOSA, P. CIUCIU, V. VAN WASSENHOVE, A. GRAMFORT. *Decoding perceptual thresholds from MEG/EEG*, in "Pattern Recoginition in Neuroimaging (PRNI) (2014)", Tubingen, Germany, June 2014, https://hal.archives-ouvertes.fr/hal-01032909

[29] V. BORGHESANI, F. PEDREGOSA, E. EGER, M. BUIATTI, M. PIAZZA. *A perceptual-to-conceptual gradient of word coding along the ventral path*, in "Pattern Recognition in Neuroimaging", Tubingen, Germany, IEEE, June 2014, https://hal.inria.fr/hal-00986606

[30] E. DOHMATOB, A. GRAMFORT, B. THIRION, G. VAROQUAUX. *Benchmarking solvers for TV-l1 least-squares and logistic regression in brain imaging*, in "Pattern Recoginition in Neuroimaging (PRNI)", Tübingen, Germany, IEEE, June 2014, https://hal.inria.fr/hal-00991743

[31] A. FRAU-PASCUAL, T. VINCENT, F. FORBES, P. CIUCIU. *Hemodynamically informed parcellation of cerebral FMRI data*, in "ICASSP - IEEE International Conference on Acoustics, Speech and Signal Processing", Florence, Italy, IEEE, May 2014, pp. 2079-2083 [*DOI :* 10.1109/ICASSP.2014.6853965], https://hal.inria.fr/hal-01100186

[32] A. FRAU-PASCUAL, T. VINCENT, J. SLOBODA, P. CIUCIU, F. FORBES. *Physiologically Informed Bayesian Analysis of ASL fMRI Data*, in "BAMBI 2014 - First International Workshop on Bayesian and grAphical Models for Biomedical Imaging", Boston, United States, M. J. CARDOSO, I. SIMPSON, T. ARBEL, D. PRECUP, A. RIBBENS (editors), Lecture Notes in Computer Science, Springer International Publishing, September 2014, vol. 8677, pp. 37 - 48 [*DOI :* 10.1007/978-3-319-12289-2_4], https://hal.inria.fr/hal-01100266

[33] D. KEATOR, S. GHOSH, C. MAUMET, G. FLANDIN, N. NICHOLS, E. THOMAS, G. BURNS, R. BRUEHL, C. CRADDOCK, B. FREDERICK, K. GORGOLEWSKI, D. MARCUS, M. HANKE, C. HASELGROVE, K. HELMER, A. KLEIN, M. MILHAM, R. POLDRACK, F. MICHEL, J. STEFFENER, Y. SCHWARTZ, R. STONER, J. TURNER, N. KENNEDY, J.-B. POLINE. *Developing and using the Neuroimaging and Data Sharing Data Model: the NIDASH Working Group*, in "20th Annual Meeting of the Organization for Human Brain Mapping", Hamburg, Germany, June 2014, http://www.hal.inserm.fr/inserm-01101842

[34] B. NG, M. DRESSLER, G. VAROQUAUX, J.-B. POLINE, M. GREICIUS, B. THIRION. *Transport on Riemannian Manifold for Functional Connectivity-based Classification*, in "MICCAI - 17th International Conference on Medical Image Computing and Computer Assisted Intervention", Boston, United States, Springer, September 2014, https://hal.inria.fr/hal-01058521

[35] R. PHLYPO, B. THIRION, G. VAROQUAUX. *Deriving a multi-subject functional-connectivity atlas to inform connectome estimation*, in "Medical Image Computing and Computer-Assisted Intervention", Boston, United States, P. GOLLAND, N. HATA, C. BARILLOT, J. HORNEGGER, R. HOWE (editors), Springer International Publishing, September 2014, vol. 8675, pp. 185-192, MICCAI 2014 preprint [*DOI :* 10.1007/978-3-319-10443-0_24], https://hal.inria.fr/hal-00991124

[36] B. THIRION, G. VAROQUAUX, O. GRISEL, C. POUPON, P. PINEL. *Principal Component Regression predicts functional responses across individuals*, in "MICCAI", Boston, United States, Springer, September 2014, https://hal.inria.fr/hal-01015173

[37] F. YEPES, F. PEDREGOSA, B. THIRION, Y. WANG, N. LEPORE. *Automatic pathology classification using a single feature machine learning - support vector machines*, in "SPIE Medical Imaging 2014", San Diego, United States, February 2014, vol. 9035, n^{o} 1, 24 p. [*DOI :* 10.1117/12.2043943], https://hal.archives-ouvertes.fr/hal-00975872

Scientific Books (or Scientific Book chapters)

[38] B. NG, M. TOEWS, S. DURRLEMAN, Y. SHI. *Shape Analysis for Brain Structures: A Review*, in "Shape Analysis in Medical Image Analysis", S. LI, J. M. R. S. TAVARES (editors), Lecture Notes in Computational Vision and Biomechanics, Springer, 2014, vol. 14, https://hal.inria.fr/hal-00925536

[39] X. PENNEC, P. FILLARD. *Statistical Computing On Non-Linear Spaces For Computational Anatomy*, in "Handbook of Biomedical Imaging: Methodologies and Clinical Research", N. PARAGIOS, J. DUNCAN, N. AYACHE (editors), Springer, 2014, forthcoming, https://hal.inria.fr/inria-00616201

Other Publications

[40] A. ABRAHAM, E. DOHMATOB, B. THIRION, D. SAMARAS, G. VAROQUAUX. *Region segmentation for sparse decompositions: better brain parcellations from rest fMRI*, September 2014, 8 p. , Sparsity Techniques in Medical Imaging, https://hal.inria.fr/hal-01093944

[41] N. CHAUFFERT, P. WEISS, J. KAHN, P. CIUCIU. *Gradient waveform design for variable density sampling in Magnetic Resonance Imaging*, December 2014, https://hal.inria.fr/hal-01095320

[42] S. MEDINA. *Unsupervised Clustering of Neural Pathways*, Universidad de Buenos Aires, February 2014, 75 p. , https://hal.inria.fr/hal-00908433

[43] F. PEDREGOSA, F. BACH, A. GRAMFORT. *On the Consistency of Ordinal Regression Methods*, June 2014, https://hal.inria.fr/hal-01054942

References in notes

[44] S. MOELLER, E. YACOUB, C. A. OLMAN, E. AUERBACH, J. STRUPP, N. HAREL, K. UĞURBIL. *Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high*

spatial and temporal whole-brain fMRI, in "Magn Reson Med", May 2010, vol. 63, n^{o} 5, pp. 1144–1153, http://dx.doi.org/10.1002/mrm.22361