NNT: 2015SACLS159

Thèse de Doctorat
de
l'Université Paris-Saclay
préparée à
l'Université Paris-Sud

École Doctorale n°580
Sciences et Technologies de l'Information et de la communication

Spécialité de doctorat : Informatique

Par

M. Alexandre Abraham

Apprentissage d'atlas fonctionnels du cerveau modélisant la variabilité inter-individuelle

Thèse présentée et soutenue à Saclay, le 30 Novembre 2015:

Composition du Jury :

M. Hervé Delingette	Directeur de recherche, Inria	Examinateur, Président du jury
M. Hugues Talbot	Professeur, ESIEE	Rapporteur
M. Mark Woolrich	Professeur, OHBA	Rapporteur
Mme Catherine Oppenheim	Professeur, CH Sainte-Anne	Examinateur
M. Nikos Paragios	Professeur, École Centrale	Examinateur
M. Bertrand Thirion	Directeur de recherche, Inria	Directeur de thèse
M. Dimitris Samaras	Professeur, Stony Brook University	Co-encadrant de thèse
M. Gaël Varoquaux	Chargé de recherche, Inria	Co-encadrant de thèse

Version du 9 février 2016

NNT: 2015SACLS159

PhD Thesis
of
University of Paris-Saclay
prepared at
University of Paris-Sud

Doctoral School n°580
Information and Communication Sciences and Technologies

PhD Specialty : Computer Science

By

Mr. Alexandre Abraham

Learning functional brain atlases modeling inter-subject variability

Thesis presented and defended in Saclay, the 30th of November 2015:

Composition of the Jury:

Mr. Hervé Delingette	Research Director, Inria	Examiner, Jury president
Mr. Hugues Talbot	Professor, ESIEE	Reviewer
Mr. Mark Woolrich	Professor, OHBA	Reviewer
Mrs. Catherine Oppenheim	Professor, CH Sainte-Anne	Examiner
Mr. Nikos Paragios	Professor, École Centrale	Examiner
Mr. Bertrand Thirion	Research Director, Inria	Doctoral advisor
Mr. Dimitris Samaras	Professor, Stony Brook University	Co-supervisor
Mr. Gaël Varoquaux	Research Associate, Inria	Co-supervisor

Version as of February 9, 2016

Acknowledgements [1]

[1] I hope that I have cited everybody. Forgotten ones are eligible for a free dinner.

I would like to express my gratitude to my thesis supervisors Dimitris Samaras, that never let me down and even brought me to the USA, Gaël Varoquaux, that guided in the fog of research during these 3 years, and Bertrand Thirion, that incited me to be always rigorous! I would also like to thank my reviewers, Hugues Talbot and Mark Woolrich that were brave enough to read my thesis till the end and give me their useful insights, as well as Nikos Paragios, Hervé Delingette and Catherine Oppenheim for being part of my thesis committee. Thanks also to Cameron Craddock, Michael Milham, Adriana Di Martino, Christine De Lorenzo and Ramin Parsey for their warm welcome in their respective labs. Special mention to Thierry Géraud and Nicolas Sabouret for introducing me in the world of research.

My greatest thanks goes to the Parietal team whose spirit shaped my PhD. Namely, (I apologize in advance for people that I may forget), Philippe Ciuciu – still not sure how to pronounce it –, Virgile Fritsch – the most stable and balanced guy I've ever met –, Michael Eickenberg – for being so irrelephant –, Yannick Schwartz – our human ressources specialist –, Fabian Pedregosa – best PhD defense buffet ever –, Benoît Da Mota – for petaflopping our algorithms –, Sergio Medina – who brought the sun into the team –, Viviana Siless – for the background songs –, Mehdi Rahim – serious contender for the title of most stable and balanced guy –, Elvis Dohmatob – ça va dans le bon sens! –, Aina Frau Pascual – for being so shy –, Andrés Idrobo – for the columbian "sugar" –, Arthur Mensch – for taking over the dictionary learning development –, Kamalakar Dadi – the new nilearn specialist –, Ronald Phlypo – for bringing me back home after the MICCAI 2014 dinner. I'm sure he has some unforgettable memories because I have none –, Philippe Gervais – for cleaning the code of nilearn –, Solveig Badillo – that never invited me to her parties –, Nicolas Chauffert – what happens in Porquerolles... –, Salma Bougacha – for her neverending complaints –, Loïc Estève – for handling all infrastructure related problems of nilearn –, Olivier Grisel – for showing us how futile is our work compated to deep learning –, Kyle Kastner – for finding slothes in images –, Jaques Grobler – they're taking the hobbits to Isengard! –, Danilo Bzdok – for pointing me some papers –, Ana Luìsa Pinho – for the special PhD gifts –, and also Bernard Ng, Jacob Schreiber, Giorgio Patrini, Hubert Pellé, Léonard Blier, Konstantin Shmelkov, Torsten Ruest, Nicolas Zilber, and Fernando Yepes.

Big thanks also to all other people from the Neurospin lab! In particular, Elodie Cauvet – that made me join the team –, Vincent Frouin – for helping with our problems when he could –, Catherine Wacongne – for her patience when I had meetings with Gaël in her office –, Baptiste Gauthier – for being the maximum bound of all depression tests –, Valentina Borghesani – for being the maximum bound of all optimism tests –, Murielle Fabre – for the chinese lessons –, Sandrine Lefranc – the best bus buddy –, Antoine Grigis – the second best bus buddy –, Edouard Duchesnay – for our collaborations –, and Jennifer Boisgontier – that struggled with ASD too –.

Special thanks to all people in charge of administration that deserve their own section. Namely, Régine Bricquet – the Parietal's virtuose of the administration. She makes all the struggle go away, it's almost magic –, Marie Domingues – that always helped me understand the Inria rules –, and Giovanna Santoro – that took care of me for my first contract –. And obviously nothing would have been the same without the everyday smiles of Joana and Elodie. Thanks also to Denis for helping when we had technical difficulties during the defenses.

And obviously all the people that supported me in my everyday life. First, my family: Audrey, my wife, Christophe and Delphine, Éléa, Hugo, Zoé, Zofia and Bernard, Christian and Marie, Yoann and Christine, Antonin, and Alexandre. All my friends, Julien, Pierre, Olivier, Éric and Marie, Fabrice, Marie and Sébastien, Leslie, Morgan, Jules, Kévin, Florian, Sly, LLB, Nicuvéo, Yabo and the other Test.ers, Lydie, Jules, Laure and Alexandre – chatting buddies during my thesis redaction –, Carlo and Pamela, and Armand – another bus buddy! –.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Apprentissage d'atlas fonctionnels du cerveau modélisant la variabilité inter-individuelle
Mots-clés: IRMf au repos, apprentissage de dictionnaire, connectivité fonctionnelle, biomarqueurs

Résumé : L'Imagerie par Résonance Magnétique fonctionnelle (IRMf) est une source prometteuse de biomarqueurs permettant le diagnostic de troubles neuropsychiatriques sur des sujets non coopératifs. Un connectome est communément estimé en établissant un atlas de régions cérébrales représentatif de l'organisation fonctionnelle puis en étudiants la corrélation entre leurs signaux.

La première étape, la définition de régions, est réalisée à l'aide d'une analyse statistique de l'activité cérébrale spontanée au repos. Pour les extraire, nous utilisons une approche d'apprentissage de dictionnaire multi-sujets intégrant une pénalité imposant compacité spatiale et parcimonie. Nous extrayons les unités fonctionnelles de base des réseaux fonctionnels extraits à l'aide de techniques de segmentation inspirées du domaine de la vision. Nous montons à l'échelle sur de gros jeux de données en utilisant une stratégie d'optimisation stochastique.

L'absence de vérité de terrain est un frein à l'évaluation de la pertinence de nos modèles. Pour pallier ce problème, nous introduisons des métriques évalouant la stabilité et la fidélité des modèles générés. Appliquées à plusieurs sous-ensembles d'un grand jeu de données, nous montrons que nos méthodes de définition de régions expliquent mieux les données et sont plus stables à travers les sujets que les méthodes de référence.

Nous intégrons ensuite notre méthode de définition de régions dans un pipeline entièrement automatisé afin d'estimer un connectome et de l'utiliser dans des tâches de prédiction. Nous étudions sa pertinence empirique sur une tâche clinique de diagnostic des troubles autistiques et montrons la première prédiction de trouble psychiatrique à travers différents sites d'acquisition et sur des sous-ensembles d'homogénéité variable. Nos résultats de prédiction sont supérieurs à l'état de l'art. Nous démontrons que la prédiction sur des sites inconnus est aussi efficace que la prédictions sur site connus à condition d'avoir suffisamment de sujets. Les neurophénotypes extraits sont compatibles avec la littérature.

Enfin, par une analyse post-hoc des résultats, nous montrons que la définition de région est l'étape la plus importante du pipeline et que l'approche que nous proposons obtient les meilleurs résultats. Nous fournissons également des recommandations sur les méthodes les plus performantes pour les autres étapes du pipeline.

Title: Learning functional brain atlases modeling inter-subject variability
Keywords: resting state fMRI, dictionary learning, functional connectivity, biomarkers

Abstract: Resting-state functional Magnetic Resonance Imaging (fMRI) holds the promises to reveal functional biomarkers for neuropsychiatric disorders applicable to non-cooperative patients. To extract such biomarkers, the standard approach is, first, to establish an atlas of the cerebral areas defining the functional organization of the brain, and then to study the correlation of their brain signals. They form the functional connectome.

The first step of building a connectome requires defining the regions that constitute brain functional units. This is done by performing a statistical analysis of the dynamics of spontaneous brain activity during resting-state. We introduce a method combining spatially-structured and sparsity-inducing penalties in a multi-subject dictionary learning approach to automatically extract brain networks from rest-fMRI. A stochastic optimization strategy enables scaling to big datasets. We then show how computer vision inspired segmentation techniques can be used to threshold automatically and break down these networks into connected functional units.

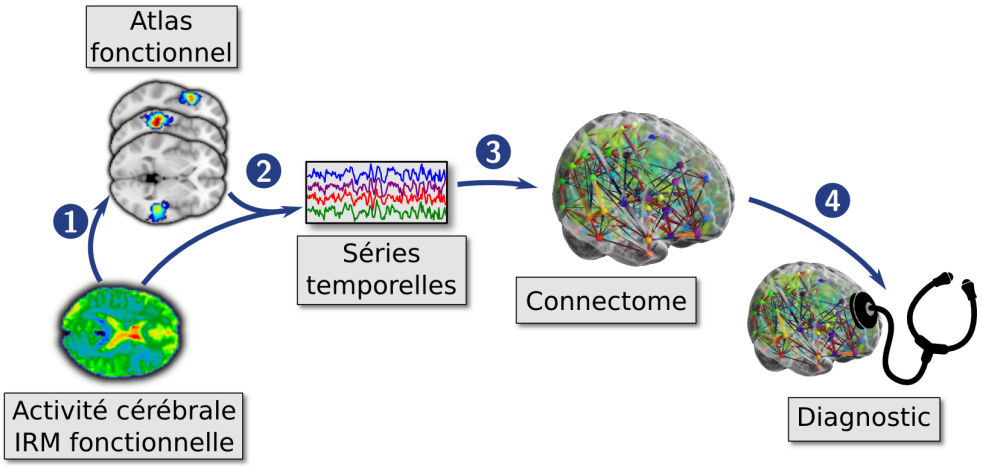
To compensate for the lack of ground truth, we introduce two metrics that aim at scoring the stability and the data fidelity of the generated models. Using these metrics, we show that our methods better explain the data and are more stable across subjects than reference decomposition or clustering methods.

We then integrate this region-definition method in a fully-automatic prediction pipeline, to build connectomes from the data and use them in classification tasks. We study its empirical relevance on the clinical task of predicting autism spectrum disorders. We demonstrate the first prediction of pyschiatric condition across different scanning sites and apply it on subsets of participants of variable homogeneity. We exhibit prediction scores higher than state of the art and show that, given a sufficient number of individuals in the training set, prediction across sites is as efficient as traditional prediction. We also extract autism neurophenotypes compatible with the litterature.

Finally, we show that region definition is the most important step of the pipeline and that our approach is the best performer. We also explore the other steps of the pipeline and give recommandations on how to choose a prediction pipeline.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Résumé : Apprentissage d'atlas fonctionnels du cerveau modélisant la variabilité inter-individuelle



L'Imagerie par Résonance Magnétique fonctionnelle (IRMf) capture le fonctionnement du cerveau et enregistre donc des données précieuses pour étudier les troubles neuro-psychiatriques. Mesuré au repos (sans tâche particulière à effectuer), ce signal renseigne sur l'activité de fond du cerveau et permet l'estimation d'un connectome : une représentation des intéractions fonctionnelles entre régions cérébrales (aussi appelée connectivité fonctionnelle). La mesure de repos est particulièrement adaptée à l'études des neuropathologies, car elle s'applique aux sujets handicapés, non-coopératifs ou sédatés. Il a d'ailleurs été prouvé que certaines pathologies modifient cette connectivité. Malheureusement, la plupart des études neuro-psychiatriques se cantonnent à quelques dizaines de sujets en raison du coût d'acquisition et de limitations logicielles. Pour pallier ce problème, de récentes initiatives ont permis la diffusion de jeux de données de neuroimagerie de grande taille (plus de 1000 sujets) en regroupant des données provenant de nombreux sites cliniques. Ces gros volumes de données sont une opportunité fantastique pour les études neuro-psychiatriques, dont les conclusions obtenues sur quelques dizaines de sujets peuvent être soumises à caution. Toutefois, une telle aggrégation n'est pas sans conséquence puisqu'elle ajoute une variaibilité propre à chaque site dans les données.

A la croisée des mathématiques, de l'informatique et des neurosciences, j'ai contribué par mon travail de thèse à des innovations sur 3 aspects différents: *i)* au niveau algorithmique, j'ai proposé une nouvelle méthode pour extraire des régions cérébrales à partir de données d'imagerie fonctionnelle prenant en compte la variabilité inter-individuelle, *ii)* dans le domaine applicatif, j'ai utilisé ces régions cérébrales avec des modèles de machine learning pour améliorer sensiblement l'état de l'art sur le diagnostic d'autisme et j'ai réalisé une analyse statistique complète de ces résultats afin de trouver la chaîne de traitement optimale, et *iii)* dans le domaine informatique, j'ai développé un logiciel Python open-source permettant une collaboration plus aisée entre neuro-scientifiques et informaticiens proposant des implémentations performantes d'algorithmes propres au domaine et capable de traiter de grands jeux de données.

Apprentissage de dictionnaire avec régularisation par variation totale parcimonieuse

Les performances des méthodes de machine learning dépendent grandement de la dimension des données d'apprentissage. Les données IRMf étant composées d'une succession de 150 à 1000 images 3D comptant 100 000 voxels cérébraux, il est impossible de les analyser directement en raison du fléau de la dimension. Pour s'affranchir de cet effet, nous réduisons la dimension du problème en segmentant le cerveau en sous-unité fonctionnelles, les régions cérébrales.La méthode de référence utilisée pour cette tâche est l'ICA, qui extrait des composantes temporellement indépendantes. Toutefois,

l'ICA extrait des composantes couvrant tout le cerveau et doit donc être suivie d'un seuillage et d'une extraction de composantes connexes pour obtenir des régions. C'est pourquoi l'indépendance temporelle imposée par l'ICA a été remise en question au profit de la parcimonie spatiale, généralement appliquée à l'aide de méthode d'apprentissage de dictionnaire. Toutefois, la parcimonie imposée seule ne donne pas des cartes cérébrales satisfaisantes : les voxels sélectionnés dans les composantes sont disséminés à travers le cerveau. j'ai donc ajouté donc à la parcimonie une régularisation structurée basée sur un a priori neuroscientifique. En effet, de précédentes études sur l'anatomie et les protocoles de tâches nous informent que les régions cérébrales sont formées de neurones connexes et sont en général petites (entre 0.5% et 1% du cerveau sur les atlas de référence).

Contribution

Dans la continuité du développement d'une méthode d'apprentissage de dictionnaire multi-sujet au sein de mon laboratoire, j'ai ajouté à cette méthode un a priori neuroscientifique structuré : la variation totale. Couplée à la parcimonie, elle promeut atlas composés de petites régions compactes [1]. L'application d'une telle contrainte, très coûteuse, est rendue possible par la structure multi-sujet de la méthode qui applique la contrainte au niveau du groupe et non de chaque sujet individuellement, ainsi que par une optimisation algorithmique. Enfin, j'ai proposé une méthode d'extraction de régions à partir de cartes cérébrales, pouvant autant fonctionner à partir de composantes d'ICA que d'apprentissage de dictionnaire, afin de segmenter les régions proches, cas fréquent au niveau de la jointure inter-hémisphérique [2].

Résultats

L'absence de vérité de terrain est un frein à l'évaluation de la pertinence de nos modèles. Pour pallier ce problème, j'ai introduit des métriques quantitatives évaluant la stabilité et la fidélité des modèles générés [5]. Appliquées à plusieurs sous-ensembles d'un grand jeu de données, nos méthodes de définition de régions expliquent mieux les données et sont plus stables à travers les sujets que les méthodes de référence. Nous avons également demandé à des psychiatres spécialistes d'apporter une évalutation qualitative subjective des atlas générés par notre méthode. Ceux-ci ont été systématiquement privilégiés aux autres méthodes.

Diagnostic des troubles du spectre autistique

L'autisme est un spectre de troubles recouvrant plusieurs aspects de la vie quotidienne particulièrement difficile à diagnostiquer : sa définition même évolue avec le temps au fil des découvertes réalisées par les psychiatres. Récemment, des études provenant de 17 sites ont été agrégées en un jeu de données de plus de 1000 sujets appelé ABIDE. J'ai donc entrepris une expérience de diagnostic en collaboration avec Adriana Di Martino, en charge du projet ABIDE pour l'université de New York ainsi que Michael Milham et Cameron Craddock du Child Mind Institute. La procédure standard de diagnostic à partir de données IRMf commence par l'estimation des régions fonctionnelles (1) pour lesquelles on estime une activité moyenne en corrigeant d'éventuelles sources de bruit comme les mouvements (2), puis on estime un connectome cérébral (3) qui sert de base à la classification (4). Plusieurs méthodes existent pour chacune de ces étapes et aucun consensus n'a été trouvé sur le pipeline optimal. L'état de l'art en matière de diagnostic sur ABIDE est de 63%.

Contribution

Mon premier travail sur le pipeline de prédiction a été de le rendre complètement automatique en utilisant une validation croisée imbriquée. Cela permet d'obtenir des résultats optimaux de façon complètement automatique. J'ai étudié sa pertinence empirique sur une tâche clinique de diagnostic des troubles autistiques et montré la première prédiction de trouble psychiatrique à travers différents sites d'acquisition et sur des sous-ensembles d'homogénéité variable. Afin de rechercher le pipeline de prédiction optimal, j'ai exploré pour chaque étape plusieurs des méthodes couramment

utilisées. Les tester toutes conduit à une explosion combinatoire : c'est pourquoi les études précédentes se focalisent sur une seule étape mais obtiennent donc des résultats qui ne se généralisent pas. Afin d'être capable de supporter cette explosion, j'ai utilisé un cluster de calcul et ai ainsi pu traiter plus de 100 000 pipelines, soit plus de 10 ans d'études sur un ordinateur standard (8 coeurs, 32Gb RAM). J'ai ensuite réalisé une étude post-hoc des résultats qui ont mené à plusieurs conclusions importantes [4].

Résultats

Les résultats de prédiction (68%) sont supérieurs à l'état de l'art sur ce jeu de données. L'analyse post-hoc révèle que la prédiction en condition clinique est aussi efficace que la prédiction sur des sites connus à condition d'avoir suffisamment de sujets. J'ai également déterminé que l'étape la plus importante du pipeline est le choix de l'atlas et que l'atlas extrait par l'approche que j'ai proposé améliore significativement les résultats de prédiction par rapport aux méthodes d'extraction de référence et aux atlas existants. J'ai également pu fournir des recommandations sur les méthodes les plus performantes pour les autres étapes du pipeline. Les neurophénotypes propres à l'autisme extraits par mon modèle sont cohérents avec les hypothèses psychiatriques sur l'autisme et avec les résultats présents dans la littérature. Nous avons ensemble soumis une publication au journal NeuroImage [4].

Développement logiciel

En parallèle de mon travail de thèse, j'ai été le developpeur principal de nilearn, un package Python destiné à faciliter l'application de méthode de machine-learning sur des données de neuroimagerie et ainsi permettre une collaboration plus efficace entre experts en machine learning et neuroscientifiques. En effet, les logiciels existants sont spécifiques à certaines applications et leur interface est particulièrement adapté aux neuroscientifiques. En se basant sur le célèbre package scikit-learn, nilearn rend possible l'intégration directe d'algorithmes conçus par des experts techniques pour des applications neuroscientifiques. De plus, nilearn fournit des implémentations de plusieurs algorithmes de référence du domaines (ICA, Searchlight) rapides et optimisées pour un grand volume de données. Nilearn participe aussi à rendre la science plus reproductible puisque qu'une analyse entière peut-être retranscrite dans un unique script et mise à la disposition de la communauté. A ce jour, nilearn est adopté par une quantité croissante de laboratoire de recherche incluant par exemple le célèbre Montreal Neurological Institute. Certains de mes travaux de thèse ont déjà été intégrés dans nilearn et le reste le sera bientôt. Une introduction aux concepts sous-jacents de nilearn a déjà fait l'object d'un article de journal [3].

[1] Extracting brain regions from rest fMRI with Total-Variation constrained dictionary learning, *Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, Gaël Varoquaux*, MICCAI - 16th International Conference on Medical Image Computing and Computer Assisted Intervention - 2013, Sep 2013, Nagoya, Japan. Springer, 2013

[2] Region segmentation for sparse decompositions: better brain parcellations from rest fMRI, *Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, Gaël Varoquaux*, Sparsity Techniques in Medical Imaging, Sep 2014, Boston, United States. pp.8

[3] Machine Learning for Neuroimaging with Scikit-Learn, *Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Muller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, Gaël Varoquaux*, Frontiers in Neuroscience, 8, 2014.

[4] Toward Robust Functional-Connectivity Biomarkers of Autism, *Alexandre Abraham, Michael Milham, Adriana Di Martino, Cameron Craddock, Dimitris Samaras, Bertrand Thirion and Gaël Varoquaux*, Neuroimage, submitted

[5] Total-Variation Regularized Multi-subject Dictionary Learning, *Alexandre Abraham, Dimitris Samaras, Bertrand Thirion and Gaël Varoquaux*, Transactions on Medical Imaging - IEEE

Contents

1 Introduction **13**

1 Context of this thesis . . . 13

2 Layout of the manuscript . . . 14

2 Background: functional magnetic resonance imaging to study the brain **17**

1 Understanding brain diseases and mental disorders . . . 17

2 Functional magnetic resonance imaging . . . 17

3 Brain atlas extraction . . . 21

4 The Autism Brain Imaging Data Exchange (ABIDE) dataset . . . 23

3 Total-Variation Multi-Subject Dictionary Learning **27**

1 Introduction – atlas estimation methods . . . 27

2 A dictionary learning approach to segment regions . . . 28

3 Sparse TV penalization to enforce compact regions . . . 31

4 Computational optimization . . . 34

5 Effects of parameters on atlas extraction . . . 39

6 Results – Extracted brain atlases . . . 40

7 Conclusion . . . 43

4 Extracting brain regions from networks maps 45

1 *Introduction – Region extraction as a post-processing on brain atlas* . . . 45

2 *Region extraction methods* . . . 46

3 *Results – Extracted brain regions* . . . 51

4 *Conclusion* . . . 51

5 Selecting the best atlas: metrics for stability and data faithfulness 53

1 *Introduction – Interesting properties of brain atlases* . . . 54

2 *Stability with regards to training sets* . . . 54

3 *Measuring data faithfulness* . . . 57

4 *Model selection plots: exploring the trade-off* . . . 58

5 *Results – Best atlas estimation and region extraction methods* . . . 59

6 *Validation for a prediction purpose* . . . 63

7 *Conclusion* . . . 64

6 Learning predictive biomarkers of a psychiatric disease from resting state fMRI 71

1 *Introduction – Prediction from resting state functional MRI* . . . 71

2 *Biomarkers extraction pipeline* . . . 73

3 *Experiments – Predicting autism spectrum disorders* . . . 76

4 *Prediction accuracy in intra-site and inter-site settings* . . . 78

5 *Functional biomarkers* . . . 79

6 *Conclusion* . . . 81

7 What's in a pipeline? Important choices for prediction 85

1 *Post-hoc analysis of classification results* . . . 85

2 Best atlas extraction method 87
3 Best covariance matrix estimation 88
4 Best predictor 89
5 Conclusion 89

8 Conclusion **95**

1 Contributions 95
2 Perspectives 96

Bibliography **99**

1 | Introduction

1 Context of this thesis

The overarching goal of functional neuroimaging is the understanding of the cognitive organization of the brain and of its dynamics, in typical condition and when affected by a disease or a mental disorder. Studying the neural activity during various tasks or during rest allows us to segment the brain into several brain functional units. Besides understanding the brain, a clinical application is the extraction of biomarkers and the establishment of neurophenotypes – such as functional connectivity patterns – characterizing perturbations of brain function in brain diseases and mental disorders.

Functional connectivity between brain networks observed during resting state functional Magnetic Resonance Imaging (fMRI) is a promising modality to retrieve the underlying organization of the brain as it can be measured on several impaired subjects, as opposite to task-driven experiments. However, it suffers from several practical problems.

First, because of the prohibitive cost of image acquisition, most of the functional MRI studies so far are based on small datasets and may suffer of biases in subject selection and lack of generalization power. Recently released datasets, such as ABIDE that contains more than 1000 subjects, solved this problem by aggregating heterogeneous datasets acquired across several sites. But the methods developed to process a small number of individuals do not always scale to big datasets. Moreover, the results obtained on these datasets do not always transfer to bigger ones.

Then, resting-state functional MRI data suffers from low signal-to-noise ratio. The number of scans (usually between 100 and 1000) is also much smaller than the number of voxels in the brain (around 100000). This is a burden for the unsupervised learning approaches commonly used to segment the brain, that must rely on dimensionality reduction or on a strong prior on the data.

Finally, scaling up to large datasets and using sophisticated methods also raises problems of reproducibility of the experiments. Despite the numerous standard software available in this field, each study uses its own methods and prediction pipeline for clinical diagnosis, different cross-validation schemes to measure the generalization performance or select the parameters involved, and different metrics to evaluate the performance of the models. This wide

variability in experimental design makes it hard to confront the results from one study to another.

In this thesis, we propose to tackle all these challenges to reliably extract biomarkers and neurophenotypes. We take advantage of the recently released big datasets to enhance the results obtained by our unsupervised learning approach. We propose a new method to uncover the functional organization of the brain on large datasets while taking inter-subject variability into account. We demonstrate its efficiency on a dataset of more than 800 subjects. We also show that extracting connected brain regions from atlases is beneficial to both model evaluation and prediction.

By establishing a standard pipeline with no free parameter and diffusing its implementation, we promote the reproducibility of the experiments. In addition, a post-hoc analysis of the results allows one to easily evaluate the impact of a given method on a prediction task.

2 Layout of the manuscript

Chapter 2 introduces the functional magnetic resonance modality and the required background knowledge to study it in the case of resting-state studies. We highlight the characteristics of the extracted signals and the challenges faced when working on them. It is also the occasion to introduce autism spectrum disorders, the disease we study in this thesis and the dataset on which our experiments are based. We show that this particular disorder is not well understood and can benefit from the insight provided by functional neuroimaging. All the following chapters are contributions.

Chapter 3 proposes a new method to extract brain functional units from resting-state fMRI. Built upon the Multi-Subject Dictionary Learning approach, we introduce a new structure inducing regularization to extract neurologically plausible brain atlases. We go through the optimizations performed to make it scale to big datasets and show that the extracted atlas competes with the reference method for this task – spatial group independent component analysis (ICA). In particular, this method manages to extract components matching neuroanatomy in highly noisy areas of the brain.

Chapter 4 introduces region extraction methods. Matrix decomposition methods – TV-MSDL or ICA for example – sometimes fail at separating brain functional units, especially when they are close. Using component extraction methods coming from the computer vision domain – but adapted to our problem –, we show that a post-processing of brain maps increases the visual quality of extracted brain atlases.

Chapter 5 introduces metrics aimed at ranking brain atlases. As of today, no ground truth exists regarding the brain functional organization. As such, evaluating the performance of brain atlas extraction methods is hard and rely on prior knowledge. Here, we propose two complementary metrics to evaluate different aspects of brain atlases. We use them to draw conclusions about atlas estimation and region extraction methods. Finally, we validate them by using a prediction task as surrogate criterion.

Chapter 6 tackles the problem of prediction on unknown sites or, as we call it, inter-site prediction. Large datasets composed of an aggregation of smaller studies brings up the problem of site-specific confounds. In particular, in a clinical setting, a diagnosis tool must be able to process patients coming from sites not known at training time. Here, we propose a fully-automatic pipeline for prediction from resting-state fMRI. We evaluate ot on the classical intra-site setting and on inter-site prediction. We show that our pipeline gives comparable results for both tasks if given enough subjects during training.

We also address the problem of experiment reproducibility. In fact, our versatile pipeline allows one to replace any step by a custom algorithm. It solves the problem of parameter selection by setting them using an internal cross-validation. Combined with the nilearn[1] package, it allows one to reproduce an experiment using a single Python script.

[1] http://nilearn.github.io

Chapter 7 presents a post-hoc statistical analysis of the results obtain in chapter 5. Through this analysis, we are able to evaluate the importance of each step in the pipeline, and, for each step, we are able to select the best method. By observing these results on different levels of stratification, we are able to determine the strengths and weaknesses of the different approaches. Finally, we dig inside the predictor to find the neurophenotype of autism spectrum disorders.

2 | Background: functional magnetic resonance imaging to study the brain

1 Understanding brain diseases and mental disorders

Numerous conditions are affecting the brain. Brain diseases can be caused by biological causes: viruses, bacteria, trauma, stroke, etc. They can be explored using neuroimaging techniques such as X-rays or by medical examination. Mental disorders are cognitive disorders that can be caused by a brain disease. Their diagnosis may be difficult when it relies on the subjective perspective of the patient.

Diagnosing mental disorder is the role of psychiatry and psychology. In particular, the diagnostic and statistical manual of mental disorders (DSM) describes the symptoms and criteria for each of these disorders (more than 80 in the last edition).

In order to understand a disorder, one must build a model of how the human mind reacts in front of some stimuli. For example, depressed people tend to ruminate, *i. e.* remember bad experiences of the past, instead of focusing on possible solutions. An hypothesis can be that depressed people tend to focus on bad feelings. Joormann and Gotlib (2008) proved it using a simple experiment. Two lists of words were shown to individuals. They were then asked to forget one of the lists. After that, they were given a word and they had to tell if the word was present in the remaining list. Researchers observed that depressed people tended to remember the words of the second list if it conveyed bad feelings.

A model proposes a potential process but it needs to be tested and validated, for example using functional neuroimaging to pinpoint brain processes.

2 Functional magnetic resonance imaging

Several types of functional imaging techniques have been developed. Electroencephalography and magneto-encephalography measure the superficial cortical neural activity of the brain with a high temporal resolution. Functional magnetic resonance imaging (fMRI) uses strong magnetic fields to measure

oxygen flow in the brain that correlates with its activity. Finally, invasive techniques have been developed such as positron emission tomography that relies on a radioactive tracer to track glucose consumption.

2.1 Blood-oxygen-level dependent (BOLD) contrast imaging

When a brain area is solicited, the brain fires chemical signals to provide it with oxygen and sugar. Nearby blood capillaries dilate to increase the quantity of flowing blood and provide these resources. This phenomenon is called the haemodynamic response.

As a result, we expect a higher concentration of oxygenated hemoglobin in a given brain area soon after its activation. Using NMR, it is possible to differentiate oxygenated hemoglobin from desoxygenated hemoglobin and to measure their concentration: this is called the BOLD signal. By studying this signal, we are able to determine which brain areas are activated at a given time.

This method is subject to several physical and physiological noises. First, some artifacts may be induced by radio transmitters or other equipments. Then, spurious activations are naturally introduced by the blood vessels present in the brain, heart beats and breathing movements. Finally, the brain can be shifted if the subject makes large movements in the scanner.

2.2 Preparation and analysis of fMRI data

Raw fMRI images are not interpretable with bare eyes. In particular because we are interested in small signal co-variation between voxels and not by the values themselves. Human eye, however, is good at perceiving global artifacts in the data such as movements, ghost or scanner coils. Quality assessment of preprocessed fMRI data is done by eye and by relying on dedicated medical imaging software.. In order to prepare the data for further statistical analysis, some preprocessing steps are required.

Data acquisition The resolution of fMRI is usually between $1mm^3$ and $3mm^3$. In a single 3D scan, the brain represents 50 000 to 300 000 voxels[1]. A run contains usually from 100 to 1000 scans. Functional MRI scans are acquired by slices, usually in the axial direction. The time required to acquire one slice is called echo time (TE) and is in the order of tens of milliseconds. The time required to acquire a whole 3D volume is called repetition time (TR) and is in the order of seconds. Typical values for a 3D volume of 60 slices are TE=33ms and TR=2s for a 3 Tesla scanner.

[1] Voxel stands for volume element. It refers to a point in a 3D image, just as pixel refers to a point in a 2D image.

Motion correction. Head movement has a big impact on fMRI. A movement with an amplitude higher than the voxel resolution (*i. e.* 2 to 3 millimeters) can shift the signal of the entire brain. Moreover, the worst impract of motion is inflow effects, *i. e.* artefactual signals. In the scanner, the head of the subject is

fixed using cushion pads to avoid movements and the subject is asked to stay as still as possible. Yet, it is impossible to completely avoid head movement. In order to remove the effect of movement, the 3D scans are realigned on a reference scan – usually the one in the middle of the sequence – using rigid body transformation (translation and rotation, without change of scale).

Slice timing correction As stated before, brain slices are not acquired at the same time. This introduces a shift in the haemodynamic response associated to each of them. This problem can be solved by interpolating the signal of each slice so that all of them can be considered as acquired at the same time. This step is optional as, in practice, it does not bring significant improvement. In a preliminary experiment performed on 50 individuals, we did not observe differences in term of prediction score with or without slice timing correction.

Normalization to a reference template. Each brain is of different size and shape. In order to compare brain activations across several individuals, we need to normalize them by registration to a common template. This template can be a reference template used in the community (MNI[2] for example). It is also possible to compute a template directly from the data. Once a template is chosen, for each subject, we perform two successive normalizations. First, the anatomical scan acquired in the subject is registered to the MNI template. Then, the fMRI data are registered to the anatomical scan. After that, the two transformation matrices are combined in order to normalize the fMRI data to the template.

[2] MNI stands for Montreal Neurological Institute. It is also the name of a reference anatomical template used for fMRI registration

Analysis Forward inference made on fMRI data (*e.g.* prediction of brain activation from the stimuli) can be conceptualized as the encoding of perceptual, motor or cognitive parameters into brain signals. The inverse model, that predicts behavioral data from brain activation is called *decoding*. This thesis focusses on the decoding of brain signals. Two paradigms allow to study brain signals. Either we study them in controlled condition on a particular task, this is the task paradigm, or we study the spontaneous activity of the brain in order to uncover its organization, this is the resting-state paradigm.

2.3 The task paradigm

Using an experimental design, it is possible to relate the BOLD signal with specific tasks performed by the subject. For example, a sound can be played in the left or the right ear of the subject. By comparing brain activation between resting state and when the sound is playing, we can isolate the auditory cortex of the brain.

Statistically, we do that by crafting a design matrix corresponding to the experiment: one column of the matrix represents one particular activation.

Columns corresponding to known artifacts of the BOLD signal, such as heart beats or movements, can be added in the design matrix in order to regress out the part of the signal related to them. We then use a general linear model to recover the brain maps corresponding to each of the columns in the design matrix. These are called β-maps:

$$y = X\beta + \varepsilon$$

with y the BOLD signal, X the design matrix, β the activation maps and ε the noise of the model.

However, task-based paradigms require both a careful experimental setup to control for psychological confounds in the fulfillment of the task, and the active and focused participation of the subject. In clinical settings, when working with impaired patients, these constraints have limited the usefulness of these tools.

For example, stroke patients suffer from a wide range of symptoms depending on the brain region affected by stroke. This disorder is hard to study in a task-based paradigm since the patient's problem is not to be able to perform some tasks. However, detecting an abnormal brain connectivity pattern is possible using resting-state fMRI (Varoquaux et al., 2010a).

2.4 Resting-state fMRI

Resting state fMRI (or rest fMRI) uses the same acquisition method as task fMRI. However, instead of giving a particular task to the subject, he or she is asked to let his mind wander without sleeping. By studying this background activity of the brain, it is possible to uncover its underlying organization (Raichle, 2010). Depending on the protocol, the subject can be asked to keep his eyes closed or to look at a fixation cross. The fixation cross prevents random eye movements and helps the subject not to sleep.

In resting-state fMRI, we do not study the signal of each voxel itself but the interactions between the brain voxels. In particular, we study the functional connectivity of the brain, *i.e.* the similarity of activation patterns between brain regions that share a common functional role. Since there is no design matrix in rest fMRI, one must be careful to properly regress out physiological noises or spurious correlations may appear between brain regions, in particular longitudinally (Power et al., 2012; Van Dijk et al., 2012).

A first approach of functional connectivity is the *voxel to voxel* approach in which the similarity is measured between each pair of voxels. This method is not only computationally expensive, given the number of voxels in the brain, but it is also unfounded from the statistical standpoint: it requires the estimation of millions of parameters, much more than the number of observations supports. As a consequence, some form of dimensionality reduction – a feature selection or extraction – is necessary to study connectivity.

Measure of symmetry. One way to simplify the problem of functional connectivity is to focus on one particular aspect of brain connectivity. An hypothesis commonly made on brain organization is its symmetry. In fact, several brain networks extracted from the brain using a task-driven approach are symmetric (auditory cortex, motor cortex, etc.). Voxel-Mirrored Homotopic Connectivity (VMHC, (Zuo et al., 2010)) is a measure of the symmetry of brain activation. It consists in measuring the similarity between each voxel in the left hemisphere and its counterpart in the right hemisphere. It has been proven to decrease in cocaine addicted individuals (Kelly et al., 2011). This feature selection approach is one way to make the functional connectivity problem tractable.

Nearest neighbors. A simple dimensionality reduction approach consists in considering the connectivity between one voxel and its nearest neighbors. The common approach is to consider a sphere of a given radius around the voxel. Region homogeneity (ReHo, (Zang et al., 2004)) measures the homogeneity of brain regions by measuring the similarity of activation patterns between one voxel and its neighbors. Originally applied in a task paradigm, it has also been applied to resting state fMRI to extract biomarkers for schizophrenia (Liu et al., 2006), ADHD (Cao et al., 2006), Alzheimer's disease (Liu et al., 2008), depression (Yao et al., 2009), and ASD (Paakki et al., 2010). A similar approach is used in the Searchlight approach (Kriegeskorte et al., 2006) where each voxels receive a score corresponding to the prediction accuracy obtained by the voxels and its neighbors for a given problem.

Voxels aggregation. A common dimensionality reduction approach used in machine learning consists in reducing the dimensionality of the data by aggregating similar features. Assuming that voxels with similar activation patterns share the same functional role, we group similar voxels into components called brain networks. The most popular method for this task is spatial group independent component analysis (ICA). After dimensionality reduction, one time series is computed for each region and the correlation inside or between brain networks is used to build neurophenotypes.

3 Brain atlas extraction

Here, we describe brain atlases that are considered as reference atlases in the community. The first anatomical atlases were established long before the discovery of functional neuroimaging. However, they are still widely used in the functional connectivity community as we expect the brain functional organization to follow anatomical features. Some functional atlases have also become references in the domain. Finally, several data-driven approaches exist to estimate an atlas direclty from the data.

Anatomical atlases. Brodmann (1909) established the oldest brain atlas still used today (See Figure 2.1 left). By using a dye that colors cells' RNA, he colored several brains (human and animal) and reported the observed segmentation of the cortex. The functional role of some of these areas was discovered and this atlas remains a reference today. The automated anatomical labeling (AAL, Tzourio-Mazoyer et al. (2002), Figure 2.1 middle) is the name of a procedure in the SPM software to automatically label brain regions. It is based on a brain segmentation performed on the T1 image of a single subject coming from the Montreal Neurological Institute. The sulci of the brain have been delineated manually and then closed to form brain regions.

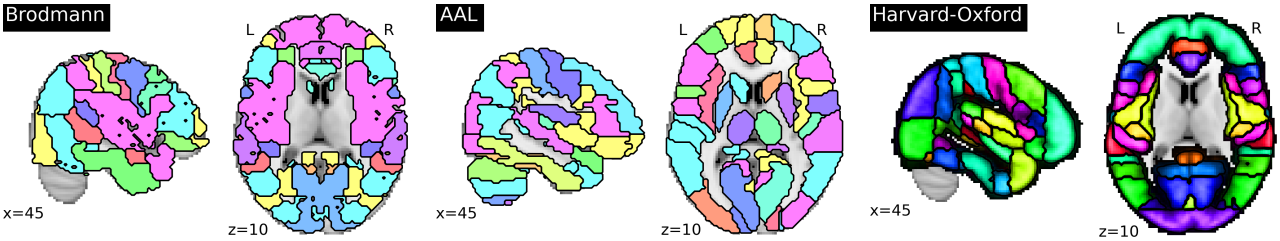


Figure 2.1: **Reference anatomical atlases. Left:**The Brodmann atlas (Brodmann, 1909). **Middle:** The automated anatomical labeling atlas. **Right:** The Harvard-Oxford atlas.

Finally, Harvard-Oxford (HO, Desikan et al. (2006), Figure 2.1 right) is an atlas obtained using a semi-automated tool on 37 subjects T1 scans and based on brain gyri. It is composed of 69 regions and is also aligned on MNI space.

Functional atlases. Smith et al. (2009) (Figure 2.2 left) uses BrainMap task data and an independent component analysis run on resting-state data to provide three atlases of different size for each paradigm (10, 20 and 70 brain networks). Yeo et al. (2011) (Figure 2.2 middle) uses a clustering approach to segment the brain cortical surface of more than 1000 subjects. The cortical surface is divided into 1175 regions of interest and each voxel is then added to the closest region as stated in Lashkari et al. (2010). The data are then aggregated into two atlases: one having 7 brain networks and one having 17. Craddock et al. (2012) (Figure 2.2 right) uses a spectral clustering approach on the voxel-to-voxel affinity matrices of 41 subjects to extract brain regions. Atlases of increasing sizes, from 10 to 400 clusters, are available.

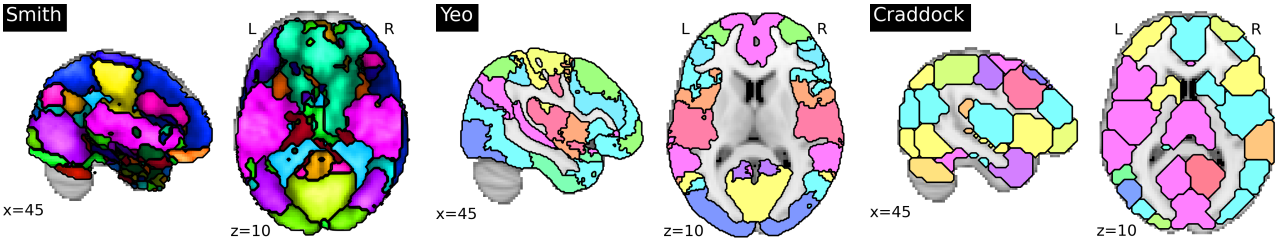


Figure 2.2: **Reference functional atlases. Left:**The Smith 2009 resting-state network atlas. **Middle:** Yeo 2011 atlas. **Right:** Craddock 2012 atlas.

4 The Autism Brain Imaging Data Exchange (ABIDE) dataset

4.1 Understanding autism spectrum disorders

Autism in society. According to the centers for disease control and prevention (CDC), autism spectrum disorders (ASD) are the fastest-growing developmental disorder[3]. Prevalence increased from 6.7‰to 14.7‰between 2000 and 2010. It affects boys five times more than girls. Diagnosis has a great impact on the patient's life and society. In fact, special education and behavioral therapies are provided to autistic children along with residential care during adulthood. The cost of these services is estimated to $2.4 million per patient (Buescher et al., 2014), most of it during adulthood ($0.6 million during childhood, $1.8 million during adulthood). Autism services cost $236 to $262 billion annually to US government. Diagnosis is decisive since early diagnosis can reduce lifelong cost of about two third (Järbrink, 2007).

[3] http://www.cdc.gov/ncbddd/autism/data.html

Symptoms and Diagnosis Discovered in 1943 (Kanner et al., 1943), autism spectrum disorders (ASD) were originally defined by symptoms affecting three cognitive aspects: *i)* impairment in language learning *ii)* impairment in social interactions and/or *iii)* repetitive behaviors, as stated in the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV-TR, see Figure 2.3) (Association et al., 2000). Constant efforts were put into understanding this spectrum of disorders (Wing and Gould, 1979; Baron-Cohen et al., 1985; Lord et al., 2000), but the limit between healthy and autistic individuals remains fuzzy and controversial. If standard tests (such as ADOS, ADI or SRS) have been developed to quantify the severity of each symptom, in the end, the diagnosis is left to the psychiatrist or psychologist.

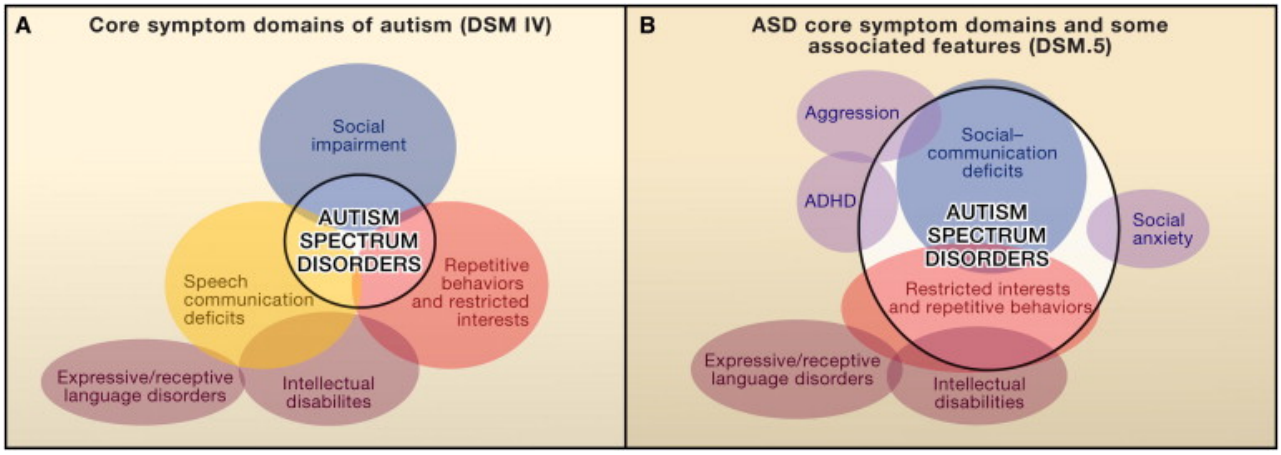


Figure 2.3: **Autism spectrum disorders symptoms as defined in the diagnostic and statistical manual of mental disorders. Left:** fourth edition. **Right:** fifth edition. Figure from (Lord, 2011).

Recent findings dissociated language disorders of ASD. For example, using diffusion tensor imaging, (Verhoeven et al., 2012) showed abnormal microstructural organization in Broca's and Wernicke's area in both ASD and healthy participants. This led to a new definition of ASD as presented in new DSM-V

(Association et al., 2013) (see Figure 2.3) which removed the language aspect from ASD and conserved only two symptoms (social interaction and behavioral problems). As a consequence, subtypes of ASD (like Asperger's syndrom that does not imply language problems) have been merged in the global notion of autism spectrum disorders. In its current definition, Autism Spectrum Disorder is characterized by early emerging persistent social communication and interaction deficits alongside restricted and repetitive behaviour and/or interests

Standard tests. Standardized tests are used both to diagnose autism and to follow the progress of a subject during therapy. The autism diagnostic observation schedule (ADOS, (Lord et al., 1989)) is the gold standard in term of ASD. It consists of a semi-structured assessment of communication, social interaction, and play (or imaginative use of materials) for individuals suspected of having autism or other pervasive developmental disorders. Based on these interactions, scores are attributed for social and language skills of the patient. However, ADOS is not well suited for children that cannot speak, this is why it is usually coupled with autism diagnostic interview (ADI, (Le Couteur et al., 1989)) that consists of a structured interview with the parents of the patients. Unlike ADOS, ADI (or ADI revised) takes the patient's history into account. The social responsiveness scale (SRS, (Constantino and Gruber, 2002)) measures social skills of the patient through an interview given to the child's parents and teachers. Other general tests can also be used to quantify ASD symptoms: Vineland adaptive behavior scale (Sparrow et al., 1989) measures the ability of individuals with developmental disorders to handle everyday tasks, and the Wechsler intelligence scale for children (Wechsler, 1949) measures the cognitive abilites of an individual. Regular IQ tests can also be used.

Why is diagnosing ASD difficult? Early diagnostic of ASD is crucial. Studies have shown that they can be reliably diagnosed at the age of 2 (Lord, 1995). However, diagnostic must be made by a trained psychiatrist with experience in this kind of diagnosis (Stone et al., 1999). Autism tests based on the observed behavior of the child by his parents (such as ADI) are less reliable since parents may not be qualified to spot specific behaviors proper to this disease and shouldn't be used alone. The main difficulty in diagnosing ASD is that it covers a wide range of symptoms as depicted by DSM classification and by the variety of tests used to measure its severity.

Reliable diagnosis of autism is a first step towards its comprehension. The hope is that neuroimaging may help us understand this disease and its multifactorial causes. A possible application could be to find the treatment best suited for each individual (Howard et al., 2005).

4.2 Using neuroimaging to diagnose ASD

The neuroanatomy of autism. The neuroanatomy of ASD has already been studied at different levels, usually on small single-site datasets. However no consensus has been reached yet. At the brain level, an abnormal growth of the brain in early childhood has been observed (Courchesne et al., 2001; Amaral et al., 2008) but not for all children (Hazlett et al., 2012). However, whether this effect concerns all tissues is still subject to debate as (Herbert et al., 2003) locates this growth in the white matter exclusively while (Lotspeich et al., 2004; Palmen et al., 2005) locate it in the grey matter. Several studies reports a reduced size of corpus callosum, along with structural hypoconnectivity, in small ASD studies (Egaas et al., 1995; Piven et al., 1997; Just et al., 2007; Alexander et al., 2007). If a recent study confirms these findings in ABIDE (Di Martino et al., 2014), another one claims that these differences are not significant (Lefebvre et al., 2015). Cortical density differences are also subject to debate. For example, (Boddaert et al., 2004) observed a lower cortical density in posterior superior temporal sulcus (pSTS) on 21 individuals, while (Haar et al., 2014) reports a higher density in the same area on the ABIDE dataset.

Finally, at neuron level, *Boddaert et al.*(Boddaert et al., 2009) suggested that ASD may be caused by a problem in myelination of axons caused by the abnormal growth of the brain while *Casanova et al.*(Casanova et al., 2006) explains ASD by perturbations in the columnar structure of neocortex. If the community is slowly converging, a lot of conclusions are still subject to debate which shows how hard it is to understand ASD.

Resting-state functional connectivity. Rest fMRI applied to autism is still at an early stage due to the lack of large datasets. Dysfunctions in the Default Mode Network and regions linked to language and emotions have been reported as well as non-focal effects such as increased lateralization and global hypoconnectivity. ASD are particularly challenging disorders to study, as they encompass a whole spectrum of disorders with various causes. As such it can really benefit from multi-site studies.

Using the ABIDE dataset, Di Martino *et al.* (Di Martino et al., 2014) exhibit significant functional-connectivity differences between ASD and typical controls. A global hypoconnectivity is observed, particularly in the temporal lobe, while hyperconnectivity is only observed between subcortical and primary parietal sensorimotor areas. In a first predictive-modeling study on the ABIDE dataset, Nielsen et al. (2013) obtained a 60% classification accuracy on a subset of 964 individuals. Brain regions giving best accuracy were related to language or attention: parahippocampal and fusiform gyri, insula, medial prefrontal cortex, posterior cingulate cortex, and Wernicke Area. The implication of the default mode network, but also language related ROIs (such as Broca's area and superior temporal sulcus) along with attentional network have been confirmed in several other studies listed in figure 6.1.

The ABIDE dataset. ABIDE is a dataset composed of 539 patients suffering from Autism Spectrum Disorders (ASD) and 573 typical controls gathered from 16 international sites. It features one rest fMRI session per subject, acquired on different 3T scanners with TR ranging from 1500ms to 3000ms. Subjects were asked to let their mind wander without falling asleep, eyes-closed or with a fixation cross, depending on the acquisition site.

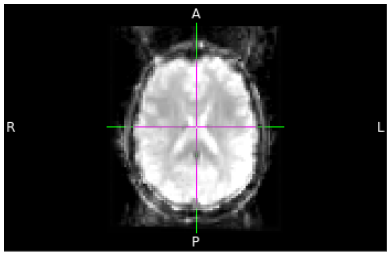
We use the ABIDE dataset preprocessed by Preprocessed Connectomes Project[4] based on the CPAC pipeline. Preprocessing includes slice timing correction, motion correction and intensity normalization. Preprocessed data were validated using normative metrics on anatomical and functional scans such as entropy focus criterion, smoothness of voxels, measure of fractional displacement, etc. Four human experts also performed visual inspection of the data and excluded any acquisition with missing part of brain, high movement peaks, ghosting artifacts or scanner coils (see Figure 2.4). From the 1112 initial participants, 871 were kept after quality checking (QC). We also performed detrending and standardization of the time series before running the analysis.

Prediction task on ABIDE. In this manuscript, we present a new brain atlas extraction method – chapters 3, 4 and 5 – along with a new pipeline to predict the condition of individuals in the ABIDE dataset – chapters 6 and 7. ABIDE is a post-hoc aggregation of data from different sites. As such, the assessments used to measure autism severity vary across sites. As a consequence, autism severity scores are not directly quantitatively comparable between sites. The diagnostic of ASD, on the other, is reliable. This is why we chose it as prediction task.

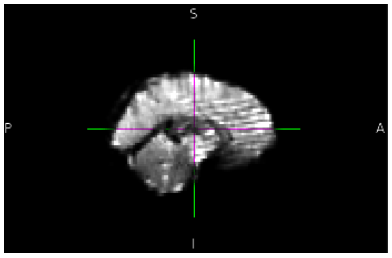
The data-processing pipeline used for prediction is described in chapter 6. In addition to the classical task of diagnosing individuals coming from a known site, we perform the task of diagnosing individuals coming from an unknown site, *i.e.* an unknown environment. For this purpose, we leave a whole site out as testing set. We call it inter-site prediction.

We take advantage of the number of acquisition sites in ABIDE by introducing a new cross-validation scheme, leave-one-site-out, that is closer to clinical setting.

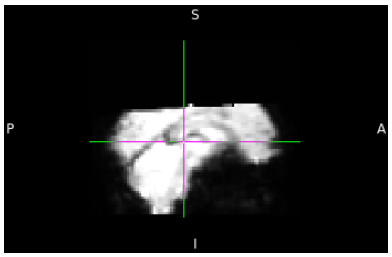
[4] http://preprocessed-connectomes-project.github.io/abide/



Ghost artifact (Caltech 51480): at the top and the bottom of the brain, a pale replication of the brain is visible, introducing artifacts in the signal of some voxels. Contrast has been increased for this example.



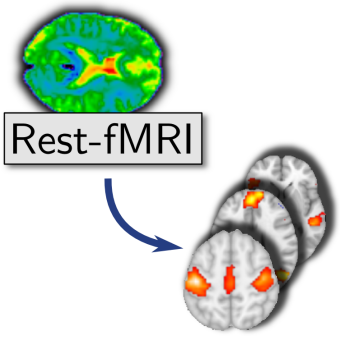
Scanner coil (UCLA 51294): stripes caused by acquisition process are visible. Contrast has been increased for this example.



Brain cut (Leuven 50727): A part of the brain is missing.

Figure 2.4: Artifacts encountered in ABIDE dataset

3 | Total-Variation Multi-Subject Dictionary Learning



In this section, we present a new method to extract functional atlases. It uses a multi-subject dictionary learning approach to scale to large datasets and integrate a structured regularization to improve the estimation the brain components. We combine total variation and sparse regularization, and show that it extracts spatially connected blobs that are plausible regarding to prior knowledge in the neuroscientific field. To validate our method, we compare the extracted atlases to the reference method used for this task, group ICA, and show that our approach is better at recovering anaomical structures in brain areas highly affected by noise.

The work presented in this chapter has been published in:

Extracting brain regions from rest fMRI with Total-Variation constrained dictionary learning, *Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, Gaël Varoquaux*, MICCAI - 16th International Conference on Medical Image Computing and Computer Assisted Intervention - 2013, Sep 2013, Nagoya, Japan. Springer, 2013

1 Introduction – atlas estimation methods

The covariance structure of functional networks, observed at rest using functional Magnetic Resonance Imaging (fMRI) signals, is a promising source of diagnostic or prognostic biomarkers, as it can be measured on impaired subjects, such as stroke patients (Varoquaux et al., 2010a). However, as stated in section 2.4 of chapter 2, the statistical analysis of this structure requires the choice of a reduced set of brain regions (Varoquaux and Thirion, 2014). These should *i)* cover the main resting-state networks (Beckmann and Smith, 2004; Yeo et al., 2011); *ii)* give a faithful representation of the original signal, *e.g.* in

the sense of compression or explained variance; *iii)* be defined in a way that is resilient to inter-subject variability.

References atlases, such as AAL (Tzourio-Mazoyer et al., 2002) or Harvard-Oxford (Desikan et al., 2006), are composed of brain regions extracted from anatomical features of the brain. But brain regions can also be learned on rest fMRI data. In particular, unsupervised learning methods can decompose a 4d signal, such as rest fMRI, into a set of spatial structures and their corresponding time series. These spatial structures are spatial brain maps composed of brain functional units.

Independent Component Analysis (ICA) is the reference method to extract spatial maps from rest fMRI (Beckmann and Smith, 2004) but the temporal independence imposed on the timeseries has been recently questioned (Daubechies et al., 2009). Promising developments rely on penalized dictionary learning to output more contrasted maps (Varoquaux et al., 2011). However, while the maps highlight salient localized features, post-processing is required to extract connected regions. (Kiviniemi et al., 2009) use ICA maps to manually define this parcellation from resting-state networks. A complementary approach is to rely on voxel clustering that creates hard assignments rather than continuous maps (Yeo et al., 2011; Blumensath et al., 2012; Thirion et al., 2014).

Here we bridge the gap between the two strategies by introducing a structured penalization, common to clustering approaches, to the dictionary learning. The main contributions are *i)* the adaptation of dictionary learning to produce well-formed brain regions and *ii)* the computational improvement to the corresponding estimation procedures.

2 A dictionary learning approach to segment regions

Dictionary learning belongs to the family of matrix factorization methods, along with independent component analysis (ICA). These unsupervised methods rely on the same principle of matrix decomposition but apply different contraints on the estimated matrices.

Matrix factorization methods are routinely used to extract structured spatial patterns from resting-state fMRI data. These patterns are then interpreted in terms of functional networks or regions. *Spatial Group Independent Component Analysis* (Group ICA) is the most popular method to process resting-state fMRI. It is based on a linear mixing model to separate different signals preceded by a principal component analysis (PCA) to reject noise (Beckmann and Smith, 2004). In the following section, we describe how dictionary learning can be used for the same purpose and extended to multi-subject problems.

2.1 Preliminary: Dictionary Learning

A dictionary is a set of k basis functions, called dictionary elements, that can be used to decompose a signal. A typical example is wavelets. In the following

work, we see how to build a dictionary $\mathbf{V}$ of spatial components that can be used to decompose resting-state brain signals. This approach tends to be robust to noise because white noise cannot be represented by a sparse combination of elements of the dictionary.

We first formalize the problem of matrix decomposition. The decomposition of $\mathbf{Y} \in \mathbb{R}^{n \times p}$ in k components is written:

$$\mathbf{Y} = \mathbf{U}\mathbf{V}^\mathsf{T} + \mathbf{E}, \quad \mathbf{E} \sim \mathcal{N}(0, \sigma\mathbf{I})$$

with $\mathbf{V} \in \mathbb{R}^{p \times k}$ being the dictionary, $\mathbf{U} \in \mathbb{R}^{n \times k}$ the associated components and $\mathbf{E}$ the residuals of the model which are modeled as white Gaussian noise.

The dictionary learning approach can be used to solve this problem and relies on the alternated minimization of $\mathbf{U}$, in a data fitting step, and $\mathbf{V}$, in a sparse dictionary update step[1].

[1] Note that this formulation of the dictionary learning problem is different from the traditional one encountered in computer vision. In our setting, we want to impose sparsity directly on the dictionary and not on the loadings of the model

Data fit:

$$\hat{\mathbf{U}} = \arg\min_{\mathbf{U}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^\mathsf{T}\|^2, \quad \text{s.t.} \|u_l^s\|_2^2 = 1$$

Sparse dictionary update:

$$\hat{\mathbf{V}} = \arg\min_{\mathbf{V}} \|\mathbf{Y} - \mathbf{U}\mathbf{V}^\mathsf{T}\|^2 + \alpha\Omega(\mathbf{V})$$

with Ω a sparsity-inducing regularization and α its associated weight.

2.2 Multi-Subject Dictionary Learning (MSDL)

(Varoquaux et al., 2011) proposed an extension of the dictionary learning approach to several subjects by introducing a hierarchical probabilistic model. Following the standard dictionary learning model, the n-long time series observed on p voxels for subject s $\mathbf{Y}^s \in \mathbb{R}^{n \times p}$ are written as the linear combination of k subject-specific dictionary elements, that are spatial maps $\mathbf{V}^s \in \mathbb{R}^{p \times k}$. For resting-state brain activity, we model the loadings $\mathbf{U}^s \in \mathbb{R}^{n \times k}$ by specifying their covariance.

$$\forall s \in 1...S,\ \mathbf{Y}^s = \mathbf{U}^s{\mathbf{V}^s}^\mathsf{T} + \mathbf{E}^s, \quad \mathbf{E}^s \sim \mathcal{N}(0, \sigma\mathbf{I}), \quad \mathbf{U}^s \sim \mathcal{N}(0, \mathbf{I}) \tag{3.1}$$

In addition, the subject-specific maps $\mathbf{V}^s$ are generated from population-level latent factors, the spatial patterns written as brain maps $\mathbf{V}$:

$$\forall s \in 1...S,\ \mathbf{V}^s = \mathbf{V} + \mathbf{F}^s, \quad \mathbf{F}^s \sim \mathcal{N}(0, \zeta\mathbf{I}) \tag{3.2}$$

Finally, we specify the prior distribution on $\mathbf{V}$: $\mathcal{P}(\mathbf{V}) \propto \exp(-\xi\Omega(\mathbf{V}))$, where Ω is typically a norm or a quasi-norm[2] We estimate the models presented in Equations 3.1 and 3.2 in a maximum a posteriori framework. The corresponding learning strategy is a minimization problem comprising a subject-level data-fit term, a term controlling subject-to-group differences, and a group-level penalization:

[2] A quasi-norm is a norm for which the triangle inequality is defined as

$$\|x + y\| \leq k(\|x\| + \|y\|)$$

for a given $k > 1$.

$$\underset{\mathbf{U}^s,\mathbf{V}^{s\mathsf{T}},\mathbf{V}}{\arg\min}\ \frac{1}{S}\sum_{s\in S}\frac{1}{2}\left(\left\|\mathbf{Y}^s-\mathbf{U}^s\mathbf{V}^s\right\|_{\text{Fro}}^2+\mu\left\|\mathbf{V}^s-\mathbf{V}\right\|_{\text{Fro}}^2\right)+\mu\,\alpha\,\Omega(\mathbf{V}),$$
$$\text{s.t.}\|u_l^s\|_2^2=1 \quad (3.3)$$

where $\mathbf{V}\in\mathbb{R}^{p\times k}$ is the set of group-level maps and Ω is a convex regularizer. $\mu=\sigma/\zeta$ is a parameter that controls the similarity between subject-level and group-level maps while α sets the amount of regularization enforced on the group-level maps. Note that μ weights the last two terms, this is used further to simplify the resolution of the problem.

This problem is not jointly convex with respect to $\{\mathbf{U}^s\}$, $\{\mathbf{V}^s\}$ and $\mathbf{V}$, but it is separately convex and (Varoquaux et al., 2011) relies on an alternate minimization strategy, optimizing separately (3.3) with regards to $\{\mathbf{U}^s\}$, $\{\mathbf{V}^s\}$ and $\mathbf{V}$ while keeping the other variables fixed.

Minimization with respect to U:

$$\hat{\mathbf{U}}^s=\underset{\mathbf{U}^s}{\arg\min}\left\|\mathbf{Y}^s-\mathbf{U}^s\mathbf{V}^s\right\|_{\text{Fro}}^2,\quad \text{s.t.}\|\mathbf{u}^s\|_2^2=1 \quad (3.4)$$

If we assume that $\mathbf{V^s}$ and $\mathbf{V}$ are fixed, we end up solving (3.4). This problem is solved by an ordinary least squares. To optimize memory usage, we implement it using a block coordinate descent (Jenatton et al., 2010). While optimize the l^{th} line of matrix $\mathbf{U}$, we assume the rest of the matrix fixed:

$$\forall l\in 1...n,\quad \hat{\mathbf{u}}_l=\mathbf{u}_l+\|\mathbf{v}_l\|_2^{-2}(\mathbf{Y}-\mathbf{U}\mathbf{V}^{\mathsf{T}})\mathbf{v}_l \quad (3.5)$$

Minimization with respect to $\mathbf{V^s}$:

$$\hat{\mathbf{V}}^s=\underset{\mathbf{V}^s}{\arg\min}\left\|\mathbf{Y}^s-\mathbf{U}^s\mathbf{V}\right\|_{\text{Fro}}^2+\mu\left\|\mathbf{V}^s-\mathbf{V}\right\|_{\text{Fro}}^2 \quad (3.6)$$

If $\mathbf{U^s}$ and $\mathbf{V}$ are fixed, we end up solving (3.6). This problem is solved by applying a ridge regression on the variable $\mathbf{V_s}-\mathbf{V}$.

$$\hat{\mathbf{V}}^s=\mathbf{V}+(\mathbf{Y}^s-\mathbf{U}^s\mathbf{V}^{s\mathsf{T}})^{\mathsf{T}}\mathbf{U}^s(\mathbf{U}^{s\mathsf{T}}\mathbf{U}^s+\mu\mathbf{I})^{-1} \quad (3.7)$$

Minimization with respect to V:

$$\hat{\mathbf{V}}=\underset{\mathbf{V}}{\arg\min}\ \frac{1}{2}\left\|\mathbf{V}^s-\mathbf{V}\right\|_{\text{Fro}}^2+\alpha\,\Omega(\mathbf{V}) \quad (3.8)$$

If $\mathbf{U^s}$ and $\mathbf{V^s}$ are fixed, we end up solving (3.8). Note that μ is absent in this optimization problem since it weights both terms. This amounts to computing a proximal operator, which can be seen as a denoising operator, or an implicit gradient step in Ω of length α:

$$\forall l\in 1...p,\quad \underset{\alpha\Omega}{\text{prox}}(\mathbf{v}_l)\overset{\text{def}}{=}\underset{\hat{\mathbf{v}}_l}{\arg\min}\ \frac{1}{2}\left\|\mathbf{v}_l-\hat{\mathbf{v}}_l\right\|_2^2+\alpha\,\Omega(\hat{\mathbf{v}}_l) \quad (3.9)$$

The implementation of TV-MSDL provided in (Varoquaux et al., 2011) is shown in listing 3.1.

Listing 3.1: Simple implementation of Multi Subject Dictionary Learning presented in (Varoquaux et al., 2011)

Input:
- $\{\mathbf{Y}^s \in \mathbb{R}^{n \times p}, s = 1, ..., S\}$, the n-long time series observed on p voxels for subject s
- $\mathbf{V}_{init}$, an initial guess for group maps $\mathbf{V}$

Output:
- $\mathbf{V} \in \mathbb{R}^{p \times k}$, the set of group-level maps
- $\{\mathbf{V}^s \in \mathbb{R}^{p \times k}\}$, the subject-specific spatial maps
- $\{\mathbf{U}^s \in \mathbb{R}^{n \times k}\}$, the k associated time series

$E_0 \leftarrow \inf$, $E_1 \leftarrow \inf$, $i \leftarrow 1$
while $E_i - E_{i-1} > \epsilon E_{i-1}$ **do**
  for all s in S **do**
  Update $\mathbf{U}^s$ (block coordinate descent (Jenatton et al., 2010)):
    for $l = 1$ to k:
      $\mathbf{u}_l^s \leftarrow \mathbf{u}_l^s + \|\mathbf{v}_l^s\|_2^{-2}(\mathbf{Y}^s - \mathbf{U}^s{\mathbf{V}^s}^\top)\mathbf{v}_l^s$
      $\mathbf{u}_l^s \leftarrow \mathbf{u}_l^s / \max(\|\mathbf{u}_l^s\|_2, 1)$
    end for
  Update $\mathbf{V}^s$ (ridge regression):
    $\mathbf{V}^s \leftarrow \mathbf{V} + (\mathbf{Y}^s - \mathbf{U}^s\mathbf{V}^\top)^\top \mathbf{U}^s(\mathbf{U}^{s\top}\mathbf{U}^s + \mu\mathbf{I})^{-1}$
  end for
  Update $\mathbf{V}$ (proximal operator):
    $\mathbf{V} \leftarrow \underset{\frac{\lambda}{S}\Omega}{\text{prox}}(\frac{1}{S}\sum_{s=1}^{S}\mathbf{V}^s)$
  Compute new value of energy:
    $E_i \leftarrow \varepsilon(\mathbf{U}^s, \mathbf{V}^s, \mathbf{V})$
  $i \leftarrow i + 1$
end while

3 Sparse TV penalization to enforce compact regions

We want to define a small set of regions that represent well brain-activity signals. Dictionary learning does not produce in itself regions, but continuous maps. Enforcing sparsity, *e.g.* via an ℓ_1 penalty ($\Omega(\mathbf{v}) = \|\mathbf{v}\|_1$) on these maps, implies that they display only a few salient features that may not be grouped spatially. (Varoquaux et al., 2011) use a spatial smoothness prior (ℓ_2 norm of the image gradient) in addition to the sparsity prior to impose spatial structure on the extracted maps. However, while smoothness is beneficial to rejecting very small structures and high-frequency noise, it also smears edges and does not constrain the long-distance organization of the maps.

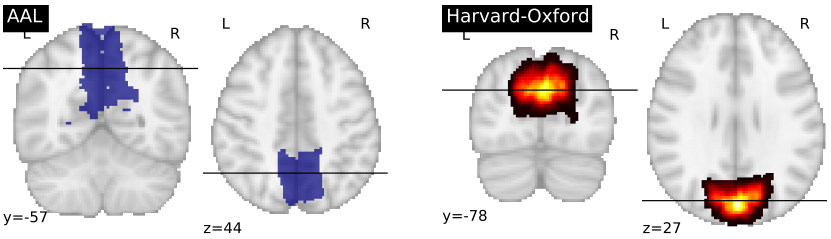


Figure 3.1: **Precuneus in different atlases.** Left: precuneus region in AAL (Automated Anatomical Labeling). Right: precuneus regions in Harvard Oxford probabilistic atlas.

3.1 Spatial properties of functional units

Anatomical atlases, such as AAL (Tzourio-Mazoyer et al., 2002) and Harvard-Oxford (Desikan et al., 2006) are considered as references for neuroimaging studies. This success may be due to some geometrical traits shared by their ROIs: They are small and consist of neighbouring voxels.

Functional units are small. In these atlases, the average number of nonzero voxels in each region is less than 1% of the whole brain volume (AAL: 0.86%, Harvard Oxford: 0.63%). The small size of the components can be enforced by using sparse regularization, usually the ℓ_1 norm. Figure 3.1 shows the precuneus in AAL and Harvard-Oxford, the region is small compared to the rest of the brain.

Functional units are connected. In these atlases, units are represented as connected regions, *i.e.* sets of neighboring voxels. Agglomerative clustering methods, such as Ward's clustering, enforce cluster voxels to be neighbors. The precuneus shown in Figure 3.1, we can see that the voxels composing the precuneus region are all connected.

In conclusion, a good atlas extraction method will promote these traits when extracting brain regions.

3.2 Total Variation

A simple formulation of the segmentation in an energy minimization problem is to penalize total variation (TV) (Rudin et al., 1992; Chambolle et al., 2010) that tends to produce plateaus, i.e. regions of perfectly flat activation. Briefly, the total variation is defined as the norm of the gradient of the image:

$$\mathrm{TV}(\mathbf{v}) = \sum_i \sqrt{(\nabla_x v)_i^2 + (\nabla_y v)_i^2 + (\nabla_z v)_i^2}$$

Considering the image gradient as a linear operator ∇:

$$\mathbf{v} \in \mathbb{R}^p \rightarrow (\nabla_x \mathbf{v}, \nabla_y \mathbf{v}, \nabla_z \mathbf{v}) \in \mathbb{R}^{3p}, \quad \mathrm{TV}(\mathbf{v}) = \|\nabla \mathbf{v}\|_{2,1}$$

where the $\ell_{2,1}$-norm groups (Kowalski, 2009) are the x, y, z gradient components at one voxel position.

Original picture | Ward's clustering | Total-Variation denoising

Figure 3.2: **Effect of Total-Variation denoising on natural image.** *Left:* Original image of Lena. *Middle*: Ward's clustering applied on Lena. *Right*: Total-Variation denoising applied on Lena. A blurring effect is applied on the image but the strong edges are not affected. The picture exhibits large colored blobs as a clustering, but the values and edges are smoother.

3.3 Sparse TV

Going beyond TV regularization, we want to promote regions comprising many voxels, but occupying only a fraction of the full brain volume. For this we combine ℓ_1 regularization with TV (Baldassarre et al., 2012; Gramfort et al., 2013). We define $\tilde{\nabla}_\rho$ as the augmented gradient operator $\mathbb{R}^p \to \mathbb{R}^{4p}$, consisting of a concatenation of the operator ∇ and the scaled identity operator $\rho\,\mathbf{I}$:

$$\mathbf{v} \in \mathbb{R}^p \to \tilde{\nabla}_\rho \mathbf{v} = ((1-\rho)\nabla_x \mathbf{v}, (1-\rho)\nabla_y \mathbf{v}, (1-\rho)\nabla_z \mathbf{v}, \rho \mathbf{v})$$

along with its augmented divergence operator $\tilde{\nabla}_\rho^\mathsf{T}$:

$$\mathbf{Z} \in \mathbb{R}^4 p \to \tilde{\nabla}_\rho^\mathsf{T} \mathbf{Z} = (1-\rho)(\nabla_x^\mathsf{T} \mathbf{z}_1 + \nabla_y^\mathsf{T} \mathbf{z}_2 + \nabla_z^\mathsf{T} \mathbf{z}_3) - \rho \mathbf{z}_4$$

The proximal operator corresponding to the sparse TV is

$$\begin{aligned} \text{prox}(\mathbf{w}, \mathbf{v}) &= \arg\min_{\mathbf{v}} \frac{1}{2}\|\mathbf{w}-\mathbf{v}\|_2^2 + \alpha\left(\|\nabla \mathbf{v}\|_{2,1} + \rho\|\mathbf{v}\|_1\right) \\ &= \arg\min_{\mathbf{v}} \frac{1}{2}\|\mathbf{w}-\mathbf{v}\|_2^2 + \alpha\left\|\tilde{\nabla}_\rho \mathbf{v}\right\|_{2,1} \end{aligned} \tag{3.10}$$

where , and the $\ell_{2,1}$ norm uses an additional set of groups on the new variables.

Note that the structure of the resulting problem is exactly the same as for TV, thus we can rely on the same efficient algorithms (Beck and Teboulle, 2009) to compute the proximal operator. Finally, in an effort to separate as much as possible different features on different components, we impose positivity on the maps. This constraint is reminiscent of non-negative matrix factorization (Lee and Seung, 1999) but also helps removing background noise formed of small but negative coefficients (as in Figure 3.3 extracted from (Varoquaux et al., 2011)). It is enforced using an algorithm for constrained TV (Beck and Teboulle, 2009). The optimality of the solution can be controlled using the dual gap (Boyd and Vandenberghe) defined by the difference between the primal, *i. e.* the proximal operator in our case, and the dual problem. We define the dual problem of the proximal (see (Michel et al., 2011) for the proof):

$$\text{prox}^*(\mathbf{w}, \mathbf{v}) = \arg\max_{\mathbf{v}} \frac{1}{2}\left(\|\mathbf{w}\|_2^2 - \|\mathbf{v}\|_2^2\right) \tag{3.11}$$

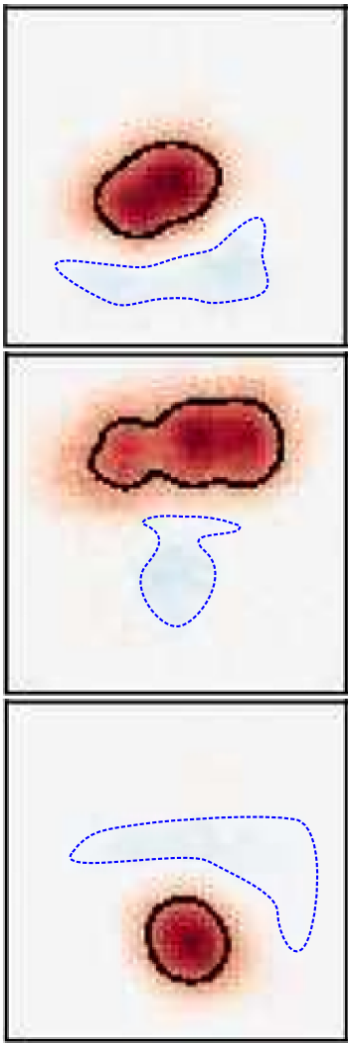


Figure 3.3: **Maps of a toy dataset extracted using MSDL.** The noise composed of negative values is highlighted in blue. Figures from (Varoquaux et al., 2011).

We use it to compute the dual gap:

$$\begin{aligned}\delta_{\text{gap}}(\mathbf{v}, \mathbf{w}) &= \text{prox}(\mathbf{w}, \mathbf{v}) - \text{prox}^*(\mathbf{w}, \mathbf{v}) \\ &= \frac{1}{2}\|\mathbf{w} - \mathbf{v}\|_2^2 + \alpha \left\|\tilde{\nabla}_\rho \mathbf{v}\right\|_{2,1} - \frac{1}{2}\left(\|\mathbf{w}\|_2^2 - \|\mathbf{v}\|_2^2\right)\end{aligned} \tag{3.12}$$

The implementation of the method solving the Sparse TV problem is presented in listing 3.2.

Listing 3.2: Algorithm solving the Sparse TV proximal problem.

Input:
- $\mathbf{V} \in \mathbb{R}^{p \times k}$, the set of group-level maps
- $\{\mathbf{V}^s \in \mathbb{R}^{p \times k}\}$, the average spatial maps over all subjects
- α, the regularization weight
- ρ, the ratio of ℓ_1 norm
- L, a Lipschitz constant of Δf
- $\delta_{\text{gap}} tol$, optional, stopping criterion on dual gap

Output:
- $\mathbf{V} \in \mathbb{R}^{p \times k}$, the set of group-level maps

$\mathbf{V}_0 \leftarrow \frac{1}{S}\sum_{s=1}^{S} \mathbf{V}^s$, $\mathbf{Z}_0 \leftarrow 0$, $t_0 \leftarrow 1$, $i \leftarrow 1$, $i \leftarrow 1$

while $\delta_{\text{gap}}(\mathbf{V}_0, \mathbf{V}_i) > \delta_{\text{gap}} tol$:
- $\mathbf{Z} \leftarrow \mathbf{Z}_{i-1} + \frac{1}{L\alpha}\tilde{\nabla}_\rho(\mathbf{W}_i)$
- $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4) \leftarrow \mathbf{Z}$
- Dual projection:
  - $\mathbf{Z}_i \leftarrow (\text{Proj}_{\ell_{2,\infty}}(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3), \text{Proj}_{\ell_\infty}(\mathbf{z}_4))$
- $t_i \leftarrow \frac{t_{i-1} + \sqrt{1 + 4t_{i-1}^2}}{2}$
- Accelerated proximal gradient:
  - $\mathbf{Z}_i \leftarrow \mathbf{Z}_i + \frac{t_{i-1} - 1}{t_i}(\mathbf{Z}_i - \mathbf{Z}_{i-1})$
- $\mathbf{V}_i \leftarrow \mathbf{V} - \alpha\tilde{\nabla}_\rho^\top \mathbf{Z}_i$
- i ← i + 1

end while

4 Computational optimization

The naive implementation presented above requires loading the whole dataset into memory. With the advent of bigger and bigger datasets in neuroimaging, atlas extraction methods must be able to process datasets containing more than 1000 subjects, *i.e.* several hundreds of gigabytes. In this section, we present several improvements in order to make MSDL scale to big data.

4.1 Stochastic coordinate descent

The algorithm outlined in (Varoquaux et al., 2011) to minimize (3.3) is an alternate minimization using a cyclic block coordinate descent. The time

required to update the $\{\mathbf{U}^s, \mathbf{V}^s\}$ parameters grows linearly with the number of subjects, and becomes prohibitive for large populations. For this reason, rather than a cyclic choice of coordinates to update, we alternate between selecting a random subset of subjects to update $\{\mathbf{U}^s, \mathbf{V}^s\}$ and updating $\mathbf{V}$. This stochastic coordinate descent (SCD) strategy draws from the hypothesis that subjects are similar and a subset brings enough representative information to improve the group-level maps $\mathbf{V}$ while bringing the computational cost of an iteration of the outer loop of the algorithm down. More formally, the justification of this strategy is similar to the stochastic gradient descent approaches: the loss term in (3.3) is a mean of subject-level term (Bottou, 2004) over the group; we are interested in minimizing the expectation of this term over the population and, for this purpose, we can replace the mean by another unbiased estimator quicker to compute, the subsampled mean.

In order to start the optimization with reasonable values, we do the first iteration on all the subjects. Similarly, to obtain a result optimized with respect to all the training set in the end, the last iteration is run on all the subjects. In order to avoid optimizing relatively to the same subjects at each iteration, one subject cannot be selected for two consecutive iterations [3].

[3] The importance of this precaution depends on the setting of the experiment. In a setting where half of the dataset is selected at each iteration, some subjects may be selected at every iteration while other others may never be. We can also imagine the worst case where the same subjects are selected at every iterations: We end up fitting the model on a subsample of our dataset, which is not acceptable.

4.2 Optimization of the proximal problem

Masking. The computation of spatial regularization, whether it be with Smooth- Lasso[4] or TV penalization, implies computing spatial gradients of the images. However, in fMRI, it is most often necessary to restrict the analysis to a mask of the brain: out-of-brain volumes contain structured noise due *e.g.* to scanner artifacts. This masking imposes to work with an operator ∇ that has no simple expression. This is detrimental to computational efficiency because *i)* the computation of the proximal operator has to cater for border effects with the gradient for voxels on the edge of the mask –see *e.g.* (Michel et al., 2011) for more details– *ii)* applying ∇ and ∇^T imposes inefficient random access to the memory while computing gradients on rectangular image-shaped data can be done very efficiently. For these reasons, we embed the masked maps $\mathbf{v}$ into "unmasked" rectangular maps, on which the computation of the proximal term is fast: $\mathbf{M}^{-1}(\mathbf{v})$, where $\mathbf{M}$ is the masking operator. In practice, this amounts to using $\mathbf{M}(\mathbf{z})$ with $\mathbf{z} = \text{prox}_\Omega(\mathbf{M}^{-1}(\mathbf{w})))$ when computing $\text{prox}_\Omega(\mathbf{w})$, and correcting the energy with the norm of $\mathbf{z}$ outside of the mask. Indeed, in the expression of the proximal operator (3.9), $\|\mathbf{M}^{-1}(\mathbf{w}) - \mathbf{z}\|^2 = \|\mathbf{w} - \mathbf{M}(\mathbf{z})\|^2 + \|\mathbf{M}^{-1}(\mathbf{M}(\mathbf{z})) - \mathbf{z}\|^2$ where the first term is the term in the energy (3.3) and the second term is the correction factor that does not affect the remainder of the optimization problem (3.3).

[4] As a reminder, Smooth-Lasso is the Lasso term with an ℓ_2 penalty on the gradient:

$$\Omega_{\text{SL}} = \|\mathbf{v}\|_1 + \frac{1}{2}\mathbf{v}^\mathsf{T}\mathbf{L}\mathbf{v}$$

with $\mathbf{L}$ the Laplacian operator defined on the 3D grid of voxels as defined in (Varoquaux et al., 2011)

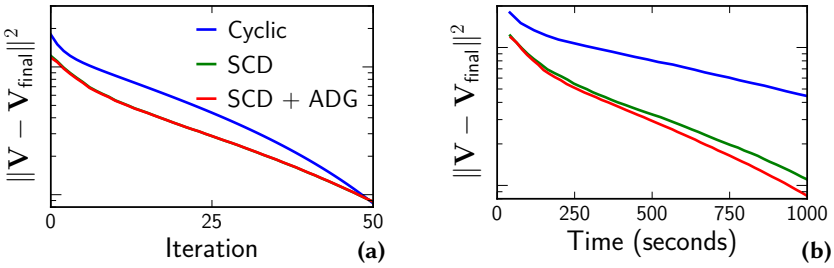
Adaptive dual gap. We use the fact that in an alternate optimization it is not always necessary to optimize to a very small tolerance all the terms for each execution of the outer loop. In particular, the final steps of convergence of

TV-based problems can be very slow. The dual-gap (3.12) gives an upper bound of the distance of the objective to the optimal. We introduce an adaptive dual gap (ADG) strategy: at each iteration of the alternate optimization algorithm, we record how much the energy was decreased by optimizing on $\{\mathbf{U}^s, \mathbf{V}^s\}$ and stop the optimization of the proximal operator when the dual gap reaches a third of this value. As shown in Figure 3.4, we found this to be much faster than fixing an accuracy from the beginning.

4.3 Implementation optimization

Parallelization. Since subject data $\mathbf{U}^s$ and $\mathbf{V}^s$ are optimized independently in each subject, it is possible to run the *for* loop of the procedure in parallel.Subject wise optimization being the most time-consuming step of the algorithm (see Figure 3.4), parallelization can speed up the computation by several orders of magnitude. We also parallelize the optimization of the proximal operator on each map of the group atlas $\mathbf{V}$ independantly.

Out-of-core computation. When a dataset does not fit in memory, we have to store it in an external memory (typically the hard drive) and load chunks of data when we need them. In the particular setting of multi-subject experiments, the data is naturally chunked by subjects in the external memory. Since we process subjects iteratively in MSDL, this storage is ideal. For this purpose, we introduce two functions that will be used to perform out-of-core computation: *load* takes a subject id as argument and load its data from the external memory and *dispose* discards the data from memory. The drawback of this method is that it increases I/O consumption since the whole dataset must be read from the external memory at each iteration.



Time per iteration

Update	$\{\mathbf{U}^s, \mathbf{V}^s\}$	$\mathbf{V}$
Cyclic	37.3s	7.6s
SCD	11.7s	8.4s
SCD + ADG	11.3s	4.0s

(c)

Figure 3.4: Comparing different optimization strategies: cyclic block coordinate descent, as proposed by (Varoquaux et al., 2011), stochastic coordinate descent (SCD), and SCD with adaptive dual gap (ADG) on the proximal term. **(a)** distance to the optimal $\mathbf{V}$ (in log scale) as a function of the number of iterations, **(b)** distance to the optimal $\mathbf{V}$ (in log scale) as a function of time, **(c)** time spent per iteration to update $\{\mathbf{U}^s, \mathbf{V}^s\}$ or $\mathbf{V}$.

4.4 Optimized algorithm

The drawback of out-of-core computation is that I/O operations can be costly in terms of time. By combining it with stochastic coordinate descent, we drastically reduce this overhead and are still able to process big data. The combination of these two techniques is the key to process big datasets in reasonable time. For the stochastic coordinate descent, we choose to load approximately 20%[5] of the dataset at each iteration, based on preliminary experiments. Since subject optimization is realized in parallel, it is best to

[5] This value has been validated empirically on various datasets as a trade-off between convergence speed and memory usage.

choose a number of subjects that is a multiple of the number of CPU cores.
An optimized version of the algorithm is presented in Listing 3.3.

Listing 3.3: Optimized TV-MSDL algorithm.

Input:
- $\{\mathbf{Y}^s \in \mathbb{R}^{n \times p}, s = 1, ..., S\}$, the n-long time series observed on p voxels for subject s
- $\mathbf{V}_{\text{init}}$, an initial guess for group maps $\mathbf{V}$

Output:
- $\mathbf{V} \in \mathbb{R}^{p \times k}$, the set of group-level maps
- $\{\mathbf{V}^s \in \mathbb{R}^{p \times k}\}$, the subject-specific spatial maps
- $\{\mathbf{U}^s \in \mathbb{R}^{n \times k}\}$, the k associated time series

$E_0 \leftarrow \inf$, $E_1 \leftarrow \inf$, $i \leftarrow 1$, $S_0 \leftarrow \emptyset$
while $E_i \; E_{i-1} > \epsilon E_{i-1}$ **do**
  if $i = 1$ or last_iteration **then**
    $S_i = S$
  else
    $S_i \leftarrow$ pick random subjects in $S \setminus S_{i-1}$
  end if

  parallel for all s in S_i **do**
    Load subject data from external memory:
      $\mathbf{Y}^s \leftarrow \text{load}(s)$
    Update $\mathbf{U}^s$ (Block coordinate descent):
      parallel for $l = 1$ to k:
        $\mathbf{u}_l^s \leftarrow \mathbf{u}_l^s + \|\mathbf{v}_l^s\|_2^{-2}(\mathbf{Y}^s - \mathbf{U}^s {\mathbf{V}^s}^\mathsf{T})\mathbf{v}_l^s$
        $\mathbf{u}_l^s \leftarrow \mathbf{u}_l^s / \max(\|\mathbf{u}_l^s\|_2, 1)$
      end for
    Update $\mathbf{V}^s$ (ridge regression):
      $\mathbf{V}^s \leftarrow \mathbf{V} + (\mathbf{Y}^s - \mathbf{U}^s \mathbf{V}^\mathsf{T})^\mathsf{T} \mathbf{U}^s (\mathbf{U}^{s\mathsf{T}} \mathbf{U}^s + \mu \mathbf{I})^{-1}$
      $\Delta \mathbf{V}^S \leftarrow \sum_{s=1}^{S} \left(\frac{\mathbf{V}^s - \mathbf{V}^{s-1}}{2|S|\mu} \right)^2$
    Remove subject s data from memory:
      dispose($\mathbf{Y}^s$)
  end for
  Update $\mathbf{V}$ (proximal operator):
    $\overline{\mathbf{V}^s} \leftarrow \frac{1}{S} \sum_{s=1}^{S} \mathbf{V}^s$
    parallel for $l = 1$ to k:
      $\mathbf{v}_l \leftarrow$
      $\text{prox}_{\frac{\lambda}{S\mu}\Omega}(\overline{\mathbf{v}_l^s}, \text{tolerance} = \frac{\Delta \mathbf{V}^S}{3})$
    end for
  Compute new value of energy:
    $E_i \leftarrow \varepsilon(\mathbf{U}^s, \mathbf{V}^s, \mathbf{V})$
  $i \leftarrow i + 1$
end while

4.5 Experimental results

Fig. 3.4 shows speed benchmarks realized on 48 subjects, parallelizing the computation of $\{\mathbf{U}^s, \mathbf{V}^s\}$ on 16 cores. Profiling results (Fig 3.4c) show that the update $\{\mathbf{U}^s, \mathbf{V}^s\}$ is the bottleneck. Using SCD with a stochastic subset of a fourth of the dataset proportionaly decreases the time of this step and only has a little impact on convergence rate per iteration (Fig 3.4a). However, the iteration time speedup brought by SCD dramatically increases overall convergence speed (Fig 3.4b). ADG yields an additional speed up, and altogether, we observe a speedup of a factor 2, but we expect it to increase with the group size. SCD combined with ADG enable tackling large groups.

5 Effects of parameters on atlas extraction

The TV-MSDL algorithm has 3 parameters. Each of them having an impact on the aspect of the generated maps. In this section, we build an intuition of this effect by taking a look at generated maps.

5.1 μ: global amount of penalty

μ is the weight attributed to the whole penalty term including subject-to-group map similarity and the TV-ℓ_1 regularization. As a result, increasing μ increases the constraints applied on the subject maps resulting in the loss of voxels not shared by all the subjects. This effect is shown in 3.5: Sparsity penalty is increased with μ resulting in sparsier maps when its value increases.

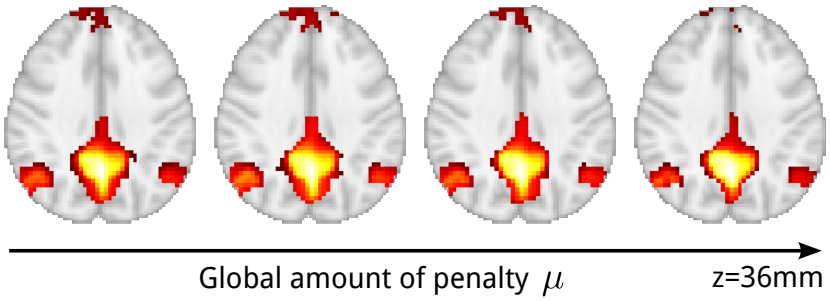


Figure 3.5: **Effect of the variation of TV-MSDL global amount of penalty μ on the default mode network extraction.**

5.2 Regularization parameters α and ρ

α is the weight of the regularization on the group maps. In our problem, it is only involved in the estimation of group maps $\mathbf{V}$. A low regularization promotes maps that explain the original signal better but may overfit if it is too low.

Since our regularization is composed of TV and ℓ_1, we can see both aspects reinforced when we increase it in Figure 3.6: the brain components get sparser but the aura around the blobs is also reduced. ρ defines the balance between

ℓ_1-norm and TV inside the regularization. As such, it is tightly linked to the α parameter.

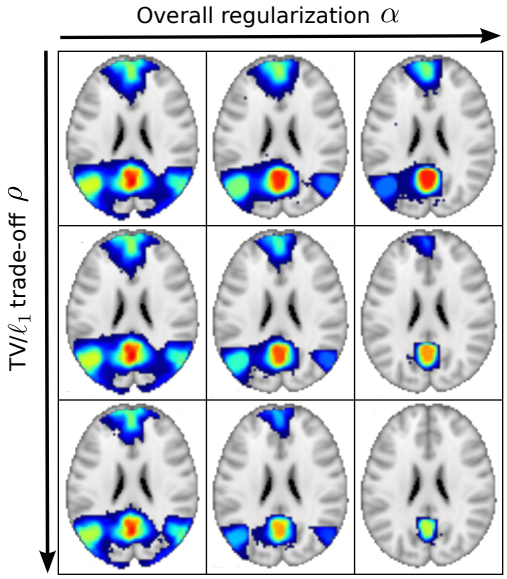


Figure 3.6: **Effect of the variation of TV-MSDL regularization parameters α and ρ on the default mode network extraction.** α is the parameter that controls the overall regularization. ρ controls the ratio of ℓ_1 norm in the TV-ℓ_1 regularization. Together, they control the expansion of brain areas across the brain. In this particular component, there is no ideal value that separate all the components of the default mode network. It is necessary to apply a post-hoc step, such as region extraction, described in the next chapter.

6 Results – Extracted brain atlases

6.1 Experiments

We run TV-MSDL on the ABIDE dataset presented in section 4. Following (Varoquaux et al., 2011), we use a dimensionality $k = 42$ that maximizes likelihood of left-out data. We consider each map as a region. Parameter μ controls the overall penalization. It has only a small impact on the resulting group-level maps and we set it to 2. ρ and α control the overall aspect of the maps. Maximizing explained variance on test data leads to setting $\rho = 2.0$ and $\alpha = 0.02$.

We compare TV-MSDL to Group independent component analysis (**ICA**), which is the reference method to extract brain maps from rest fMRI data (Beckmann and Smith, 2004). It extracts several components while enforcing the non gaussianity of their spatial maps. We use the implementation proposed in (Varoquaux et al., 2010c).

We also compare TV-MSDL to two clustering methods. **K-Means** is the historical technique used for fMRI time series clustering (Goutte et al., 1999; Thirion et al., 2014). It minimizes the ℓ_2 reconstruction error by learning a hard assignment for optimal compression. **Ward's clustering** also seeks to minimize the ℓ_2 error using agglomerative hierarchical clustering: Imposing a spatial constraint comes at no cost and it has been extensively used more recently to learn brain parcellations (Blumensath et al., 2012). As Ward's clustering directly extract blobs, we use it directly to extract 84 ROIs.

6.2 Brain atlases

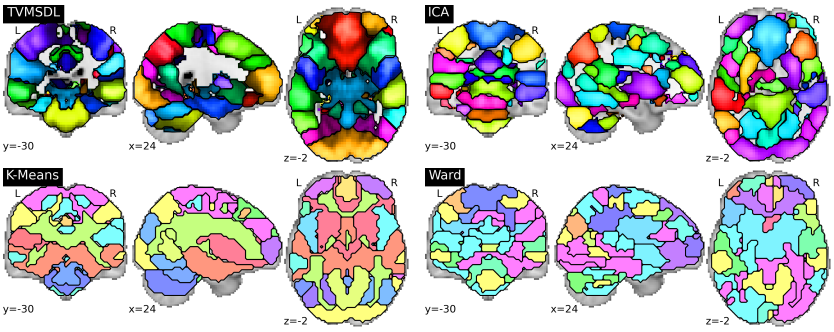


Figure 3.7: Regions extracted with the different strategies (colors are random). The representation for clustering methods K-Means and Ward is exact since each voxel belongs to a single region. For overlapping atlases such as ICA and TV-MSDL, we affected each voxel to the region of higher weight.

A flattened representation of the atlases is presented in Figure 3.7. First, we note that, unsurprisingly, algorithms with spatial constraints give more structured parcellations. This behavior is visible in the TV-MSDL regions. Regions extracted by TV-MSDL segment best well-known structures, such as the ventricles or gray nuclei. Finally, their strong symmetry matches neuroscientific knowledge on brain networks, even though it was not imposed by the model.

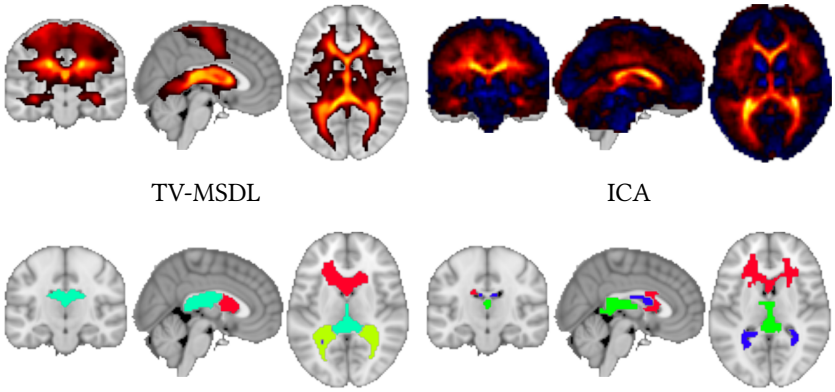


Figure 3.8: Ventricular system extracted by different atlas extraction methods.

Ventricular system Figure 3.8 shows the ventricular system extracted with different methods. ICA and TV-MSDL extractions follow closely the brain anatomy although TV-MSDL is better for extracting finer details. For example, in the coronal view, we see clearly the three ventricles while they seem to be merged in the ICA. This is probably due to the smoothing applied to the data which improves the overall structure of the atlas but blurs the small details. Clustering methods also manage to extract the ventricular system but it is divided in several regions.

Default Mode Network Figure 3.10 shows the default mode network extracted with different methods. Again, ICA and TV-MSDL are close to the

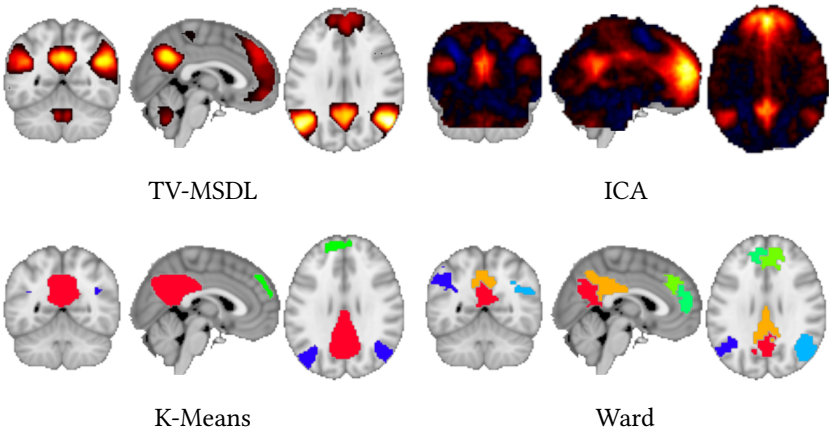


Figure 3.9: Default mode network extracted by different atlas extraction methods.

common definition of Default Mode Network. In the K-Means atlas, the PCC is a bit bigger than the other atlases but the DMN is still divided in 4 regions. In the Ward atlas, it oversegmented.

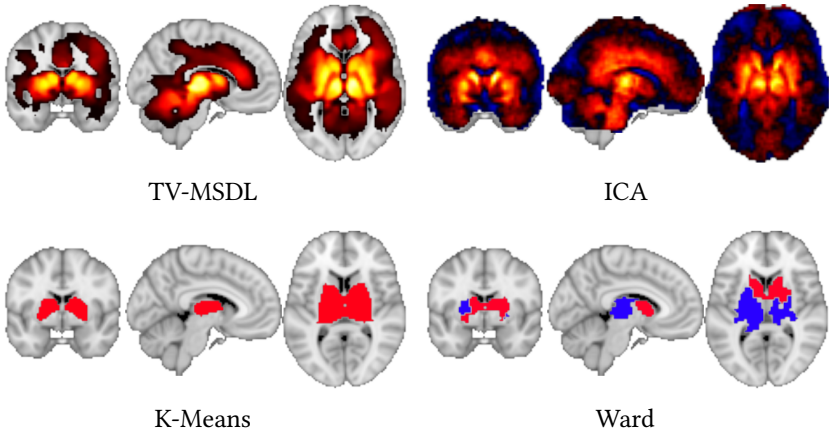


Figure 3.10: Basal nuclei extracted by different atlas extraction methods.

Basal nuclei Basal nuclei are ganglia located in the forebrain and are subject to high noise in the fMRI modality. ICA and TV-MSDL are again better at extracting these structures than clustering methods. In particular, we can see the anatomy in the weights of the maps. Again, TV-MSDL is more contrasted than ICA. K-Means and Ward are also able to retrieve the ganglia but, since they are hard assignments, they can't retrieve the underlying anatomy.

6.3 Anatomical validation

In the previous section, we assessed qualitatively that data-driven methods are better at extracting anatomical features. Here we propose a quantitative approach to measure the similarity of extracted atlases to reference atlases. We use a jackknifing approach to compute a high number of atlases and measure their Spearman correlation to reference atlases for basal ganglia. The results are reported in Figure 3.11.

We observe that TV-MSDL is the best at retrieving anatomical features in the brain. This confirms the results observed in the previous sections.

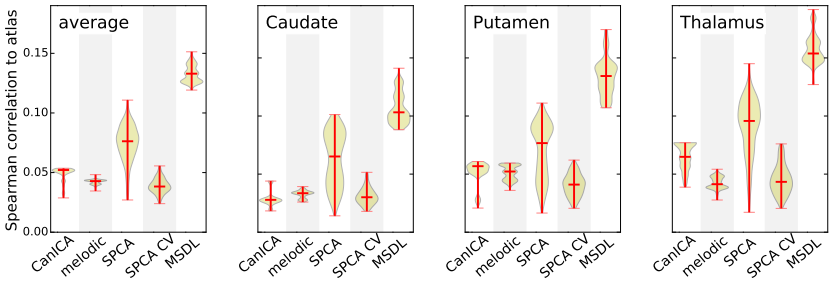


Figure 3.11: **Similarity scores of repeated atlas extraction compared to Harvard-Oxford basal ganglia reference.** The variance of the values accounts for the stability of the extracted atlas using different methods. The values themselves indicates how close the extracted basal ganglia are from the Harvard-Oxford atlas reference. SPCA is an ℓ_1-regularized decomposition method corresponding to our method with $\rho = 1$.

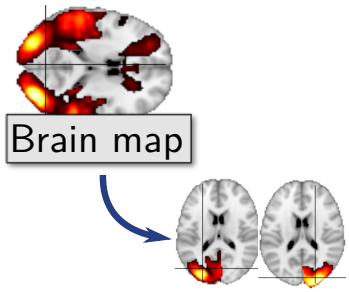
7 Conclusion

In this chapter, we introduced a new method to extract function brain maps from resting-state functional MRI dynamics. More precisely, we introduced a new regularization for the multi-subject dictionary learning approach that reveals neurologically plausible partition of the brain.

We use a sparse total-variation penalty with dictionary learning to combine the tendency of total-variation to create discrete spatial patches with the ability of linear decomposition models to unmix different effects. Careful choices of optimization strategy let our method scale to very large groups of subjects. Compared to reference methods, it improves the visual aspect of the functional networks.

Yet, our goal is to analyze the functional connectivity inside networks, between the different brain functional units that form them. In order to break down to this level, we need to extract the brain regions from these networks. Region extraction from continuous maps is discussed in the next chapter.

4 | Extracting brain regions from networks maps



The previous chapters have introduced a new brain network extraction method. Using a carefully crafted regularization, we enforce these networks to be composed of well delineated blobs. In order to study the connectivity inside the network, we need to segment these blobs. This task is trivial in the case of clustering, which generates binary maps, but not in the case of matrix decomposition methods, such as ICA or dictionary learning, that generate full brain continuous maps.

In this chapter, we present several strategies to extract these blobs. They consist in two successive steps: First we extract the foreground voxels of the brain maps, and then we extract connected components from these foreground voxels. We propose a qualitative assessment of the extracted region by visual inspection. The next chapter introduces a quantitative validation using model selection metrics.

The work presented in this chapter has been published in:

Region segmentation for sparse decompositions: better brain parcellations from rest fMRI, *Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, Gaël Varoquaux*, Sparsity Techniques in Medical Imaging, Sep 2014, Boston, United States. pp.8

1 Introduction – Region extraction as a post-processing on brain atlas

Networks extracted by ICA are full-brain and require a post-processing step to extract the salient features, i.e., brain regions, which is often done manually (Kiviniemi et al., 2009) (see Figure 4.4). To avoid post-processing and directly

extract regions, more sophisticated approaches rely on sparse, spatially structured priors (see chapter 3). Indeed, maps of functional networks display a small number of non-zero voxels, and thus are well characterized through a sparsity criterion, even in the case of ICA (Varoquaux et al., 2010b; Daubechies et al., 2009). However, sophisticated priors such as structured sparsity do not separate some networks into regions. Altogether, region extraction is unavoidable to go from brain image decompositions to Regions-of-Interest-based analysis (Nieto-Castanon et al., 2003).

A simple approach to obtain sharper maps is to use hard thresholding, which is a good sparse, albeit non convex, recovery method (Blumensath and Davies, 2009). We improve upon it by introducing richer post-processing strategies with spatial models, to avoid small spurious regions and isolate each salient feature in a dedicated region. Based on purely geometric properties, these strategies take advantage of the spatially-structured and sparsity-inducing penalties of recent dictionary learning methods to isolate regions. They can also be used in the framework of computationally cheaper ICA.

2 Region extraction methods

Clustering methods do a hard segmentation of the brain: each voxel belongs to one and only one region. Decomposition methods, such as ICA and dictionary learning do soft assignment: a voxel can belong to several functional regions. In fact, multivariate decomposition techniques most often decompose the signal of one voxel as a linear mixture of several signal components, which matches better the physical generating process. In practice, these overlapping regions are small and located in areas of low confidence. Voxels that belong to no component are left unlabeled.

Extracting regions to outline objects is a well-known problem in computer vision. For the particular problem of extracting regions of interest (ROIs) out of brain maps, we want a method that *i)* handles 3D images *ii)* processes one image while taking into account the remainder of the atlas (e.g., region extraction for a given image may be different depending on the number of other regions) and *iii)* isolates each salient feature from a smooth image in an individual ROI without strong edges or structure (see Figure 4.1). Here, we assume that a given set of *brain maps* has been obtained by a multivariate decomposition technique.

Depending on the regularization of the matrix decomposition method, extracted brain networks can be more or less sparse. For example, TV-MSDL extracts sparse components while ICA extracts full brain components (see Figure 3.8). For this reason, we perform a foreground extraction step before extracting brain blobs. An example of region extraction is shown in Figure 4.1.

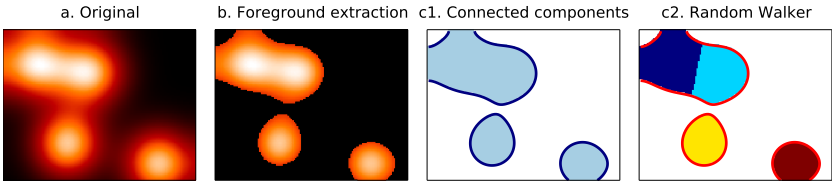


Figure 4.1: Example of region extraction. Foreground pixels (b) are extracted from the original image (a). Regions are extracted using connected component extraction (c1) or random walker (c2).

2.1 Notations

This section describes region extraction techniques. In order to define them, we represent a brain image as a graph: each voxel being a vertex of the graph. One voxel is linked to its neighbors by edges. We use the 6-neighborhood relationship as depicted in Figure 4.2).

Let $\mathbf{V} = \{\mathbf{V}_1, ..., \mathbf{V}_k\}$ be the set of brain maps (3D images), or atlas[1]. An image $\mathbf{V}_i$ is seen as a 3-dimensional graph with 6-connected vertices. For each image $\mathbf{V}_i$, we consider G_i, its associated undirected graph composed of a set of p vertices V^i and a set of q edges E^i (represented as a list of triplets of 2 vertices and the weight of the edge):

$$G_i = (V^i, E^i), \quad V^i \in \mathbb{R}^p, E^i \in (\mathbb{N} \times \mathbb{N} \times \mathbb{R})^q$$

G_i is constructed so that the vertices are ordered across all the graphs: For all graphs, the vertices of index x correspond to the points p in the original image with the same 3D coordinates.

The region extraction process is divided in two steps. The first step consists in extracting relevant voxels using a foreground extraction procedure. We define it as an extraction of subgraph in the graph formalism. The second step consists in extracting regions of interest from the foreground voxels, which is equivalent to a clustering in graph formalism.

Given a subset of vertices $V' \subseteq V$, we denote by $\langle V' \rangle$ the subgraph V' containing all vertices of V' and the subset of edges $E' \subseteq E$ connecting vertices in V':

$$\langle V' \rangle = G' = (V', \{(v_i, v_j, e_{ij}) \in E, (v_i, v_j) \in V' \times V'\})$$

In the following, we assume all vertices positive. In the case of negative components in an atlas, we split the brain map in two: its positive part and its negative counterpart, before extracting regions.

2.2 Foreground extraction

In this section, we restrain the number of voxels per map by selecting the most meaningful ones. In graph formalism, this boils down to selecting a subset of meaningful vertices in a graph. Computationally, we select foreground voxels by setting background voxels to zero.

[1] We keep the same formalism as the previous chapter.

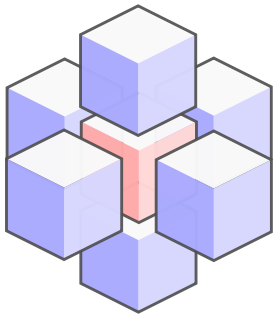


Figure 4.2: **3-dimensional 6-neighborhood:** the neigbors of the considered red voxel are shown in blue. 6-neighborhood implies that only immediate neighbors are considered neighbors. Diagonal voxels are not.

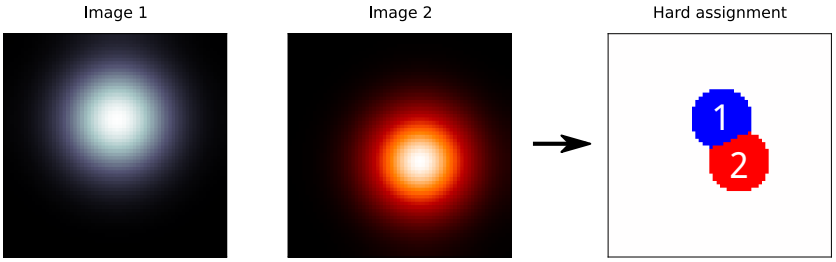


Figure 4.3: Example of hard assignment. An atlas of 2 images is presented on the left. The hard-assignment (or maximum a posteriori labeling) is presented on the right and results in the fusion of the two images. We can see that voxels are attributed to only one region.

Normalization Multivariate decomposition techniques – such as dictionary learning or ICA – assign a weight to each voxel. This weight accounts for the quantity of the component signal present in the voxel's signal. This value has a meaning inside the component but cannot be interpreted relatively to the rest of the atlas.

In a preliminary study, we experimented with normalization regarding to maximum value and to unit variance. In the particular case of brain maps, with small overlaps in low confidence values, we observed that the choice of normalization method has no significant impact on the results. In the following, the normalization to unit variance (without substracting the mean) is used.

Hard assignment. Hard assignment transforms a set of overlapping maps into a brain segmentation with no overlap between regions. In the end, each voxel will be represented by a unique brain map from the atlas. This map is the one that has the highest value for this voxel. This procedure is also called *maximum a posteriori labeling*. The result is a segmentation from which we can extract connected components, as shown in Figure 4.3.

In graph formalism, this corresponds to keeping vertices that have the maximum values across all images:

$$\text{HARD}(G_i) = \langle \{V_x^i, \forall j \neq i \in 1...k, V_x^i > V_x^j\} \rangle$$

Mathematically, these voxels can be selected using an *argmax* operation. We implement it by setting to zero all the non-foreground voxels (*i. e.* background). A naive implementation is proposed in listing 4.1.

Input:
- $\mathbf{V} \in \mathbb{R}^{p \times k}$, the set of group-level maps: k maps with p voxels

```
for x ∈ 1...p do
    max_p ← arg max((V_1)_x, ..., (V_k)_x)
    (V_{i≠max_p})_x ← 0
end for
return \B{V}
```

Listing 4.1: Implementation of hard assignment on group maps $\mathbf{V}$ extracted using TV-MSDL.

Automatic thresholding. Thresholding is the common approach used to extract ROIs from ICA. However, the threshold is usually set manually and is different for each map. In order to propose an automatic threshold choice, we consider that, on average, an atlas assigns each voxel to one region. For this purpose, we set the threshold $t^k(G)$ so that the number of nonzero voxels – across all the maps – corresponds to the number of voxels in the brain. This boils down to considering the threshold $t^k(G)$ as the $(k–1)^{\text{th}}$ k-quantile across all voxels:

$$\text{AUTO}(G_i) = \langle \{ V_x^i, \quad V_x^i > t^k(G) \} \rangle$$

However, since we are working with overlapping regions, this threshold tends to leave too much brain voxels aside. In practical, the $(k–1.5)^{\text{th}}$ k-quantile gives good results.

On Figure 5.9b, we can see that the region contours are much smoother than regions extracted using hard assignment. However, the segmentation is not still not perfect: the two blobs on the left are considered as one.

2.3 Component extraction

Our graphs are now composed of a reduced set of voxels. A simple way to extract components is to extract the connected components (or subgraphs) from the graph. This method is presented first. However, in some cases, the components have not been properly separated by the foreground extraction. In order to properly separate them, we use hysteresis thresholding and random walker.

Connected components. In graph formalism, a connected component is a subgraph where all vertices are connected. We say that two vertices are connected if there exists a path, *i.e.* a succession of edges, from one to the other:

$$\text{connected}(u, v) \Leftrightarrow \exists (x_1, ..., x_n) \in V^i, (u, x_1), (x_1, x_2), ...(x_n, v) \in E^i$$

We denote by $CC(\mathbf{V})$ the connected components of graph $\mathbf{V}$.

Hysteresis thresholding. Hysteresis thresholding is a two-threshold method where all connected components with value higher than a given threshold t_{high} are used as seeds for the regions. Neighboring voxels with values between the high threshold t_{high} and the low threshold t_{low} are added to these seed regions. In our setting, the high threshold can be seen as a minimal activation value over the regions in order to sort out regions of marginal importance. Each brain map has its own optimal value but, in practice, cross validation has shown that keeping the 10% highest foreground voxels as seeds gives the best results[2]. The automatic thresholding strategy described above is used to set the low threshold t_{low}.

[2] In terms of stability and explained variance. These metrics are introduced in the next chapter.

In graph formalism, this is equivalent to keeping only the connected components which have at least one vertex with a value over t_{high}. We define this operation as $\text{HYST}(G)$:

$$\text{HYST}(G) = \{(V', E') \in \text{CC}(\text{AUTO}(G)), \exists v' \in V, v' > t_{texthigh}\}$$

It can also be defined generatively: for each connected component, we use the seeds as starting set and add make it grow by adding neighbors which value is higher than t_{low}.

Random Walker. Random Walker is a seed-based segmentation algorithm similar to watershed. It calculates, for each point v, the probabilities to end up in each of the r seeds $s_1, ..., s_r$ by doing a random walk across the image starting from v. The original version of the algorithm (Grady, 2006) was of probabilistic nature, whereby the probability to jump to a neighboring point is driven by the gradient magnitude between them. After convergence the point is attached to the seed with the highest probability.

Random Walker can also be seen as a diffusion process. It is equivalent to hysteresis thresholding where regions that have grown enough to share a boundary are not allowed to be merged. In graph formalism, this would be equivalent as making sub-components grow by adding edges but never let an edge connect two connected components, as it may happen in hysteresis thresholding. The definition of Random Walker as a diffusion process is formalized in (Grady, 2006). The probabilities to reach each of the seeds can be computed using the Laplacian matrix $\mathbf{L}$ of the graph associated with the map.

The classical random walker approach uses the gradient between two voxels (or vertices) to determine the weight of the edge between them. In our setting, a high value in the map means a high confidence. So, instead of using the finite difference gradient, we consider the max of the image minus the lowest voxel. Therefore, diffusion is facilitated in areas of high confidence and more difficult elsewhere. Following (Grady, 2006), we use a Gaussian weighting function:

$$e_{ij} = \exp(-\beta(\max(V) - \min(v_i, v_j)))$$

We refer the reader to (Grady, 2006) for the complete description of the algorithm. In the following, we denote by $P_s(v)$, the probability of point v to end up in seed s through a random walk.

$$RW = \left\{ \langle v_i \in V \quad \text{s.t. } \arg\max_s P_s(v_i) = s, \quad \forall s \in 1, ..., r \rangle \right\}$$

If we consider the set of probability maps given for each seed, extracting the connected components can also be seen as a hard assignment on them.

We tested two strategies to select the seeds: taking the local maxima of the smoothed image (this method is called automatic Random Walker in the rest

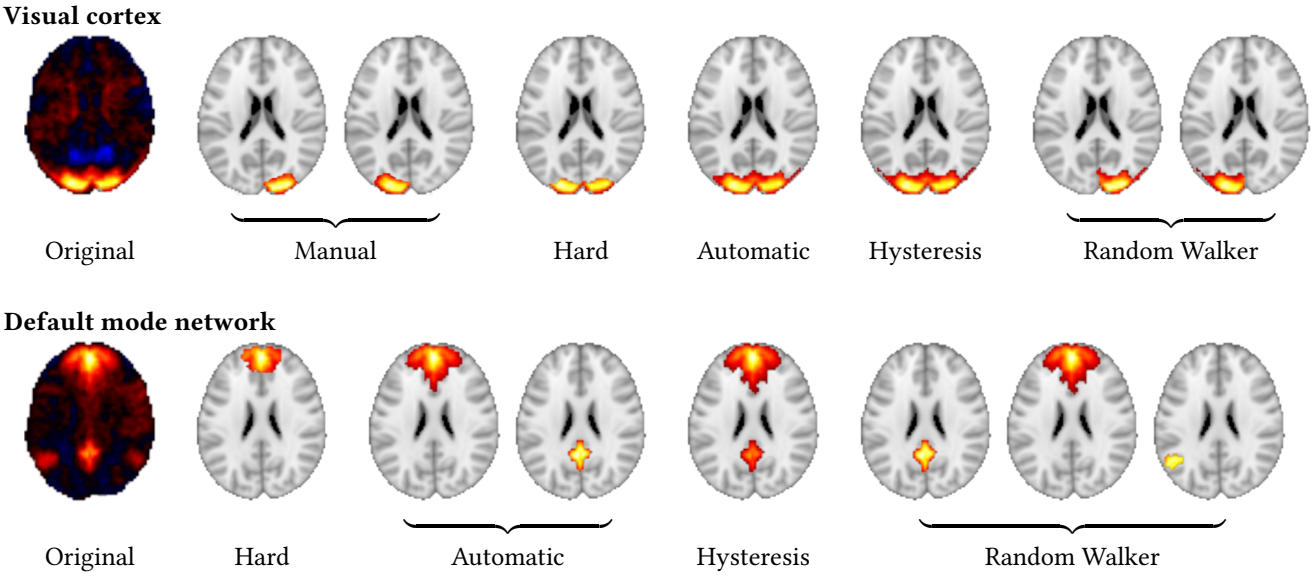


Figure 4.4: Example of region extraction in maps obtained by ICA.

of the manuscript) or taking the hard assignment of the images (denoted as hard Random Walker in the rest of the manuscript). We choose the latter as it gives significantly better results.

3 Results – Extracted brain regions

This section presents a qualitative assessment of the maps extracted with each region extraction method. A quantitative approach is used in the next chapter.

Figure 4.4 shows 2 networks out of 42 extracted thanks to spatial group ICA. The first one, the visual cortex, is located in the occipital part of the brain. It is composed of two regions located in each hemisphere, as shown in the example of manual segmentation. Because these regions tend to be linked by relatively high values, it is hard to segment them properly. Region extractions based on thresholding fail at separating them properly. The Random Walker (both approaches) is the only approach able to separate them.

The default mode network (DMN) is composed of four regions. In our experiment, no region extraction method is able to recover them all, the best being Random Walker with three regions recovered. We clearly see the weakness of hysteresis thresholding in this figure: the threshold is too high to recover the temporo-parietal junctions[3] but too low to properly seperate the posterior cingulate cortex (PCC) and the frontal component of the DMN.

[3] These regions are the smallest ones located at the extreme right and left of the brain

4 Conclusion

Functional atlases extracted using ICA or sparse decomposition methods are composed of continuous maps and sometimes fail to separate symmetric

functional regions.

Starting from hard thresholding (Blumensath and Davies, 2009), we introduce richer strategies integrating spatial models, to avoid small spurious regions and isolate each salient feature in a dedicated region. Indeed, the notion of regions is hard to express with convex penalties. Relaxations such as total-variation used in previous chapter only captures it partially, while a non-convex segmentation step easily enforces regions. We find that a Random-Walker based strategy manages to recover regions better than strategies based on thresholding operations. However, those are qualitative results and proper metrics are required to do a quantitative study of these approaches. Such metrics are presented in the next chapter.

5 | Selecting the best atlas: metrics for stability and data faithfulness

Previous chapters introduce parametrized atlas extraction methods along with post-hoc region extraction. Qualitative assessments have been made to pick a best perfomer among all. However, a quantitative approach is necessary, in particular in the context of automated atlas validation. A quantitative assessment yields an objective way to select the best model among a large number of candidate solutions. Qualitative inspection by humans are biased or not clearly guided by criteria and cannot be done when the number of assessments becomes large.

For that purpose, we introduce two metrics based on prior assumptions on the ideal qualities a brain atlas should exhibit. Unlike next chapters, we are interested in judging the quality of an atlas in terms of describing the brain, and not for a specific pathology. We use these metrics to find the best parameters in our models, select the best atlas extraction model and finally select the best region extraction method.

As it turns out, these properties can become contradictory in certain situations and it is important to evaluate the possible tradeoffs. We will introduce a formalism by which we can easily chose the best set of parameters if there is one, and otherwise report the frontier of the best possible tradeoffs.

Part of the work presented in this chapter has been published in:

Region segmentation for sparse decompositions: better brain parcellations from rest fMRI, *Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, Gaël Varoquaux*, Sparsity Techniques in Medical Imaging, Sep 2014, Boston, United States. pp.8

1 Introduction – Interesting properties of brain atlases

Gauging the success of an atlas extraction method is challenging because it is not meant to be taylored to specific needs, such as the diagnosis of a given pathology. In addition, in terms of basic functional organization of the brain, no ground truth is available to which our results could be compared. However, we can formulate hypotheses on the qualities a brain atlas should have. A brain atlas must *i)* represent the common functional organization existing across individuals and *ii)* provide an adequate basis to explain the brain activation for each individual. Based on these assumptions, we can state that a tool is suitable to represent brain function if the extracted atlases are stable with regards to which subjects we choose to extract the atlas, and if they can faithfully model the original signal.

What is stability? Extracted atlases must be similar when varying the set of subjects chosen in an homogeneous population. In fact, one of our hypotheses is that all individuals share the same brain functional organization. Based on this assumption, functional atlases estimated on different groups of individuals should be similar: This is what we call *stability*. In order to measure the similarity between several brain atlases, we rely on metrics designed to measure the similarity of clusterings developed in the field of computer vision. However, stability itself is not enough to characterize a good atlas. In fact, a method can extract the same atlas independently of the subjects used, and maximize stability, but badly represent the brain functional organization.

What is data fidelity? Extracted atlases must model accurately the underlying organization of the brain. As there is no ground truth, in the previous chapter, we relied on visual appreciation and correlation with a reference atlas. In order to measure data fidelity quantitatively, we rely on explained variance. Explained variance (also called R^2 or R squared) is used routinely to measure how well a model fits the data. In our case, we also want to measure how well the atlas is able to generalize to new subjects. For this purpose, we measure explained variance on the testing set of our experiment.

In this chapter, we introduce metrics for these two aspects and show how to use them to compare brain atlas extraction methods.

2 Stability with regards to training sets

In order to measure the stability of an atlas extraction method, we must compare atlases estimated from different groups of individuals. As we expect the underlying functional brain organization to be similar across individuals, the extracted atlases must be similar. However, *similarity* is not a well defined concept and each scientific field has its own metrics to measure it. We first

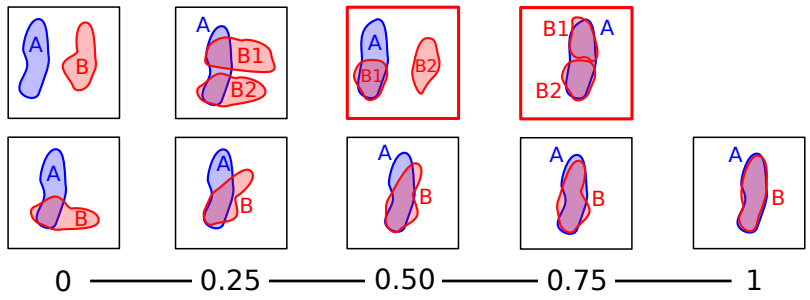


Figure 5.1: **Example of Normalized Mutual Information scores for simple cases.** Several clustering examples are depicted along with their score using the NMI metric.

define the problem and build an intuition about what a good metric is. Then, we present the metrics that we used to measure similarity on atlases.

2.1 Problem statement

Our goal is to obtain a measure of similarity between different brain parcellations. Such parcellations can be obtained using several methods: *i)* anatomical properties, as seen before with AAL and Harvard-Oxford *ii)* clustering methods, such as K-Means which produce segmentations of the brain (*i.e.* each voxel belongs to one and only one brain region) *iii)* matrix decomposition methods, such as ICA, which extract full brain components where each voxel has a particular weight.

This problem has been widely explored in computer vision and a lot of similarity metrics exist (Pfitzner et al., 2009). In order to chose one suitable for our problem, we study specific cases, define the wanted behavior in that case, and finally select the metrics that follow this behavior.

For example, in Figure 5.1, top right, we see an example of oversegmentation: a single region in atlas A corresponds to two regions in atlas B. This case is not handled properly by similarity metrics that do pairing of regions: the big region in A will be paired to one of the regions is B, ignoring completely the second one.

In a first step, we consider a similarity measure on the hard assignment of atlases, for the sake of simplicity. In a second step, we try to extend this metric to overlapping fuzzy maps, in order capture finer details.

2.2 Measuring the similarity of brain segmentations

In order to measure similarity between brain clusterings, we rely on reference metrics used in the domain of computer vision. The three reference metrics are the DICE coefficient, the Normalized Mutual Information (Vinh et al., 2010) and the Rand Index. These measures present properties more fitted to sparse atlases than spatial correlation.

In order to chose the metrics most adapted to brain segmentations, we take a look at common cases. First, in Figure 5.1, we see in the top right corner two highlighted examples. In these cases, one region extracted in a clustering is

splitted in two in another one. These examples are paramount for the choice of the metric: NMI and RandIndex, that rely on co-occurence matrices, give them different scores while the DICE similarity coefficient, that relies on pairing of regions, will not. This is why we do not use the latter. As NMI and Rand Index are very similar in other interesting cases, we chose the NMI since computing it is less memory, CPU and time consuming.

Normalized mutal information (NMI) measures the statistical independance of two clusterings. Given two hard assignments A and B with marginal entropy $H(A)$ and $H(B)$ respectively,

$$NMI(A,B) = \frac{H(A) + H(B) - H(A,B)}{\sqrt{H(A) * H(B)}} \tag{5.1}$$

$$\text{with entropy defined as} \quad H(A) = -\sum_{i=1}^{n} p(a_i)\ log\ p(a_i)$$

$$\text{and joint entropy defined as} \quad H(A,B) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(a_i,b_j)\ log\ p(a_i,b_j)$$

Several ways to normalize the mutual information exists. We use a normalization that behaves well even in corner cases as described in (Pfitzner et al., 2009)[1]. We use the implementation of the scikit-learn (Pedregosa et al., 2011) that uses this normalization too.

[1] The formulation given in equation 5.1 is labelled NMI 5 in the paper

2.3 Extension to fuzzy maps

Since the original definition of NMI only processes hard segmentation, we look for extension of this metric to sets of fuzzy maps. Such extensions exists for Normalized Mutual Information, Rand Index and DICE coefficient.

A big drawback of NMI and Rand Index is that their fuzzy extensions only works for probabilistic clusterings where *i)* each map has a maximum of 1, *ii)* and the sum of the values of each voxel weights across maps is 1. In addition, these two metrics require to compare all pairs of voxels. This yields computational problems because calculating full similarity matrices is slow and CPU expensive. Finally, these measures yield over-optimistic scores as background voxels (which cover more than 98% of each map) are taken into account, reducing the size of the effect of interest, *i.e.* region differences.

On the other hand, a fuzzy approach of the Tanimoto coefficient (or Jaccard Index, a measure very close to Dice similarity coefficient) is easy to compute and it does not have the drawbacks observed for its non-fuzzy counterpart. It has been already used in the context of neuroimaging in (Crum et al., 2006).

$$TC(A,B) = \frac{\sum_{i,j} \frac{min(a_i,b_j)}{max(a_i,b_j)}}{card(A) * card(B)}$$

Other metrics have been considered such as Earth Moving Distance or metrics proper to graph thory of fuzzy set theory. However, they were either too complicated computationally or not adapted to our problem.

2.4 Chosing a model selection metric.

We presented two metrics for measuring similarity of brain atlases. However, for the sake of simplicity, we will present the results using only one of them. Tanimoto coefficient takes more information into account than NMI since it is computed on the fuzzy maps. We see it in the distribution of the metric values across all experiments in Figure 5.2: for NMI, most scores are concentrated in high values whereas the scores of Tanimoto coefficient are more spread. For this reason, we continue with Tanimoto coefficient for the rest of the study. However, conclusions regarding best models are identical for NMI and Tanimoto coefficient.

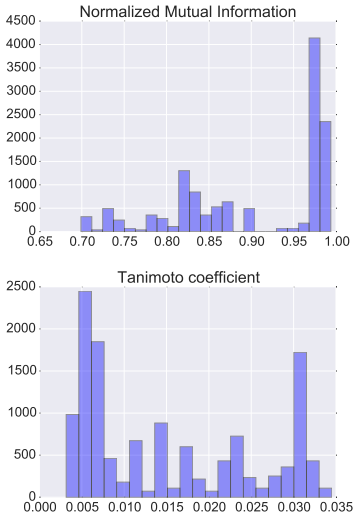


Figure 5.2: **Distribution of stability metric values on all experiments. Top:** Normalized Mutual Information. **Bottom:** Tanimoto coefficent

3 Measuring data faithfulness

3.1 Problem statement

Using an atlas to represent brain signals can be seen as a compression of these signals. In fact, instead of expressing a brain signal by the signal measure un each voxel, we aggregate voxels that share same activation signals together and use a single time series to represent them.

The goodness of fit of such method can then be quantified by measuring how far the reconstructed signal is from the original one. In the decomposition models, this quantity is called "residuals" because it represents the residual data that could not be explained by the model. In Figure 5.3, we show an exemple of such reconstruction. The goal in this figure is to have the green line as close to zero as possible. A widely used metric for this problem is the explained variance.

Principal component analysis (PCA) maximizes the explained variance on the training set. However, in the context of brain atlas extraction, we want the atlas to generalize to unknown individuals – individuals not present in the training set – as much as possible. For this purpose, we measure the explained variance on the testing set instead of the training set.

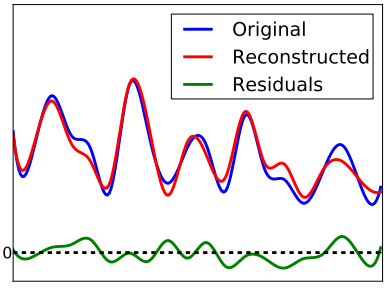


Figure 5.3: Explained variance.

3.2 Explained variance

The explained variance measures how much a model accounts for the variance of the original data. The more variance is explained, the better the model explains the original data.

In our problem, we want to know how much an atlas $\mathbf{V} \in \mathbb{R}^{p \times k}$ accounts for the variance of the original signal $\mathbf{Y} \in \mathbb{R}^{p \times n}$. For this we define the linear regression model:

$$\mathbf{Y} = \mathbf{V}\beta + \mathbf{E}$$

with $\beta \in \mathbb{R}^{k \times n}$ the regression coefficients and $\mathbf{E}$ the error term. We use an ordinary least square to estimate the regression coefficients:

$$\hat{\beta} = (\mathbf{V}^\mathsf{T}\mathbf{V})^{-1}\mathbf{V}^\mathsf{T}\mathbf{Y}$$

For convenience, we will write the signals reconstructed by the model as $\hat{\mathbf{Y}} = \mathbf{V}\hat{\beta}$. The percentage of variance explained by the model is then computed using the variance of the error term: Explained variance of these signals is then computed over the original ones.

$$\mathrm{EV}(\hat{\mathbf{Y}}) = 1 - \frac{\mathrm{Var}(\mathbf{E})}{\mathrm{Var}(\mathbf{Y})} = 1 - \frac{\mathrm{Var}(\mathbf{Y} - \hat{\mathbf{Y}})}{\mathrm{Var}(\mathbf{Y})}$$

Going further, we can expand the definition of $\hat{\mathbf{Y}}$:

$$\begin{aligned}
\mathrm{Var}(\mathbf{Y} - \hat{\mathbf{Y}}) &= \mathrm{Var}(\mathbf{Y} - \mathbf{V}\hat{\beta}) \\
&= \mathrm{Var}(\mathbf{Y} - \mathbf{V}(\mathbf{V}^\mathsf{T}\mathbf{V})^{-1}\mathbf{V}^\mathsf{T}\mathbf{Y}) \\
&= \mathrm{Var}(\mathbf{Y} - P_\mathbf{V}(\mathbf{Y})) \\
&= \|\mathbf{Y} - P_\mathbf{V}(\mathbf{Y})\|^2
\end{aligned}$$

where $P_\mathbf{V}$ is the projector onto $\mathbf{V}$. In the end, the explained variance is the norm of the projection of the data $\mathbf{Y}$ onto the components $\mathbf{V}$ that we have learnt.

4 Model selection plots: exploring the trade-off

Both metrics have values ranging from 0 to 1, 1 being the best value. These metrics can be used to compare brain atlases between them. In particular, we can use them to determine the optimal parameters of a parametrized model by comparing the resulting atlas with other models. It is possible that no *best* model stands out from the other because of a trade-off between the two metrics. In that case, the set of models that exhibit the best value for one metric w.r.t. the other is called *Pareto efficiency frontier*.

In the NPAIRS framework (Strother et al., 2002; LaConte et al., 2003), models are plotted in a two-dimensional plot with one stability as x-axis and prediction accuracy as y-axis. In a similar manner,we compare atlases in a two

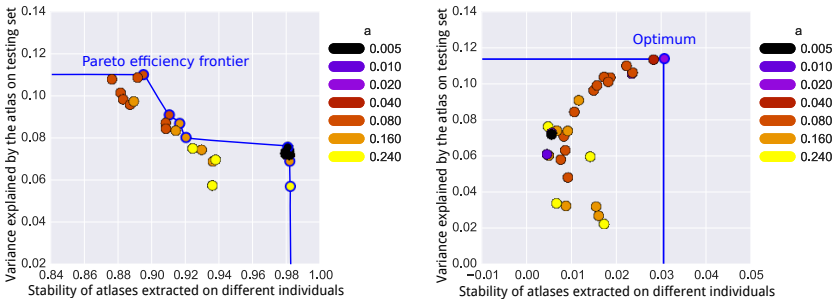


Figure 5.4: **Examples of model selection plots. Left:** a plot where several models belong to the Pareto efficiency frontier. There is no clear winner, we cannot conclude on the best value for parameter a. **Right:** a plot where only one optimum belong to the Pareto efficiency frontier. We can state that the best value for parameter a is 0.02.

dimensional plot with stability as x-axis and data faithfulness as y-axis. Each point is an atlas and any other feature can be associated with the colors of the point. Best models are at the top right of the plot. Example of such plots are shown in Figure 5.4. We observe two kinds of trend. The plot on the left exhibits a Pareto efficiency frontier with several points where an optimum cannot be determined. On the right, there is a top performer among all models and we can determine the optimal value for a.

5 Results – Best atlas estimation and region extraction methods

5.1 Finding the best TV-MSDL parametrization

We extract atlases following the protocol described in section 4. The global stability score is obtained by averaging the similarity scores computed on all pairs of atlases. The data faithfulness score of a fold is the metric applied on the concatenation of the time series of the testing set. The global score is the average of the metric across all folds.

In chapter 3, we showed the effect of the TV-MSDL parameters on the visual aspect of the maps. In this section, we study quantitatively this effect on our selection metrics and use them to find the optimal value for each parameter.

A first observation is that region extraction increases explained variance a little. Stability is not comparable since it is sensible to the number of components that is different between raw maps (42 maps) and extracted regions (84 regions).

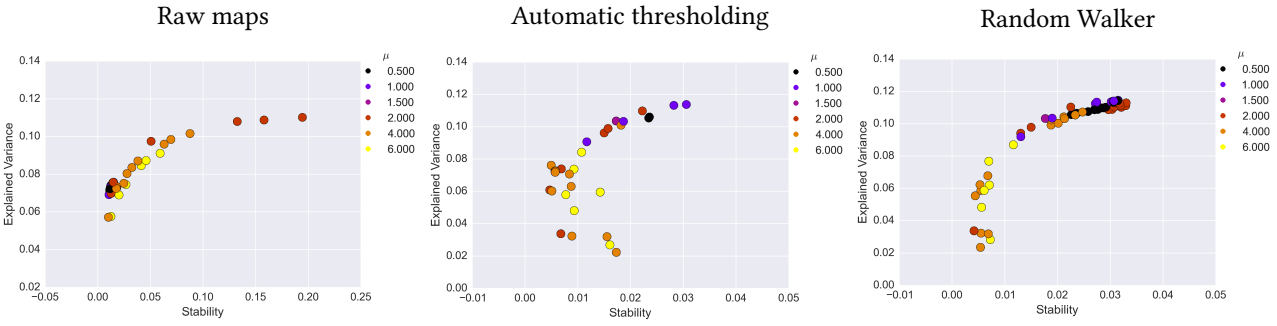


Figure 5.5: **Model selection plots for parameter μ.** Best performers are in top right corner.

Global regularization μ μ weights the entire regularization term of the TV-MSDL including the attach to group maps and the TV-ℓ_1. In Figure 5.5, we observe a common trend in our plot. We start from $\mu = 0.5$ that gives average results. Increasing to $\mu = 2.0$ improves the metrics scores. When going above 2.0, we observe that the metrics scores decrease. From these plots, we can locate the optimal value for μ between 1.5 and 2.0 Full plots (Figure 7) at the end of this chapter support this conclusion.

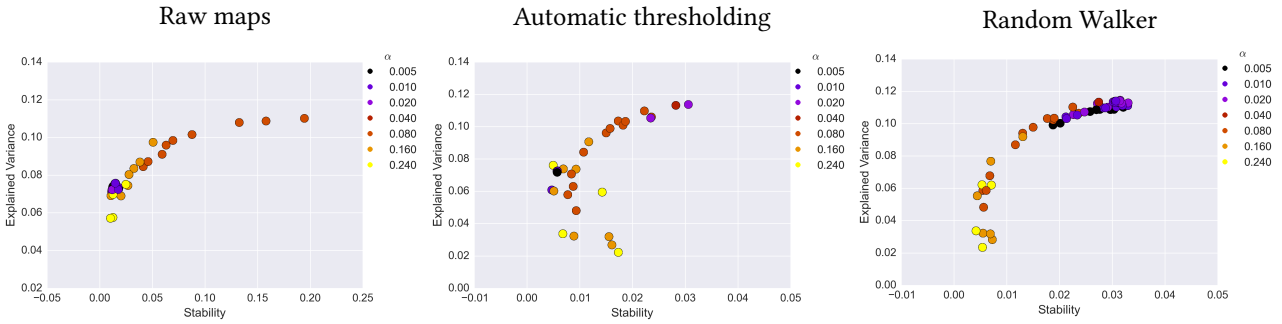


Figure 5.6: **Model selection plots for parameter α.** Best performers are in top right corner.

Maps regularization α Model selection plots for the parameter α are presented in Figure 5.6). We observe the same trend as for the global regularization parameter μ. We can locate the optimal value between 0.01 and 0.04.

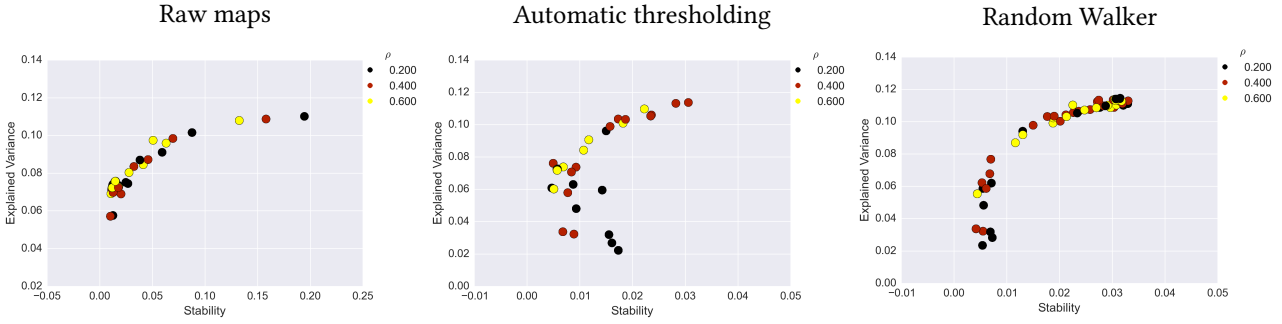


Figure 5.7: **Model selection plots for parameter ρ.** Best performers are in top right corner.

Ratio of sparsity ρ Probably because of its dependency on other parameters, it is hard to see a clear trend for this parameter (Figure 5.7). For the worst models in term of metric, we see that a higher value of ρ yields better results. But for the best models, it is unclear. In our experiences, we have seen that a lower sparsity is globally better if regions are meant to be extracted from the atlases afterward. Full plots (Figure 7) at the end of this chapter support this conclusion.

Conclusion Measuring the quality of an atlas is hard. In the case of raw atlases, NMI exhibits a Pareto efficency frontier and Tanimoto coefficient exhibits results in disagreement with NMI. However, after region extraction, a shift occurs in the model selection plots: The Pareto efficiency frontier observed in the raw maps disappears and an optimal model emerges. Our

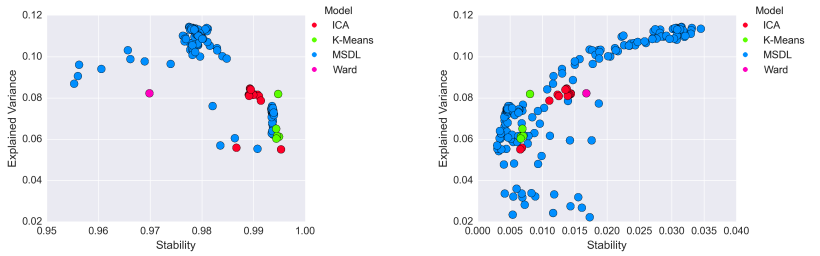


Figure 5.8: **Model selection plots for several methods. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

two metrics also designate the same model as best performer for each region extraction method. We are able to determine the optimal TV-MSDL parameters from these optimal models. Interestingly, the optimal parameterizations are the same across all region extraction methods. This result does not only allow us to determine the effect of each parameter on TV-MSDL and determine the optimal settings, it also enlightens the fact that region extraction is beneficial for comparing continuous atlases.

5.2 Finding the best model

We have shown in the previous section that our metrics matched experimental results and can be used to find the best parametrization for TV-MSDL. We apply the same metrics to other atlas extraction approaches, namely ICA and K-Means to determine which one is the best. For this task, we show both NMI and Tanimoto coefficient metrics for stability as they give different results.

Metrics validation. In the previous section, we stated that NMI and Tanimoto coefficient agree to designate the best performing parametrization. In Figure 5.8, we first see that all methods have metric scores in a similar range, which is encouraging. NMI exhibits a Pareto efficiency front while Tanimoto coefficient designates a best performer. In particular, we observe two clusters of TV-MSDL values in the NMI plot that corresponds to different region extraction methods. The most stable ones, with K-Means correspond to hard assignment and are a degenerated case of atlas extraction. For these reasons, we prefer Tanimoto coefficient to study similarity across atlas estimation methods.

Selecting the best model. In Figure 5.8, we see a trade-off between stability and data faithfulness with the NMI metrics. Each methods dominates a part of this trade-off. Interestingly, ICA is considered as less stable than K-Means for NMI but, when fuzziness and overlapping maps are taken into account, it is more stable in the Tanimoto coefficient plot. It is still a good compromise between the two metrics. TV-MSDL, depending on the region extraction method used, can be either unstable but with high explained variance, or stable with less explained variance. K-Means clustering is the most stable method and is on par with MSDL. Ward's clustering is suprisingly not stable

which contradicts our first experiments.

Conclusion. Explained variance is clearly dominated by matrix decomposition methods. Their overlapping weighted maps give them a significant advantage over the clustering methods.

However, clustering methods are stable models. According to NMI, K-Means is the most stable, and, according to Tanimoto coefficient, Ward is the best model behind TV-MSDL. Our metrics does not allow us to pick a clear winner.

5.3 Finding the best region extraction method

Effect on brain maps In a first preliminary experiment, we experimented the effect of region extraction on a specific brain map. Experiments were made on the ABIDE dataset. We selected 101 subjects suffering from autism spectrum disorders and 93 typical controls from 4 sites and computed brain atlases on 10 cross-validation iterations by taking a random half of the dataset as the train set. We extract regions from these atlases and quantify their performance on the other half of the dataset with two metrics. We investigate the two decomposition methods ICA and MSDL. Our goal is to compare the effects of region extraction on sparse and non-sparse sets of maps.

Figure 5.9 presents region extraction results using each method on the same map. In all figures, the threshold applied during region extraction is shown in a given slice to help understanding. Results for each metric are displayed on the right. We vary parameters for each model (smoothing for ICA, 3 parameters of MSDL) and, for each region extraction method, display the best 10% results across parametrization.

The regions extracted by hard assignment (Figure 5.9.a) present salient angles and their limits do not follow a contour line of the original map. The straight lines are the results of two maps in competition with each other. The 1D cut shows that the threshold applied when using hard thresholding is not uniform on the whole image. The other methods look smoother and follow actual contour lines of the original map. On this particular example, automatic thresholding (Figure 5.9.b) extracts 2 regions: a large one on the left and a very small one on the right. This is one of the drawbacks of thresholding: small regions can appear when their highest value is right above the threshold. Thanks to its high threshold, hysteresis thresholding (Figure 5.9.c) gets rid of the spurious regions but still fails to separate the large region on the left. Random Walker (Figure 5.9.d) manages to split the large region into two subregions.

Model selection plots The trends observed in Figure 5.12 are similar to the selection plots of the methods: NMI exhibits a Pareto efficiency front while Tanimoto coefficient exhibits a clear winner. In both plots, hard assignment

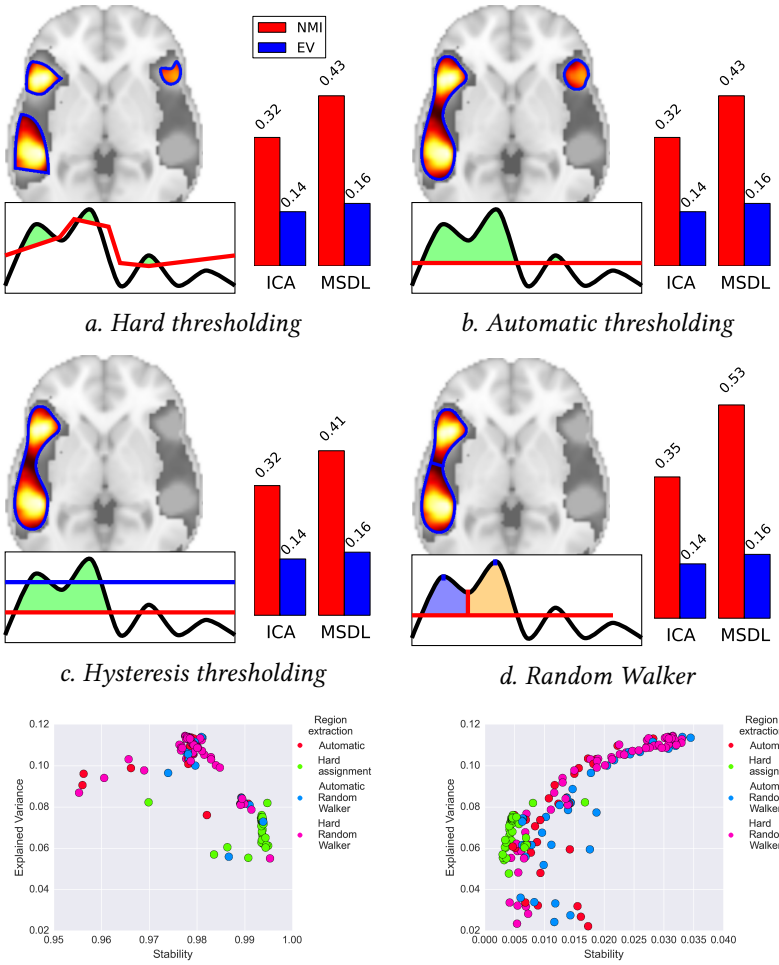


Figure 5.9: Comparison of region extraction methods (after selection of $2k$ regions). Brain maps obtained with MSDL are located on the left. The activated regions are symbolically represented below in a height map. The bars on the right of each image represent the Normalized Mutual Information and Explained variance obtained on dense maps (ICA) and sparse maps (MSDL). Random walker is the most stable method.

Figure 5.10: **Model selection plots for region extraction. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

gives poor results: it is clearly behind the other for Tanimoto coefficient. Automatic thresholding and Random Walkers cover approximately the same areas in the plot. However, hard Random Walker has a higher density in the area of high scores.

6 Validation for a prediction purpose

We use model selection metrics as heuristics to find the best model among several. In most cases, an optimal model can be determined from the values of the metrics. However, since our metrics are only heuristics, we have no guarantee that they actually select the best model.

In chapter 6, we present a prediction task we performed on the ABIDE dataset. We propose to use the classification results of this task as an exogeneous variable to validate our metrics.

Using model selection plots, we have seen that NMI and Tanimoto coefficient were in disagreement regarding the best parametrization to extract

raw maps. Figure 5.11 shows the model selection plot but adds the prediction accuracy as a third variable. While the best model is located on the Pareto efficiency frontier of NMI, Tanimoto coefficient completely fails at selecting the best model. This shows that measuring stability on raw maps is a hard problem and that in this situation, NMI is more reliable because it relies on co-occurences matrices. Please refer to Figure 7 at the end of this chapter for the full plots.

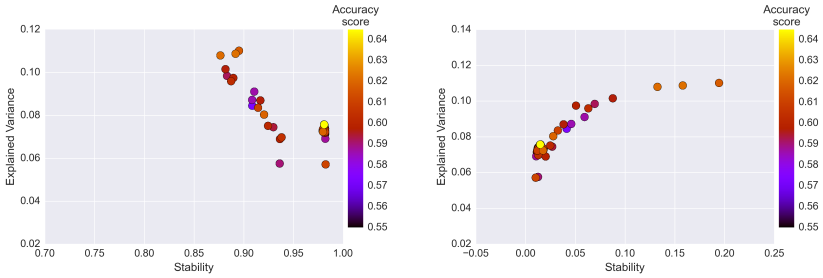


Figure 5.11: **Accuracy score of atlases depending on metrics scores. Left:** Normalized Mutual Information **Right:** Tanimoto coefficient

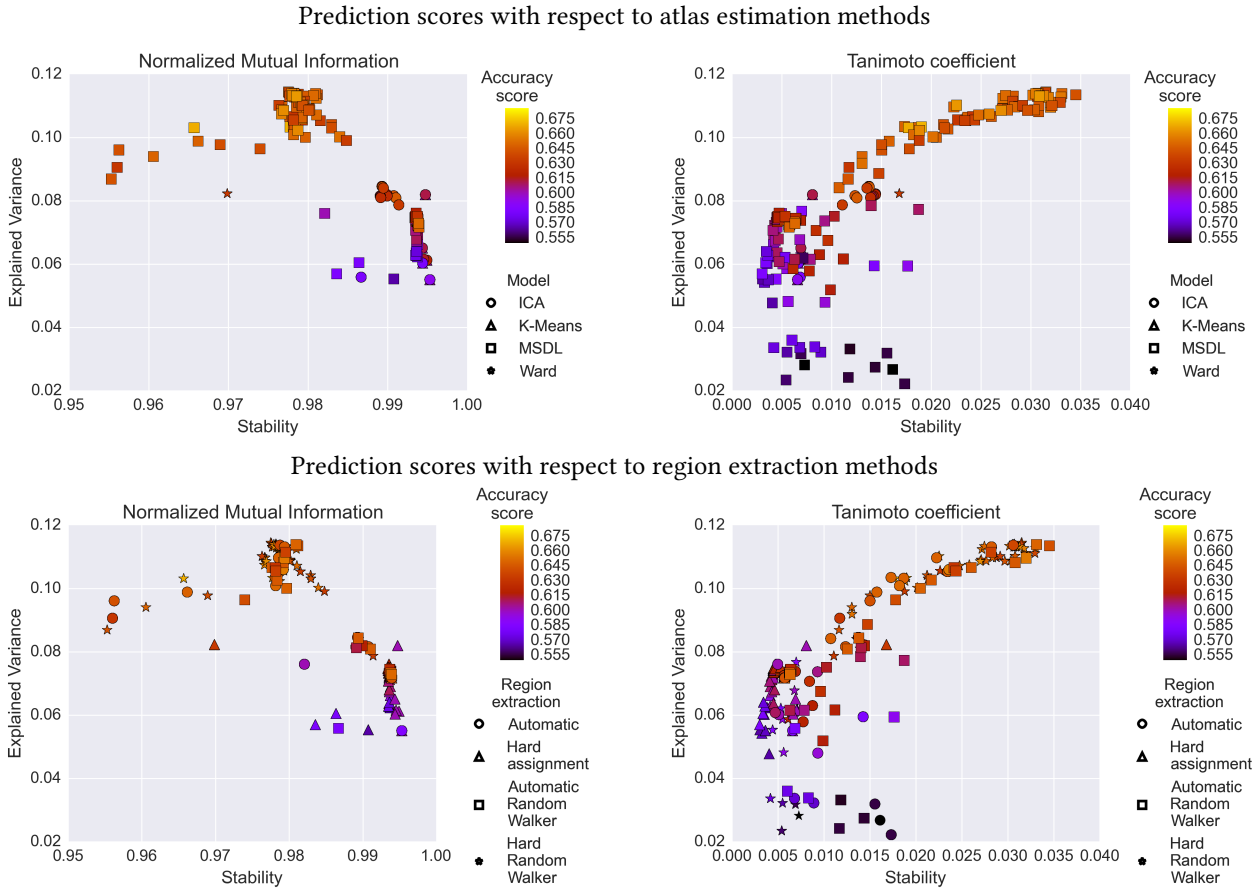
In the two other cases, namely atlas estimation method selection and region extraction selection, both metrics give satisfying results. As seen in Figure 5.12, both NMI and Tanimoto coefficient exhibit a gradient of increasing prediction accuracy. For Tanimoto coefficient, the gradient follows the metric scores: the best model according to our metrics is also the best predictor. For NMI, the gradient seems to be oriented toward the top left instead of top right. This confirms that NMI is not good at selecting models when they are of different nature.

7 Conclusion

Here, we looked for model-selection metrics to select the best brain atlas. This is paramount for resting-state fMRI analysis since no ground truth is available to validate the results.

We proposed metrics to measure the data faithfulness of a brain atlas, the Explained Variance, and two metrics to measure stability, based on the similarity metrics Normalized Mutual Information and Tanimoto Coefficient. We use them on atlases extracted using different methods and were able to use them to select the best model among several.

We validated our metrics using a prediction experiment. We found all these metrics to be useful to compare models of the same nature. When mixing models of different natures (different atlas extraction methods, or region extraction methods), the results obtained using Tanimoto Coefficient are more reliable (*i. e.* they are concordant with prediction scores and exhibit a clear trend instead of a Pareto efficiency frontier). The only exception are the raw maps: NMI still shows a Pareto efficiency front while Tanimoto coefficient gives bad scores to top performing models. It is worth noticing that, for both metrics, region extraction is beneficient to the measure of brain maps stability. This result may extend to other metrics and be valuable to measure the stability



of brain atlases out of our context.

Thanks to the model selection metrics, we were able to find a best parametrization for TV-MSDL. From our results, these metrics can be use to converge toward the best parametrization. We were also able to rank the different models and find out that matrix decomposition methods outperform clustering methods, TV-MSDL being the best one. We also ranked region extraction methods and determined that Random Walker maximized both stability (according to Tanimoto coefficient) and explained variance.

These metrics can be used for parameter selection but the best indicator in our case is experimental validation. In the following chapter, we present our prediction pipeline and apply it on ABIDE using a carefully designed cross-validation scheme to avoid overfitting. We will use the experimental results to revisit the results obtained using our metrics.

Figure 5.12: **Model selection plots for region extraction. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

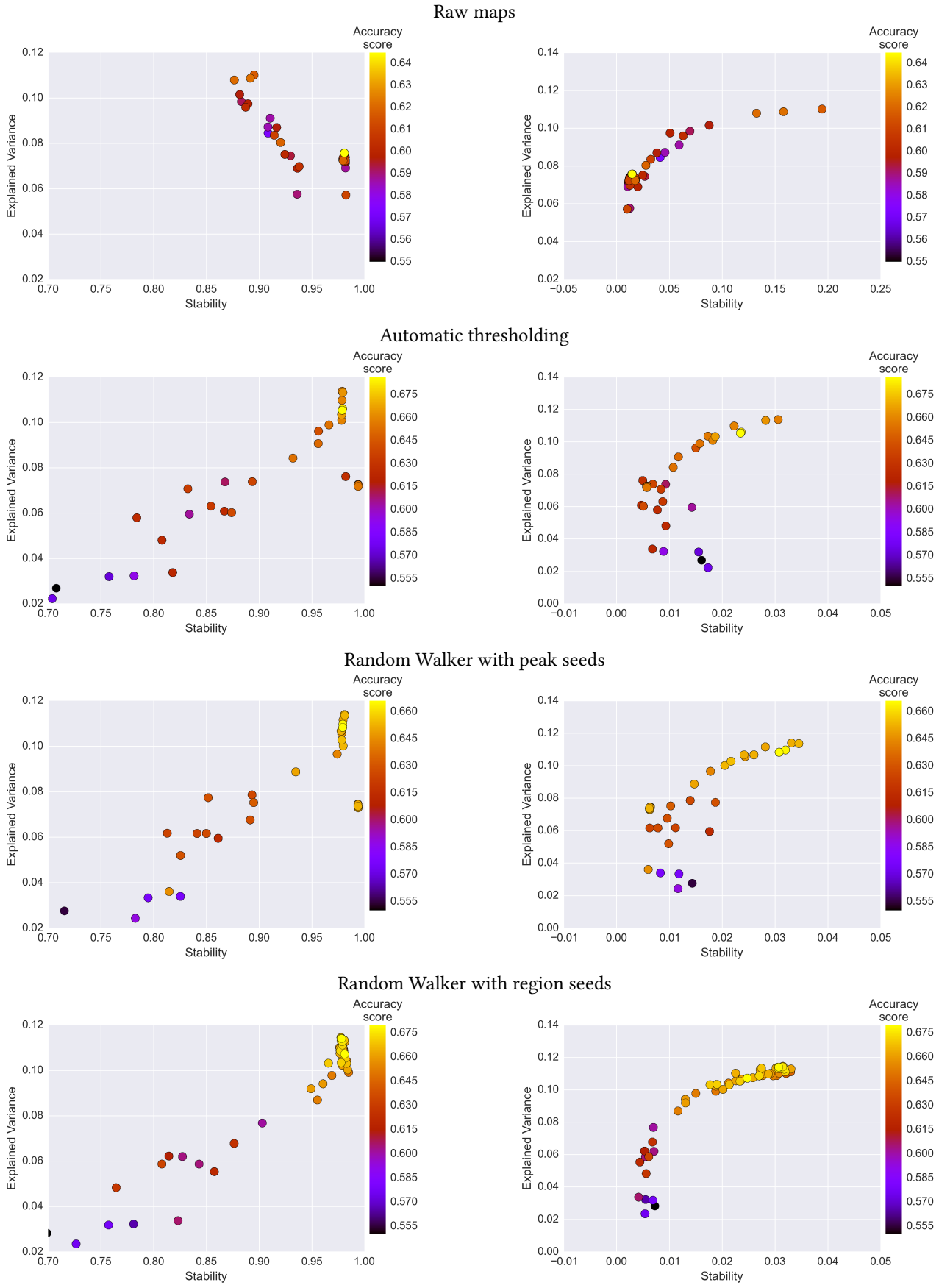


Figure 5.13: **Accuracy score of atlases depending on metrics scores. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

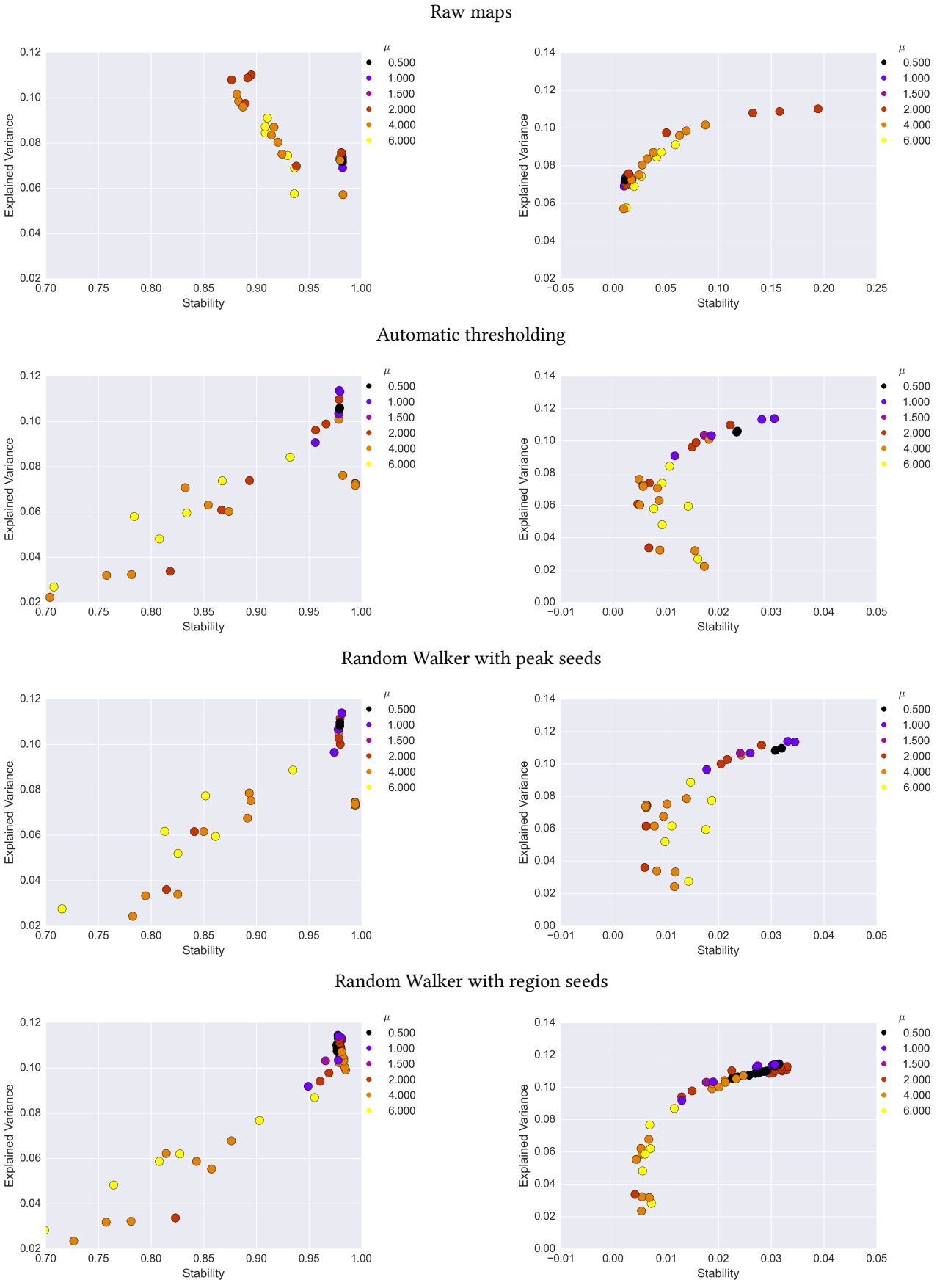


Figure 5.14: **Effects of parameter μ on model selection metrics. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

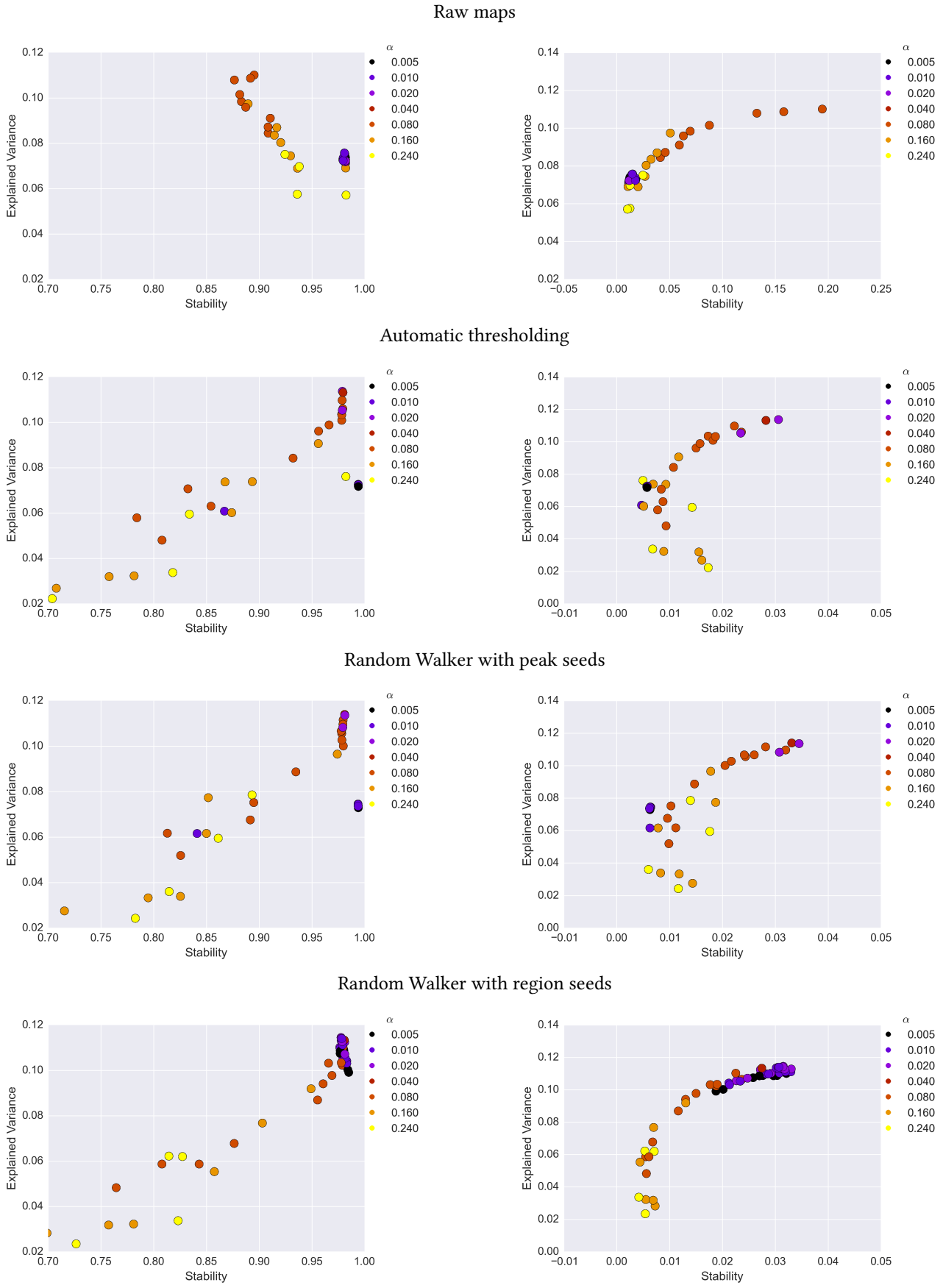


Figure 5.15: **Effects of parameter α on model selection metrics. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

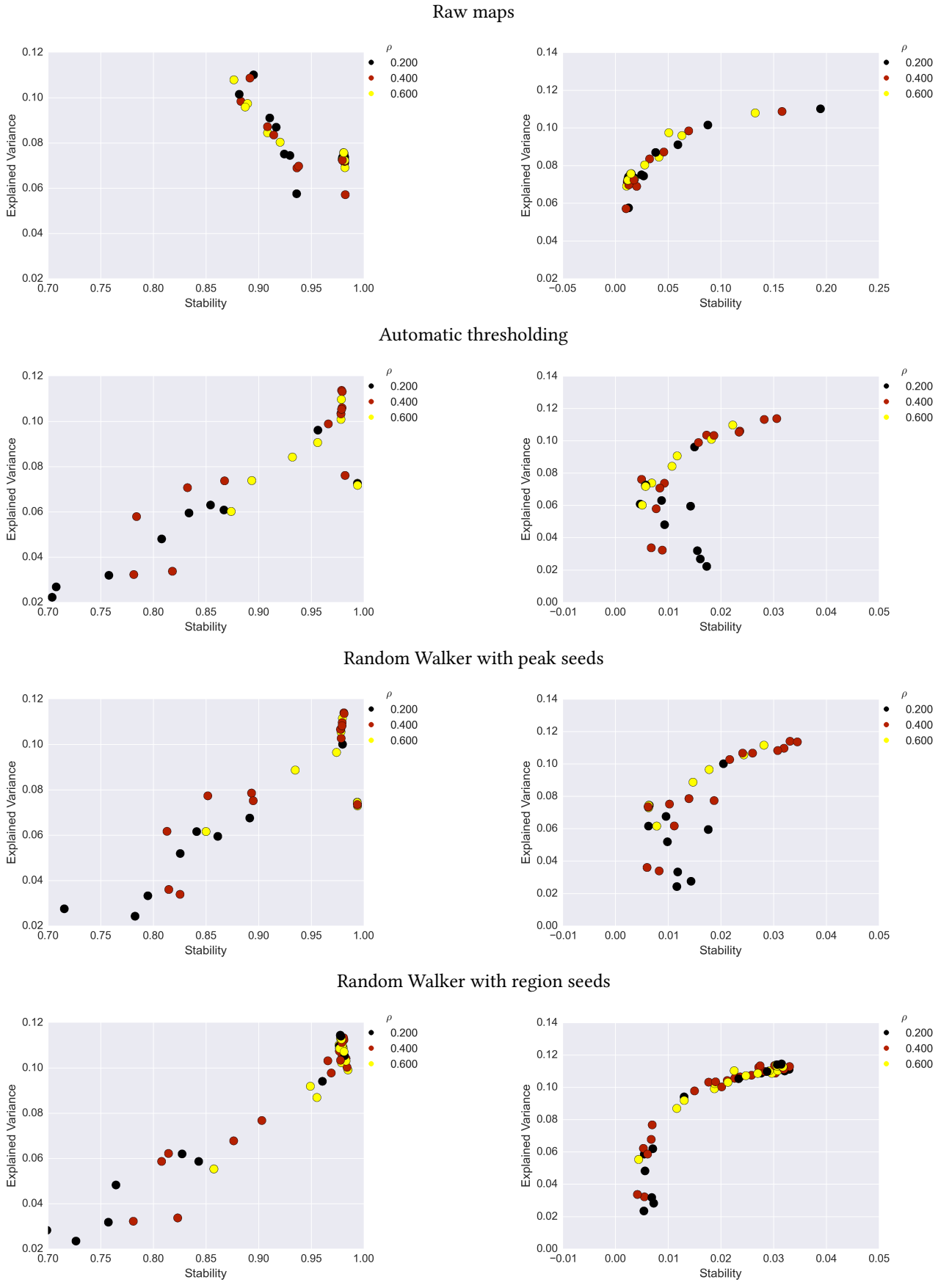
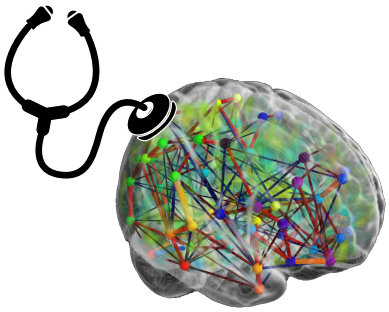


Figure 5.16: **Effects of parameter ρ on model selection metrics. Left:** Normalized mutual information. **Right:** Tanimoto coefficient.

6 | Learning predictive biomarkers of a psychiatric disease from resting state fMRI



In the previous chapters, we introduced a method to extract functional brain atlases from rest fMRI and applied it to an autism spectrum disorder (ASD) dataset, ABIDE. In this chapter, we use atlas estimation to learn predictive neurophenotypes –imaging biomarkers– from resting state fMRI data for ASD diagnosis. Such a neurophenotype is a particular connectivity pattern specific to subjects which exhibits behavioral differences of interest (Seeley et al., 2009).

We achieve prediction from fMRI data with a succession of steps. The first is the unsupervised altas extraction, followed by estimation of a functional connectome, which is then fed to a supervised predictor for diagnosis.

All the steps of this pipeline are fully automated: Model parameters are set by cross-validation. We introduce a new cross-validation scheme, closer to clinical settings, leaving a whole site out as testing set. In this chapter, we present the pipeline along with the classication results obtained on the ABIDE dataset. We achieve prediction accuracy on ASD beyond the state of the art. An extensive statistical analysis of each step of the pipeline is presented in the next chapter.

A short version of this chapter, along with the next one, has been submitted to the journal *NeuroImage*.

1 Introduction – Prediction from resting state functional MRI

Rest fMRI biomarkers have already been proven efficient for the diagnosis of Alzheimer disease (Greicius et al., 2004; Chen et al., 2011). For the Alzheimer disease, (Damoiseaux et al., 2012) even proposes an indicator for the progression of the disease. Likewise, the limbic system has been found impacted in

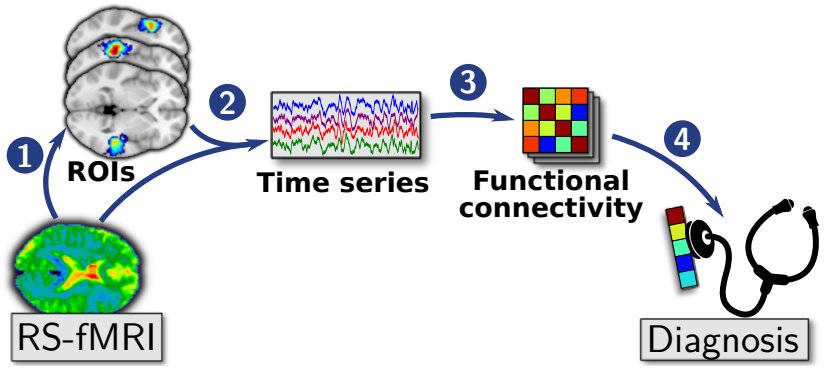


Figure 6.1: **Simple representation of the resting-state fMRI analysis pipeline.** Step 1 is the estimation of the atlas. Step 2 is the time series extraction. Step 3 is the estimation of the covariance matrix. Step 4 is the prediction of behavioral variable using the functional connectivity matrices computed in step 3.

depression (Craddock et al., 2009; Hamilton et al., 2011), functional hyperconnectivity has been observed in schizophrenic patients (Garrity et al., 2007; Zhou et al., 2007; Jafri et al., 2008; Calhoun et al., 2011), and more complex patterns in attention deficit hyperactivity disorder, autism and Down syndrome (Anderson et al., 2013).

Imaging neurophenotypes would help grounding psychiatric disease definition on brain traits rather than observed behavior. As mentioned in chapter 1, extracting such phenotypes is a challenging task because of the variability of scanning protocol and target population in the field work. In the particular case of ASD, it is even harder because the disease covers a wide variety of different symptoms.

The textbook approach to account for such inhomogeneities to use cross-validation, measuring how much models are able to generalize to unseen data. In this chapter, we propose a fully-automatic and parameter-free pipeline that relies internally on cross-validation to set its own parameters and perform a prediction task on large datasets.

Beyond the challenges associated with aggregated data, resting-state data comes with its own challenges, the largest being the lack of standard processing framework. As raw rest fMRI signals do not exhibit salient features, extracting a functional connectome requires multiple operations, as described in (Craddock et al., 2009; Richiardi et al., 2010; Shirer et al., 2012). The Conn toolbox [1] proposes an analysis pipeline for functional connectivity. Details of the data-processing pipelines vary widely in the literature (Carp, 2012), and impact strongly the resulting connectomes (Yan et al., 2013; Shirer et al., 2015). In our approach, we propose to integrate directly the atlas estimation as part of our pipeline.

To measure the generalization power of our approach, common practice relies on leave-out cross-validation strategies that remove single individuals (or random subsets), but this does not address the problem of site-specific artifacts. We also introduce a new cross-validation strategy, leaving sites out, where entire sites are used as test sets to measure the performance for sites not previously encountered in the training set. On the 871 subjects from

[1] Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2(3):125–141, 2012

Ref.	Subjects	Atlas	Results
(Nielsen et al., 2013)	964	7266 ROI	Leave one out classification 60%
(Alaerts et al., 2013)	584	pSTS	ABIDE used as validation set for task protocol
(Nebel et al., 2014)	868	Precentral gyrus	Signficant functional connectivity difference
(Plitt et al., 2014)	178	Power, Destrieux and Di Martino	75% classification accuracy
(Ray et al., 2014)	189	219 ROIs extracted from T1	Measures of graph properties
(Chen et al., 2015)	252	Power et al. (Power et al., 2011)	66% classification accuracy
(Glerean et al., 2015)	27	Voxel to voxel correlation	ABIDE used for reproducibility of results

Table 6.1: **Summary of studies using ABIDE dataset.** For each study, the number of individuals, regions and validation considered are listed along with the extracted biomarkers.

ABIDE dataset, this pipeline demonstrates the first inter-site prediction of neuropsychiatric disease.

As explained in section 4, we apply our pipeline to the classification between autistic and healthy individuals. Such task has already been explored in (Nielsen et al., 2013; Plitt et al., 2014; Chen et al., 2015). Nielsen *et al.*(Nielsen et al., 2013) presents classification results on most of ABIDE (964 participants) reaching 60% accuracy with leave-one-out cross validation. However, leave-one-out cross validation is overly optimistic on big datasets (Arlot et al., 2010). Other studies report better results albeit on smaller subsets of ABIDE such as (Plitt et al., 2014) that reaches 75% on 178 subjects or (Chen et al., 2015) that reports 66% on 252 subjects. Several studies discuss the corresponding neurophenotypes: impacted brain networks include default mode, language, attentional, and motor networks (see table 6.1).

In our experiment, we vary the inclusion criteria to validate the robustness of our approach on subsamples of different homogeneity. In the next chapter, we perform a post-hoc statistical analysis to study the importance of the modeling choices in the pipeline steps to guide future development.

2 Biomarkers extraction pipeline

Our approach –described in Figure 6.2– is composed of four *steps* for which we use different methods, or *options*. A *pipeline* is an instance of this approach with a specific choice of *options* for each step. We consider functional-connectivity maps (or components), composed of blob-shaped *brain regions* (or *regions or interests, ROIs*). A set of *ROIs*, extracted from functional-connectivity maps or not, form an *atlas*. Our approach extracts a *functional connectome* on a set of regions and trains a classifier across individuals using the edge weights of

this connectome (Varoquaux and Craddock, 2013). Unlike previous work, we integrate region learning to the pipeline (Figure 6.2). We set all the parameters using cross validation. For each step, several options are proposed. For the sake of clarity, and to save computational time, some options were excluded from the final study because of their poor results. Next chapter details the performance scores that motivated our choice.

The pipeline has been optimized to handle both the big size of the dataset (300GB) and the computational cost induced by the exploration of all the pipeline options. On a standard desktop computer (8 cores, 32GB RAM), computing one pipeline takes up to 4 hours and the whole computation of the two next chapters would take more than 10 years.

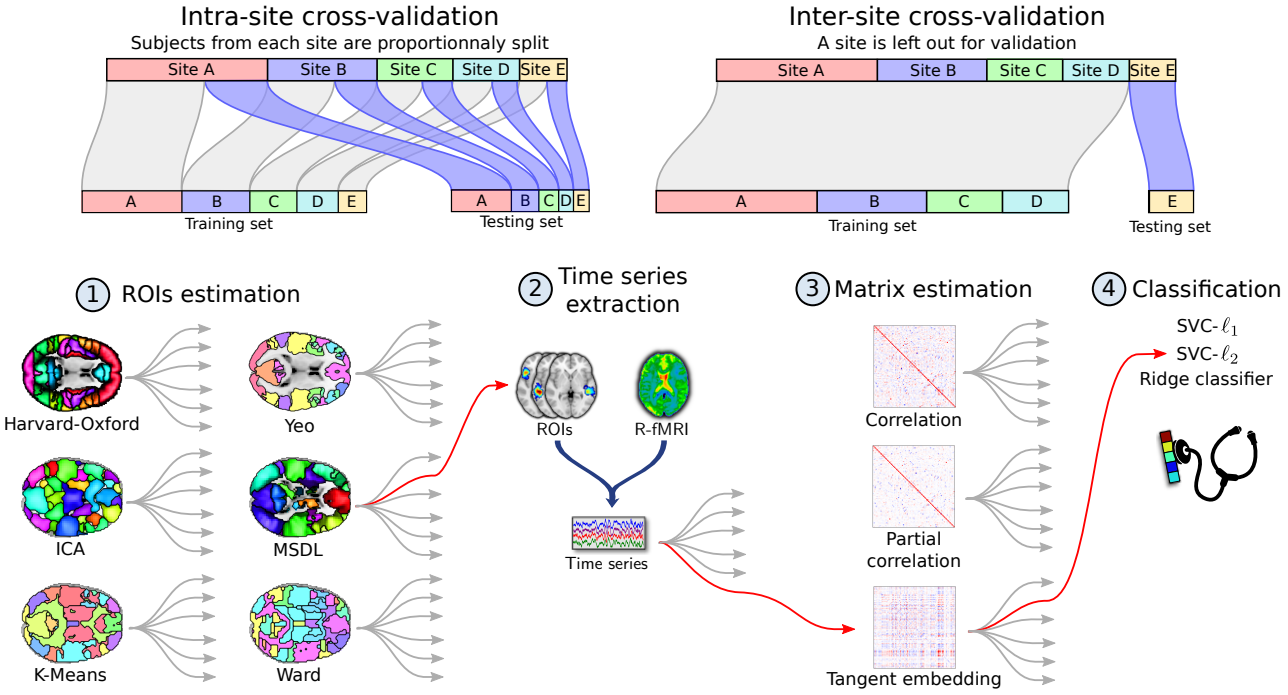


Figure 6.2: **Rest fMRI analysis pipeline.** Cross-validation schemes used to validate the pipeline are presented above. **Intra-site** cross-validation consists in randomly splitting the participants into training and testing sets while preserving the ratio of samples for each site and condition. **Intra-site** cross-validation consists in leaving out participants from an entire site as testing set. In the first step of the pipeline, ROIs are estimated from the training set. The second step consists in extracting signals of interest from all the participants, which are turned into connectivity features via a covariance estimation in the third step. These features are used in the fourth step to perform a

2.1 Step 1: Regions definition

Regions of the functional atlas can either be based on existing atlases or learned from the fMRI data. We investigate the state-of-the-art techniques to extract brain regions from rest fMRI listed in chapter 4.

Namely, we explore two clustering methods. **K-Means**, the historical technique in fMRI time series clustering (Goutte et al., 1999; Thirion et al., 2014), is a top-down approach seeking cluster assignments to minimize a variation dispersion criteria. **Ward's clustering** also seeks to minimize this criterion using agglomerative hierarchical clustering: Imposing a spatial constraint comes at no cost and has been extensively used to learn brain parcellations (Thirion et al., 2014). As Ward's clustering directly extracts blobs, we use it

directly to extract 84 ROIs. K-Means is used to extract 42 maps, from which 84 ROIs are extracted.

We also explore two decomposition methods. Independent component analysis (**ICA**) is the reference method to extract brain maps from rest fMRI data (Beckmann and Smith, 2004), seeking to maximize independence across maps. We use the ICA proposed in (Varoquaux et al., 2010c). Multi-subject dictionary learning (**MSDL**) is a decomposition method based on dictionary learning and a multi-subject learning strategy to extract a group atlas. It uses a spatial penalty to promote maps containing contiguous regions (Varoquaux et al., 2011; Abraham et al., 2013).

Unlike MSDL and Ward's clustering, ICA and K-Means do not enforce a spatial structure on their components. Applying a Gaussian smoothing on the data can inject such structure. We test a range of gaussian smoothing with full-width-at-half-maximum going from 4mm to 12mm.

In addition, we study reference functional and anatomical atlases: **Harvard Oxford** (Desikan et al., 2006), a structural atlas computed on T1 images from 36 individuals, **Yeo** (Yeo et al., 2011), a functional atlas computed using a clustering approach on a thousand individuals, and **Craddock** (Craddock et al., 2012), a functional atlas computed using constrained spectral clustering, exclusively used to study the impact of the number of regions on prediction.

2.2 Step 2: Time-series extraction

We extract one representative time series per ROI. For non-overlapping ROIs, we use the average signal of the voxels contained in each region. When brain regions overlap, as in decomposition models, we rely on least-squares regression that models the signal of one voxel as the mixture of signals from several regions. Common practice does not rely on overlapping functional maps because each region is believed to have a single functional role. However, we do not use this representation as function sharing between brain regions but as a way to compensate for normalization artifacts that grow with the number of subjects.

We regress out nuisance time series created from signals summarizing high-variance voxels (CompCor (Behzadi et al., 2007), with 5 principal components of the 2% voxels with highest variance). In addition, we extract the signals of the ROIs located in white matter and cerebrospinal fluid and regress them out to reject non-neural information (Varoquaux and Craddock, 2013).

2.3 Step 3: Connectivity matrix estimation

The length of the time series available in fMRI is insufficient to reliably compute the true covariance matrix for a given individual. We have to estimate it using a regularized covariance estimator. ℓ_2-regularized **shrinkage covariance** and ℓ_1-regularized **graphical lasso** are commonly used for this task and both have a single paramater. For the **shrinkage covariance**, **Ledoit-Wolf shrinkage**

estimator (Ledoit and Wolf, 2004) gives the parameter that optimizes the bias-variance trade-off. Cross-validation is necessary for the computation of graphical lasso parameter. In a preliminary study, all methods performed equally well. However, **Ledoit-Wolf shrinkage estimator** is the fastest and is parameter-free. To avoid a significant computational overhead, we have chosen to select only one option for this step.

From the estimated correlations we derive various connectivity measures to capture interactions between brain regions: *i)* the **correlations matrix**, *ii)* **partial correlations matrix** from the inverse covariance matrix (Varoquaux and Craddock, 2013; Smith et al., 2011), and *iii)* the **tangent embedding matrix** parametrization of the covariance matrix (Varoquaux et al., 2010a; Ng et al., 2014).

To compensate for additional variability, site, age and gender effects are regressed out from these connectivity coefficients. We excluded **covariances matrix** and **precisions matrix** (unnormalized counterparts of **correlation matrix** and **partial correlation matrix** resp.) from the study because of their insufficient performances.

2.4 Step 4: Supervised learning

We use the connectivity measures between all pairs of regions as features in a supervised learning step: we train a classifier to discriminate autistic from healthy patients. We explore support vector classification **SVC**, with an ℓ_1 or ℓ_2 penalization, **logistic regression**, with an ℓ_1 or ℓ_2 penalization, **ridge regression**, which also uses an ℓ_2 penalization and Gaussian Naive Bayes. We use the implementation of the models in Scikit-Learn (Pedregosa et al., 2011).

Gaussian Naive Bayes were excluded form the study because of their poor performance. We also removed **logistic regression** since its results were similar (although a bit lower) to **SVC**.

3 Experiments – Predicting autism spectrum disorders

In this section, we detail the experimental set up of the prediction task. We first detail the inclusion criteria of the different subsamples and also the cross-validation strategies.

3.1 Exploring Variability

Most previous studies have used less than a quarter of the ABIDE dataset. This is likely for computational reasons but also to reduce the variability due to the number of sites. In order to explore the effect of such choice on the results, we decided to apply our study not only on the whole ABIDE dataset but also on subsets of decreasing variability (see table 6.2): **all participants** (871 individuals) is the full set of individuals that passed QC after preprocessing, **biggest sites** (736 individuals) consists of the full dataset amputated from

Set	# Subj.	Sites	Selection criteria
All subjects	871	16	All subjects after QC
Biggest sites	736	11	Remove 6 smallest sites
Right handed males	639	14	All subjects without left handed and women
Right handed males, 9-18 yo	420	14	Right handed males between 9 and 18 years old
Right handed males, 9-18 yo, 3 sites	226	3	Young right handed males from 3 biggest sites

Table 6.2: **Subsets of ABIDE considered.** We explore several subsets of ABIDE with different homogeneity. *Card.* stands for the cardinality of the dataset, i.e. the number of subjects.

its 6 smallest sites, **right handed males** (639 individuals) consists of the full dataset with females, left handed participants, and small sites excluded (as in (Di Martino et al., 2014)), **right handed males, 9-18 yo** (420 individuals) is the previous group restricted to participants between 9 and 18 years old to reduce age-related variation, **right handed males, 9-18 yo, 3 sites** (226 individuals) are the individuals from the previous set belonging to the 3 biggest sites.

3.2 Cross Validation

Cross-validating an analysis consists in running it on a split of the dataset in two non-overlapping subsets: the training and the testing set. It is called a fold. The model to evaluate is then trained on then training set and is evaluated depending on its prediction score on the left out testing set. This procedure is repeated several times. Cross-validation serves two purposes: *i)* to find the best parameters of models and *ii)* to measure prediction performance, *e.g.* predicting labels on seen or unseen sites. The two cross-validation schemes used in this study are illustrated in Figure 6.2.

The first scheme, **intra-site prediction**, aims at picking training and testing sets as homogeneous as possible, in order to reduce the site variability. It is based on a cross-validation procedure called stratified shuffle split where subjects are split into training and test sets while preserving the ratio of samples for each site and condition. In our study, this means that the percentage of subjects from one site will always be the same in the training and the testing sets. The percentage of healthy and autistic participants in each set will be kept as well. In this study, we always use 80% of the dataset as training set and the remaining 20% are used as tests set. An illustration of this procedure can be found in Figure 6.2.

The second cross-validation scheme, **inter-site prediction**, reproduces clinical settings by using whole sites as testing sets. No further stratification is needed as the amount of autistic and healthy individuals in each site is balanced. However, due to the heterogeneity of the size of sites, the size of the testing set goes from 20% to 10% of the original dataset size. This cross-validation is not applicable on the subset of young right handed males from 3 biggest sites since we can only do 3 folds.

In each fold, region-extraction methods are used to learn maps independently of the testing set. For data-driven models, this step entails feature learning on up to 300 GB of data, which does not fit in memory. We reduce data size by applying a PCA dimensionality reduction for K-Means and ICA. We use a randomized SVD for computational efficiency (Martinsson et al., 2011). Using a stochastic coordinate descent approach (Abraham et al., 2013), MSDL can extract maps on a large number of subjects in a reasonable time (2 hours for 871 subjects).

4 Prediction accuracy in intra-site and inter-site settings

As shown in table 6.3, maximal performance (across pipelines) for each, intra-site and inter-site prediction approached 67.9% ($p < 0.01$, sensitivity 67%, specificity 69%) when the most inclusive subsample was examined (ABIDE subsample #1); this exceeds previously published findings with the ABIDE sample. While this result is insufficient to use fMRI as a diagnosis tool, a low score was expected since ASD cover a wide spectrum of phenotypes from fully functional individuals to patients not able to talk. We notice that intra- and inter-site prediction have similar scores. But, despite this apparent similarity of maximal performance, there are a number of noteworthy distinctions. First, across all training set sizes, the intra-site prediction strategy showed substantially lower variation ($p < 0.01$). Then, regarding the usage of more heterogenous subsamples, we did not find a significant effect of sample composition on performance.

Cross validation	Right handed males 9-18 yo, 3 sites	Right handed males 9-18 yo	Right handed males	Biggest sites	All subjects
Intra-site	66.6% ± 5.4%	65.8% ± 5.9%	65.7% ± 4.9%	67.9% ± 1.9%	66.9% ± 2.7%
Inter-site		69.7% ± 8.9%	65.1% ± 5.8%	68.7% ± 9.3%	66.8% ± 5.4%

Table 6.3: **Average accuracy scores (and standard deviation) for top performing pipelines**. This table summarizes the best results obtained for classification in for each atlas and subset using intra-site or inter-site prediction.

In Figure 6.3, the learning curves show that comparable performance between intra- and inter-site prediction strategies emerges when using more than 80% of the training set. Extended figures at the end of this chapter (see Figure 6) shows that below 278 participants in the training, the intra-site prediction is higher performing then inter-site ($p < 0.01$). Above, we don't find a statistical difference between the two.

It is worth noting, that regardless of which subsample was examined, we never observed plateaus in performance as training set size was increased. This implies that we do not reach optimal classification performance; our pipeline could benefit of additional subjects. Given a large enough training set, inter-site prediction competes with intra-site, albeit with higher variance.

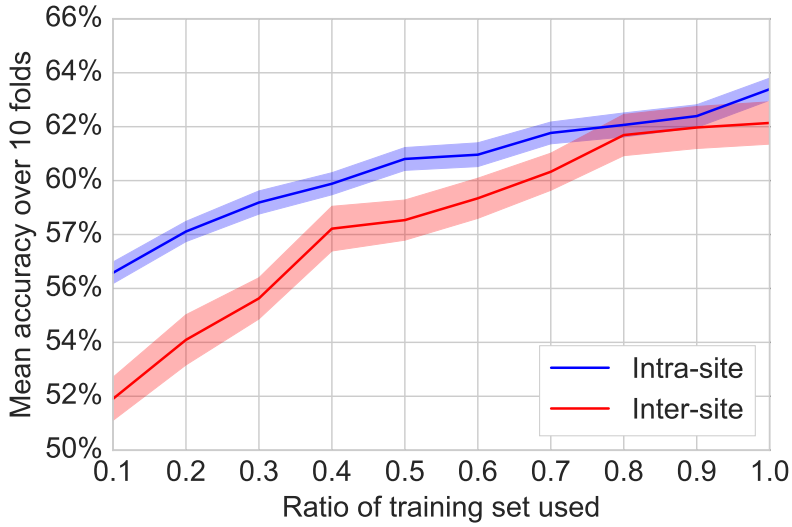


Figure 6.3: **Learning curves.** Classification results obtained by varying the number of subjects in the training set while keeping a fixed testing set. The colored band represent the standard error of the prediction. A score increasing with the number of subjects, *ie* an increasing curve, indicates that the addition of new subjects brings information.

5 Functional biomarkers

The wide exploration of the multiple options of the pipeline provides a large set of results that are analyzed in the next chapter. In this section, we analyze the observed differences of functional connectivity between ASD patients and controls.

5.1 Understanding the discriminant connections

Among all brain connections considered to diagnose subjects, some are more important than others. In order to find out which are the most important, it is common to consider the mean of correlation matrices or the weights of the classifier. In this study, we have decided to assess the importance of biomarkers using the statistical power induced by the high numbers of subjects and iterations. In particular, we run an identical pipeline on 10 different subsets of the same dataset and consider reliable the results that are repeated several times on different folds. Given its optimal results, we focused on the pipeline MSDL + tangent space embedding + SVC-ℓ_2.

Across the 10 folds, 37 ROIs are reproducible (DICE similarity score above 0.9). They cover all the gray matter at the exception of sensorimotor area and premotor cortex that does not satisfy our stability criterion. Since we use a linear classifier for the diagnosis, we are able to analyze the weights of the classifiers obtained across folds. We apply a statistical t-test on each of the weights and keep the significant ones.

5.2 Head motion

Efforts have been made to regress out the effect of movements in the data (regression of movement parameters, elimination of scan sessions when movements are too wide). It is however important to take into account that recent findings attribute differences of connectivity in longitudinal connections to head motion (Van Dijk et al., 2012). In fact, movements patterns are different between healthy and autistic patients given that the latter suffers from behavioral problems. In order to measure the size of this effect, we trained a classifier on the Fourier transform of the movement estimations made during preprocessing for all subjects in ABIDE and tried to discriminate autistic patients from healthy ones, following the same cross-validation pattern as our study. We reach the score of 62%, which is higher than state of the art results for connectivity studies, but lower than our highest scores. If movement patterns may be learned, they cannot explain all the results by themselves. However, it is likely that the difference of connectivity observed between frontal and parietal parts of the DMN is due to the effect of movements.

5.3 Autism spectrum disorders neurophenotypes

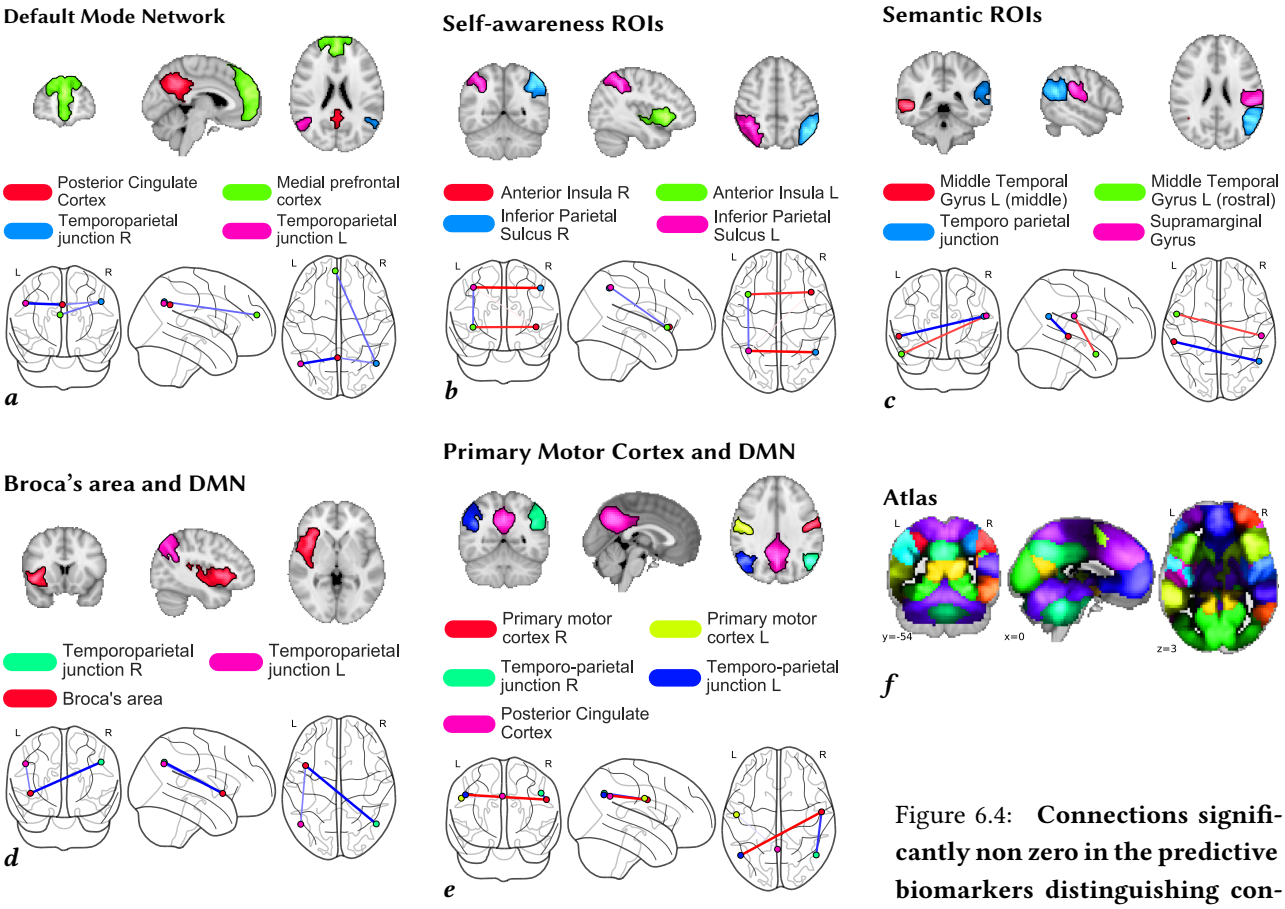


Figure 6.4: **Connections significantly non zero in the predictive biomarkers distinguishing controls from ASD patients** Red connections are stronger in controls and blue connections are stronger in ASD patients. Subfigures *a*, *b*, and *c* report intra-network differences and *d* and *e* inter-network differences. Below is the consensus atlas

Default Mode Network (DMN) (Figure 6.4.a) Connectivity in the DMN has been reported as different for several pathologies (Greicius, 2008) including ASD. Studies on adolescent or adult patients report an hypoconnectivity between regions of the DMN (Cherkassky et al., 2006; Kennedy et al., 2006) but also more complicated patterns with lower connectivity between distant regions of the DMN and stronger connectivity with close ones (Monk et al., 2009). A recent study on children (from 7 to 13 year old) reported an hyperconnectivity of the DMN (Supekar et al., 2013). Our results show a significant hyperconnectivity in the DMN, in particular between both temporoparietal junctions and PCC.

Self-awareness ROIs (Figure 6.4.b) We observe an interhemispheric hypoconnectivity between insulae and inferior parietal lobes. Those regions are part of the frontoparietal control network that is implicated in cognitive functions such as learning and memory (Iidaka et al., 2006). This hypoconnectivity has already been widely evoked in ASD (Anderson et al., 2010; Di Martino et al., 2011) and already found in the ABIDE dataset (Plitt et al., 2014; Ray et al., 2014). These ROIs also match anatomical regions of increased cortical density in ASD (Haar et al., 2014).

Semantic ROIs (Figure 6.4.c) Differences of laterality in language networks implying middle temporal gyrii have already been observed (Kleinhans et al., 2008). Although this symptom is often considered as decorrelated from ASD, we find that these regions are relevant for diagnosis.

Broca's area (Figure 6.4.d) Known for its role in language comprehension, Broca's area is implied is several other functions like action recognition and gestures. In our experiment a higher connectivity has been observed between this region and temporoparietal ROIs of DMN network in ASD patients.

Primary Motor Cortex (Figure 6.4.e) PMC has been suspected in the case of ASD because of the behavioral symptoms of the patients. As for Broca's areas, we observe that the connectivity between ROIs of the cortex is not directly disturbed. Their relationship with regions of the DMN is impacted. Right PMC's connectivities with temporoparietal junctions are disturbed: higher in the ROI of the same hemisphere and lower with the one in the opposite hemisphere. With a lower effect size, we also observe that the connectivity between left PMD and PCC is higher in ASD patients.

6 Conclusion

We have shown how to perform a prediction task using brain atlas estimation method, broached in the previous chapters, as first step. It is followed by the definition of a functional connectome upon the brain atlas and a prediction

using a discriminant model built on the corresponding connection values. Our pipeline is fully-automatic and parameter-free as it automatically sets the parameters of the models by using an internal cross-validation. By Applying it on the ABIDE multi-site autism dataset, we were able to perform a classification of autistic patients versus healthy ones and extract plausible biomarkers regarding literature on ASD. Our top prediction accuracy of 68%, which is not enough to use our pipeline for prediction, but was expected since ASD covers a wide range of symptoms which makes classification hard.

These functional results need to be validated by the classification scores obtained on ABIDE. An extensive analysis of these results is presented in the following chapter.

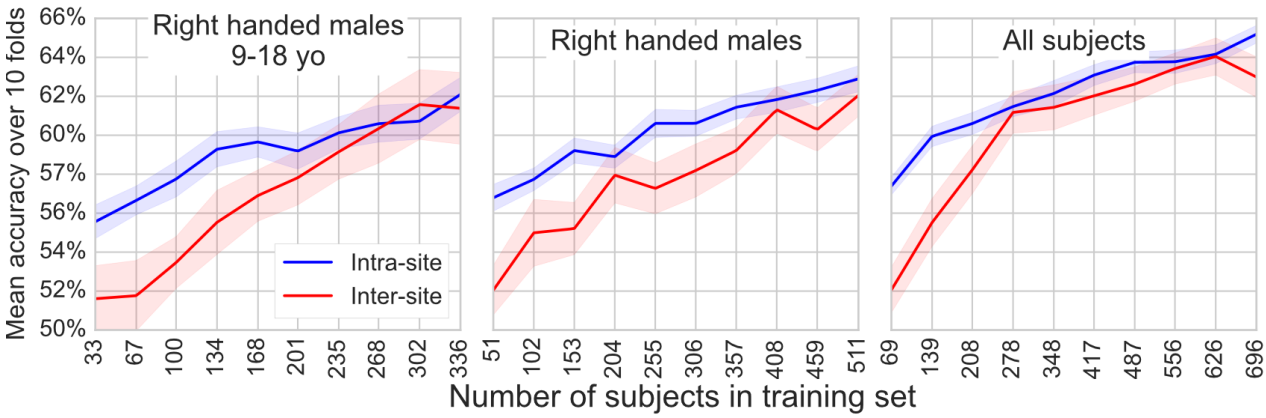


Figure 6.5: **Learning curves.** Classification results obtained by varying the number of subjects in the training set while keeping a fixed testing set. The colored band represent the standard error of the prediction. A score increasing with the number of subjects, *ie* an increasing curve, indicates that the addition of new subjects brings information.

7 | What's in a pipeline? Important choices for prediction

The pipeline presented in the previous chapter is composed of four steps. For each of these steps, several options are available. In this chapter, we answer two questions. Which step has the greatest impact on the prediction accuracy? For each step, which of the methods leads to the best accuracy? To answer these questions, we explored all the combinations of possible pipelines and ran a statistical analysis on the results.

This analysis shows that atlas estimation is the most important step of the pipeline, while the other steps are on par. We determine that the best pipeline is composed of TV-MSDL, tangent embedding covariance matrix estimation and SVC-ℓ_2.

The content of this chapter, along with the previous one, has been submitted to the journal *NeuroImage*. We present a deeper insight into partial results in this manuscript.

1 Post-hoc analysis of classification results

For statistical analysis purposes, we ran all possible pipelines and obtained a total of 1939400 pipelines. By performing a statistical analysis on these results, we determine the optimal settings for our prediction problem.

1.1 Full factorial analysis of variance

Our pipeline is composed of several steps for which several options are available. For each pipeline, we have the score of prediction on 10 folds for both cross-validation schemes and all subsamples.

Since pipeline step values are categorical and not quantitative, we use an analysis of variance (ANOVA) to determine if any option brings a statistically signficant improvement in the scores compared to the others. ANOVA compares the differences to the mean inside the groups to the difference between

the groups and gives a p-value for rejection of the null hypothesis that the results of the 2 groups are drawn from the same distribution.

Since we have several factors, we use the n-way ANOVA. This ANOVA is called *full factorial* because we explore all the combinations of all options for all the factors.

1.2 Optimal number of regions

In order to compare models of similar complexity, we restricted our study to 84 regions as in (Abraham et al., 2013). As this choice may not be optimal, we validate it with atlases from (Craddock et al., 2012), varying the number of regions (Figure 7.1).

For this aspect, we do an ANOVA with the following factors: the atlas used (Ward and Craddock atlases are used for this study to save computation time), the number of regions (10, 20, 40, 60, 80, 100, 140, 200), the matrix type (correlation, partial correlation, tangent embedding) and the predictor (SVC-ℓ_2, SVC-ℓ_1 and ridge classifier). We are interested only in region number.

The best performing models are located between 40 and 100 ROIs, with no clear winner. The score for 140 ROIs is very low. This is due to the instability of the results obtained on inter-site prediction above 100 regions. This trend can be observed when plotting this analysis for each subsample in the extended figure at the end of this chapter (Figure 7.8).

1.3 Impact of pipeline step choices

For this study, the ANOVA is realized by considering as factors the steps 1, 3, and 4 of the pipeline: *i)* the atlas extraction method (TV-MSDL, ICA, K-Means, Ward's clustering, Harvard-Oxford, and Yeo), *ii)* the covariance matrix type (correlation, partial correlation, and tangent embedding), *iii)* the predictor (SVC-ℓ_2, SVC-ℓ_1, and ridge classifier) In the ANOVA, we take as representative accuracy score the average of the 10% best performing models when varying model parameters for a pipeline. This yields scores stable and robust to potential outliers.

Figure 7.2 shows that the atlas estimation step is the most important of the pipeline and that MSDL dominates it clearly. Other data-driven methods are globally below average. Tangent embedding is overall the best covariance matrix type while correlation matrices dominate partial correlation in intra-site prediction only. Finally, we observe that ℓ_2-regularized classifiers (namely SVC-ℓ_2 and ridge classifier) outperform SVC-ℓ_1. An extended study of these effect for each subset is available at the end of this chapter(Figure 7.9). An interesting trend visible in these plots is that, when using inter-site cross-validation on a small set of subjects, data-driven methods generalize less well and reference atlases become top atlases for prediction.

Detailed pairwise comparisons (Figures 7.3, 7.5, and 7.7) confirm these trends. Results of intra-site prediction are more packed, confirming the higher

variability of inter-site prediction results (supplementary table 6.3).

2 Best atlas extraction method

Cross validation	Atlas estimator	Right handed males 9-18 yo, 3 sites	Right handed males 9-18 yo	Right handed males	Biggest sites	All subjects
Intra-site	MSDL	**66.6**% ± **5.4**%	**65.8**% ± **5.9**%	**65.7**% ± **4.9**%	**67.9**% ± **1.9**%	**66.9**% ± **2.7**%
	Yeo	60.9% ± 7.5%	62.3% ± 3.6%	64.7% ± 3.1%	64.7% ± 2.9%	66.9% ± 3.0%
	Harvard Oxford	63.6% ± 2.7%	62.1% ± 4.6%	64.8% ± 5.1%	64.8% ± 2.5%	66.4% ± 2.7%
	ICA	61.3% ± 8.7%	62.3% ± 3.0%	61.6% ± 4.3%	65.2% ± 3.1%	62.0% ± 3.4%
	K-Means	62.6% ± 5.9%	61.4% ± 5.4%	61.5% ± 3.0%	61.3% ± 3.4%	65.1% ± 1.8%
	Ward	63.2% ± 6.3%	60.1% ± 6.0%	62.2% ± 4.9%	64.6% ± 2.7%	63.7% ± 3.4%
Inter-site	MSDL		68.3% ± 7.6%	63.4% ± 6.3%	**68.7**% ± **9.3**%	**66.8**% ± **5.4**%
	Yeo		**69.7**% ± **8.9**%	64.5% ± 10.3%	61.4% ± 7.9%	61.3% ± 7.2%
	Harvard Oxford		68.1% ± 9.0%	**65.1**% ± **5.8**%	62.4% ± 5.4%	63.6% ± 6.2%
	ICA		63.1% ± 9.9%	62.5% ± 7.8%	65.0% ± 4.8%	60.9% ± 5.2%
	K-Means		62.8% ± 13.9%	61.5% ± 8.1%	61.9% ± 10.1%	60.3% ± 4.8%
	Ward		62.4% ± 11.7%	59.8% ± 6.9%	63.4% ± 5.4%	63.1% ± 4.0%

Table 7.1: **Average accuracy scores (and standard deviation) for top performing pipelines depending on atlas estimation**. This table summarizes the best results obtained for classification in for each atlas and subset using intra-site or inter-site prediction. Pipelines are sorted according to the atlas used. Best results are shown in bold.

A clear trend is that data-driven methods give best performers for high number of subjects, in particular for models with imposed spatial structure. The high performance of MSDL is probably due to its strong structured regularization. Its robustness to noise makes it a top performer even in the particular context of inter-site prediction on small datasets where other data-driven methods do not generalize well and thus reference atlases dominate.

We observe that the ranking of the methods observe in the best results matches the model selection plots: TV-MSDL is, by far, the best atlas and K-Means and ICA have close scores even if the latter is slightly better. The parameterization of TV-MSDL that yields the best result also matches the best values selected by our metrics: $\mu = 2, \alpha = 0.02, \rho = 0.6$.

Remaining site-specific artifacts cause instabilities that can be remedied by adding more subjects and sites. This means that inter-site analysis could become competitive with intra-site prediction given enough subjects and an optimal allocation of subjects across sites.

2.1 Overlapping regions predicts better

In section 2.2, we put forward the hypothesis that overlapping atlases are better at representing the signal, in particular for heterogeneous data. Based on our experiment, we compare predictions results obtained with overlapping atlases and their non overlapping counterpart (voxels belonging to several regions are assigned to the one with highest value).

Figure 7.4 shows that our hypothesis is confirmed. Overlapping atlases prediction results are higher ($p < 0.01$) than their non-overlapping counterpart.

Cross validation	Matrix type	Right handed males 9-18 yo, 3 sites	Right handed males 9-18 yo	Right handed males	Biggest sites	All subjects
Intra-site	Correlation	64.9% ± 7.6%	64.0% ± 3.0%	64.1% ± 3.7%	67.4% ± 3.6%	66.2% ± 3.8%
	Partial correlation	63.4% ± 7.3%	64.5% ± 4.0%	64.3% ± 3.4%	66.3% ± 2.3%	65.5% ± 2.2%
	Tangent embedding	**66.6**% ± **5.4**%	**65.8**% ± **5.9**%	**65.7**% ± **4.9**%	**67.9**% ± **1.9**%	**66.9**% ± **2.7**%
	Covariance	64.9% ± 7.6%	62.9% ± 4.0%	63.3% ± 2.1%		65.0% ± 2.8%
	Precision	63.8% ± 5.8%	63.8% ± 5.7%	64.3% ± 4.1%		64.3% ± 1.5%
Inter-site	Correlation		68.3% ± 7.6%	64.2% ± 7.6%	68.0% ± 7.1%	65.2% ± 10.1%
	Partial correlation		65.7% ± 5.9%	63.6% ± 6.1%	66.1% ± 4.8%	65.6% ± 5.7%
	Tangent embedding		**69.7**% ± **8.9**%	**65.1**% ± **5.8**%	**68.7**% ± **9.3**%	**66.8**% ± **5.4**%
	Covariance		68.3% ± 7.6%	64.7% ± 9.4%		64.7% ± 9.9%
	Precision		64.4% ± 5.6%	64.8% ± 6.1%		65.2% ± 7.1%

Table 7.2: **Average accuracy scores (and standard deviation) for top performing pipelines depending on matrix type**. This table summarizes the best results obtained for classification in for each atlas and subset using intra-site or inter-site prediction. Pipelines are sorted according to the matrix estimation step. Best results are shown in bold.

3 Best covariance matrix estimation

Tangent-space projection for correlation matrices gives significant better results. This enhancement comes at the cost of readability, because these matrices cannot be read as correlation matrices. Their results can however be directly interpreted (Varoquaux et al., 2010a). Correlation matrix performs better than partial correlation matrix in intra-site prediction, which may be due to the shortness of ABIDE time series.

3.1 Regression of behavioral confounds

Cross validation	Behavioral confounds	Right handed males 9-18 yo, 3 sites	Right handed males 9-18 yo	Right handed males	Biggest sites	All subjects
Intra-site	No confounds	66.0% ± 5.8%	**65.8**% ± **5.9**%	**65.7**% ± **4.9**%	**67.9**% ± **1.9**%	**66.9**% ± **2.7**%
	Site, age, gender	**66.6**% ± **5.4**%	65.3% ± 5.8%	65.3% ± 3.8%	67.2% ± 3.1%	66.3% ± 2.3%
Inter-site	No confounds		67.9% ± 8.4%	**65.1**% ± **5.8**%	**68.7**% ± **9.3**%	66.7% ± 7.2%
	Site, age, gender		**69.7**% ± **8.9**%	64.5% ± 10.3%	67.2% ± 6.4%	**66.8**% ± **5.4**%

Table 7.3: **Average accuracy scores (and standard deviation) for top performing pipelines depending on confounds regression**. This table summarizes the best results obtained for classification in for each atlas and subset using intra-site or inter-site prediction. Pipelines are sorted depending on the confound regression done on the functional connectivity matrices. Best results are shown in bold.

In pipeline step 3 (section 2.3), we indicate that we regress out age, site and gender from functional connectivity matrices. Removing site effect is logical as we know that the scanner and protocol used in the sites has an influence on the data. Age and gender regression is related to ASD: it affects mostly males and, because it is a developmental disorder, some patients tend to compensate for the symptoms above 18 years old.

Classification results (table 7.3) show no significant improvement of the confounds regression on the results. A closer look at several experiments (Figure 7.6) shows that confounds regression lowers slightly intra-site classification but increases the results of inter-site classification (probably because of site effect regression). As this is considered good practice in the litterature, we decided to apply confound regression systematically in our study.

Cross validation	Predictor	Right handed males 9-18 yo, 3 sites	Right handed males 9-18 yo	Right handed males	Biggest sites	All subjects
Intra-site	SVC-ℓ_1	61.1% ± 8.4%	61.8% ± 5.3%	62.1% ± 4.3%	64.3% ± 4.3%	63.1% ± 3.9%
	SVC-ℓ_2	66.0% ± 5.8%	65.6% ± 2.7%	**65.7% ± 4.9%**	**67.9% ± 1.9%**	**66.9% ± 2.7%**
	Ridge classifier	**66.6% ± 5.4%**	**65.8% ± 5.9%**	65.7% ± 4.2%	67.5% ± 3.0%	66.2% ± 3.8%
	LR-ℓ_1	63.0% ± 7.9%	59.4% ± 5.5%	61.4% ± 3.2%		63.7% ± 2.7%
	LR-ℓ_2	62.3% ± 5.8%	62.9% ± 4.5%	63.6% ± 3.3%		65.9% ± 2.8%
	Gaussian Naive Bayes	62.3% ± 4.8%	60.1% ± 4.4%	64.2% ± 4.6%		61.1% ± 4.4%
Inter-site	SVC-ℓ_1		67.6% ± 9.1%	64.4% ± 10.7%	66.4% ± 5.5%	63.0% ± 8.2%
	SVC-ℓ_2		**69.7% ± 8.9%**	64.5% ± 10.3%	**68.7% ± 9.3%**	**66.8% ± 5.4%**
	Ridge classifier		67.9% ± 8.4%	**64.8% ± 6.1%**	68.5% ± 9.0%	66.6% ± 5.6%
	LR-ℓ_1		68.1% ± 9.0%	64.2% ± 8.0%		62.4% ± 4.1%
	LR-ℓ_2		65.2% ± 6.3%	63.3% ± 5.6%		64.1% ± 6.9%
	Gaussian Naive Bayes		62.3% ± 8.9%	61.7% ± 9.0%		60.3% ± 9.1%

Table 7.4: **Average accuracy scores (and standard deviation) for top performing pipelines depending on predictor**. This table summarizes the best results obtained for classification in for each atlas and subset using intra-site or inter-site prediction. Pipelines are sorted according to the predictor used in the last step of the pipeline. Best results are shown in bold.

4 Best predictor

When predicting on covariance matrices, ℓ_2-regularized classifiers are the top performers in all settings. This may be due to global effects in the connectivity: A global hypoconnectivity, as already observed in ASD patients, cannot be captured by ℓ_1-regularized classifiers. In addition ℓ_2 penalization is *rotationally-invariant*, which means that it is not sensitive to arbitrary mixing of features. For instance, in our pipeline, imperfect choice of regions leads to a functional brain module being covered by several regions, which corresponds to such a mixing. Thus, a possible explanation for the good performance of ℓ_2-penalization is that it is less sensitive to the choice of regions.

5 Conclusion

In chapter 5, we proposed metrics to evaluate the quality of brain atlases. Using a prediction task, we propose here to evaluate any method of the pipeline by measuring its effect on the prediction scores.

We performed a statistical analysis on our results and drew several conclusions from that. First, we have seen the atlas extraction is the most important step of the pipeline and it is the step we should concentrate our efforts on. Then, across all pipelines, tangent embedding of covariance matrices and ℓ_2-regularized predictor were clear winners.

Finer trends also emerge from our results. In most settings, reference atlases estimated on other datasets perfoms as well or even better than data-driven methods. In particular, on datasets with a small number of subjects and high heterogeneity (right handed males subsample in inter-site prediction settings), reference atlas completely outperforms the best data-driven method. Finally, we have confirmed the trend observed in the previous chapter: inter-site prediction scores are on par with intra-site ones but with higher variance.

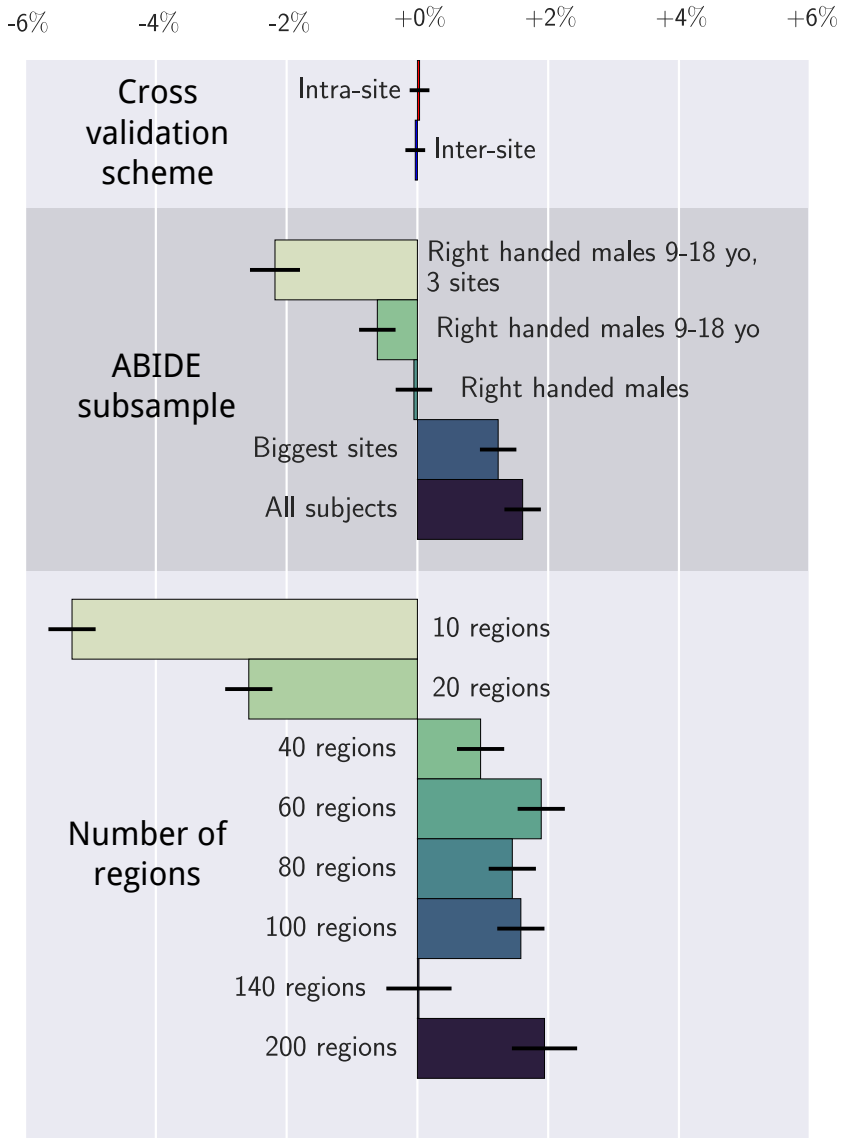


Figure 7.1: **Impact of the region number on prediction**: This plot represents the impact of the number of regions in the atlas on prediction accuracy for every subset of ABIDE and cross-validation schemes. The blocks of bars represent the cross-validation scheme, the subsets of ABIDE and the number of regions. Each bar represents the impact of the corresponding option on the prediction accuracy, relatively to the mean prediction. This effect is measured via a full-factorial ANOVA, analyzing the contribution of each step in a linear model. Each step of the pipeline is considered as a categorical variable. Error bars give the 95% confidence interval. We observe that the best number of regions is between 40 and 100 regions. The contrast between the scores obtained at 140 and 200 ROIs is due to an instability in the results.

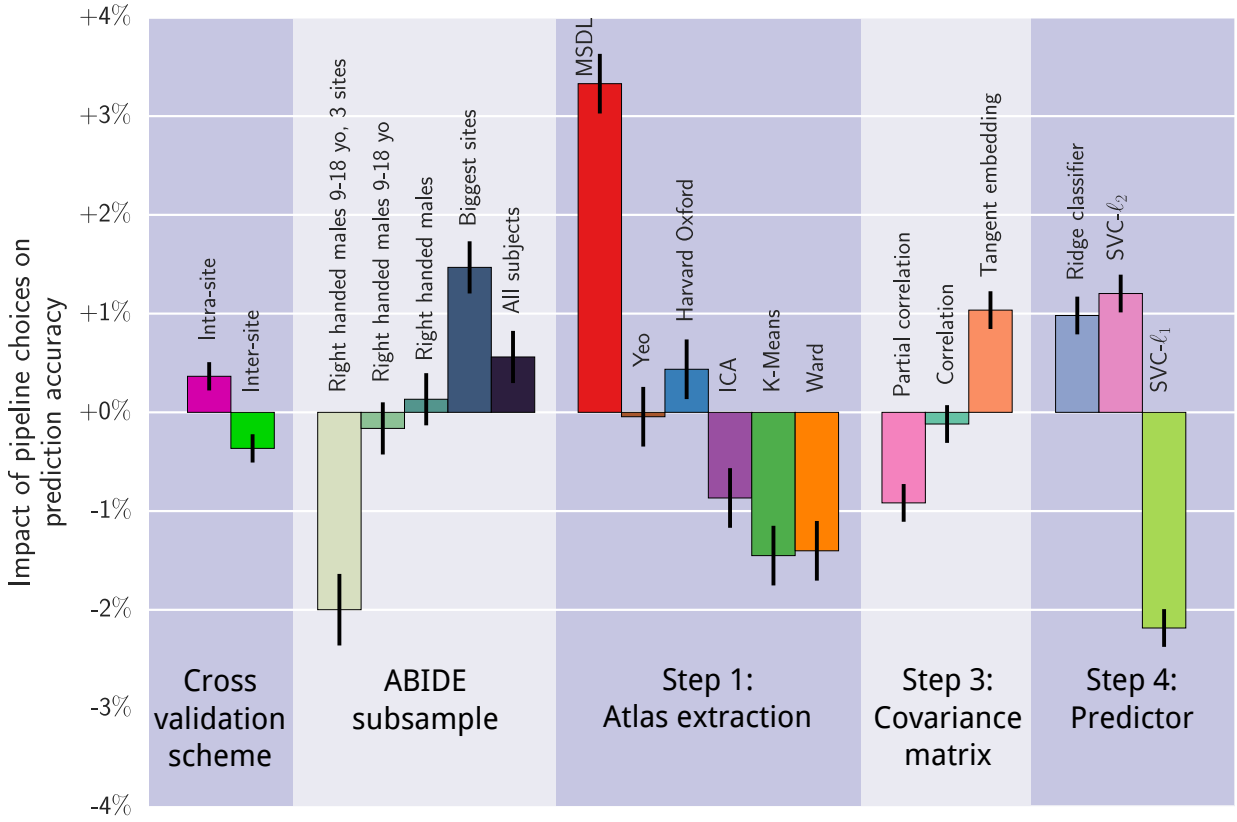


Figure 7.2: **Impact of pipeline steps on prediction.** This plot represents the impact of each step of the pipeline for every subset of ABIDE and cross-validation schemes. The blocks of bars represent the cross-validation scheme, the subsets of ABIDE and the different steps of the pipeline (namely step 1, 3 and 4). Each bar represents the impact of the corresponding option on the prediction accuracy, relatively to the mean prediction. This effect is measured via a full-factorial ANOVA, analyzing the contribution of each step in a linear model. Each step of the

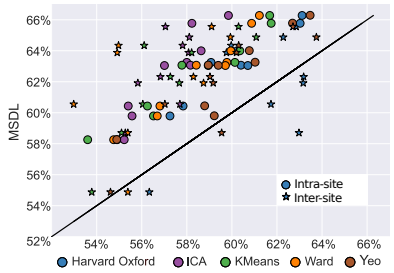


Figure 7.3: **Effect of atlas estimation on prediction accuracy.** Points above the line indicates runs where choosing TV-MSDL gives better accuracy than any another atlas estimation method. The 2 cross validation schemes are represented by stars and circles.

Figure 7.4: **Classification accuracy scores using overlapping atlases and their non overlapping counterpart.** Point above the identity line support the idea that fuzzy overlapping maps are better at predicting behavioral variable than non-overlapping.

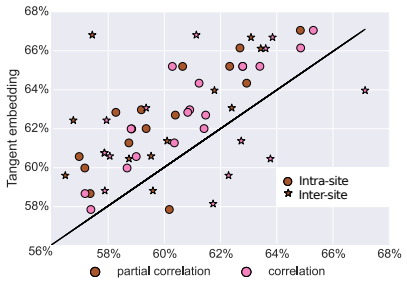


Figure 7.5: **Effect of covariance matrix type on prediction accuracy.** Points above the line indicates runs where choosing tangent embedding gives better accuracy than any other matrix types. The 2 cross validation schemes are represented by stars and circles.

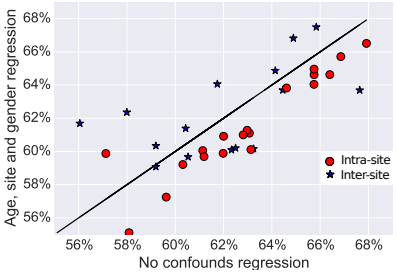


Figure 7.6: **Effect of confound regression on prediction accuracy.** Points above the line indicates runs where regressing confounds increases the prediction accuracy. The 2 cross validation schemes are represented by stars and circles.

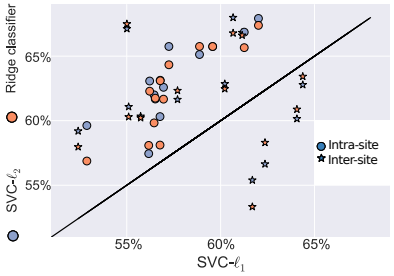


Figure 7.7: **Effect of predictor type on prediction accuracy.** Points above the line indicates runs where choosing a ℓ_2-regularized predictor outperforms SVC-ℓ_1. The 2 cross validation schemes are represented by stars and circles.

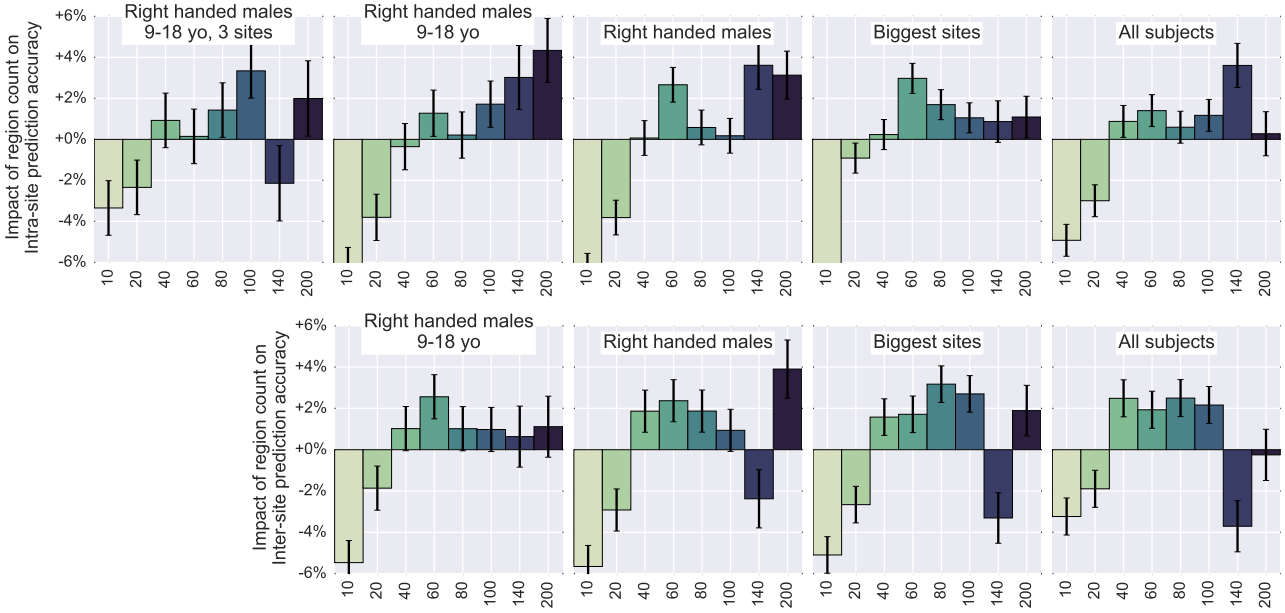


Figure 7.8: **Impact of the region number on prediction**: each bar indicates the impact of the number of regions on the prediction accuracy relatively to the mean of the prediction. These values are coefficients in a linear model explaining the best classification scores as function of the number of regions. Error bars give the 95% confidence interval, computed by a full factorial ANOVA. Atlases containing more than 40 ROIs give better results in all settings.

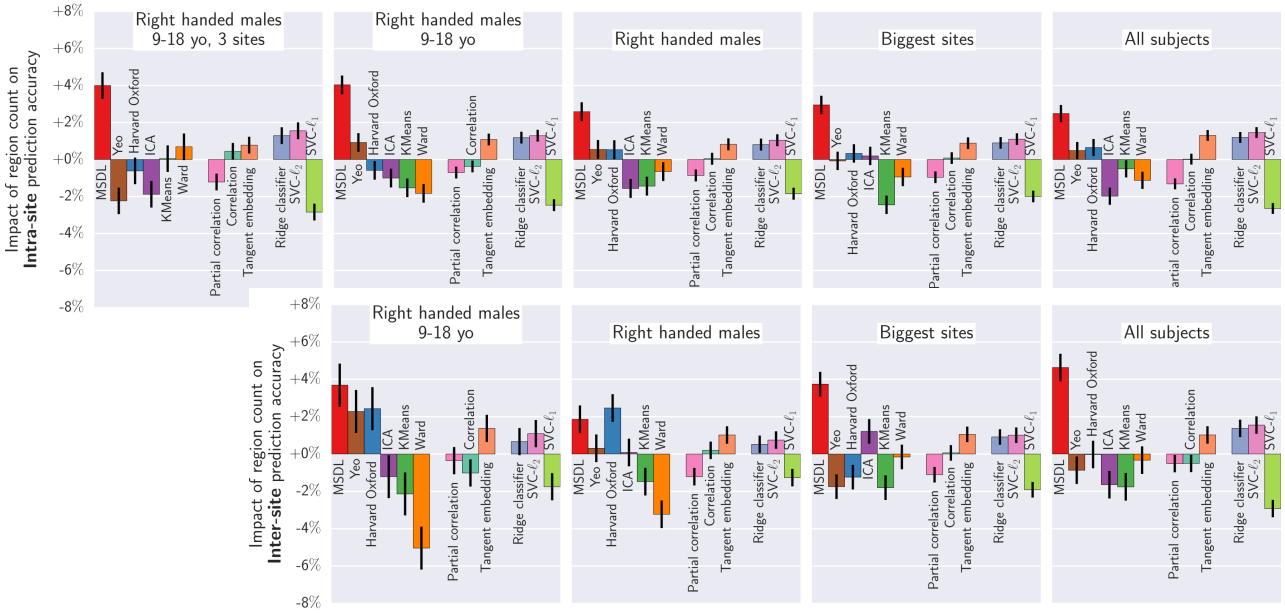


Figure 7.9: **Impact of pipeline steps on prediction.** Each plot represents the impact of each step of the pipeline for every subset of ABIDE. In each figure, a block of bars represents a step of the pipeline (namely step 1, 3 and 4). Each bar represents the impact of the corresponding option on the prediction accuracy, relatively to the mean prediction. This effect is measured via a full-factorial ANOVA, analyzing the contribution of each step in a linear model. Each step of the pipeline is considered as a categorical variable. Error bars give the 95% confidence interval. MSDL atlas extraction method gives significantly better results while reference atlases are slightly better than the mean. Among all matrix types, tangent embedding is the best on all ABIDE subsets. Finally, ℓ_2 regularized classifiers are better than ℓ_1 regularized.

8 | Conclusion

1 Contributions

Throughout this thesis, we worked on scaling up the classic rest fMRI pipeline to process big datasets. We proposed enhancements for the different steps of this procedure and proposed the tools required to evaluate them.

Structure imposing regularization for atlas extraction. We introduced a new atlas extraction approach that uses a sparse total-variation penalty with dictionary learning to combine the tendency of total-variation to create discrete spatial patches with the ability of linear decomposition models to unmix different effects. Careful choices of optimization strategy let our method scale to very large groups of subjects. The resulting regions are stable with respect to the subjects used for training, reveal a neurologically-plausible partition of the brain, and can give a synthetic representation of the resting-state correlation structure in a population.

Region extraction. Brain maps extracted using linear decomposition methods may cover the whole brain and contain uninformative voxels. We proposed an automatic thresholding strategy as an alternative to manual segmentation of the maps. We compared it to the automatic hard assignment. Because both strategies fail at splitting symmetric regions located on either side of the inter-hemispheric fissure, we used the random walker segmentation and showed that it outputs more stable atlases.

Model selection. We proposed to use similarity measures from the clustering field, namely Normalized Mutual Information and Tanimoto Coefficient, to evaluate the stability of atlas extraction methods. We also proposed to use explained variance to measure the ability of an atlas to account for the original variance of the signal. Using prediction score on a diagnosis task as target criterion, we validated these metrics as accurate to rank models depending on their prediction performance. In particular, we observed that region extraction improved the ability of stability metrics to rank atlases, and that Tanimoto coefficient was best to compare models of different nature.

Prediction pipeline. Based on the classical rest fMRI prediction procedure, we proposed a fully-automatic pipeline to predict behavioral data from rest fMRI. This pipeline integrates atlas extraction as a first step and sets its parameters using cross-validation. We performed autism spectrum disorders prediction on different subsets of the ABIDE dataset using all possible pipelines. In addition to the classical stratified cross-validation scheme (intra-site), we also predicted ASD on a completely left-over site to be closer to the clinical settings (inter-site).

A post-hoc analysis of prediction scores revealed that scores obtained in inter-site prediction are similar to intra-site but with higher variance, which is an encouraging step toward using rest fMRI in a clinical setting. We also established that region extraction is the most important step of the prediction pipeline and found out best choices for all the steps. Finally, our best pipeline obtains a prediction score above the state of the art.

Experiment reproducibility. We believe that reproducibility of experiments is important. As such, all the algorithms developed in this study are available on demand. Moreover, all the developped code is based on and will be released in the nilearn[1] (Abraham et al., 2014) Python package. Nilearn provides several tools for functional neuroimaging analysis. It features robust dataset downloading, fMRI data pre-treatment (smoothing, frequency filtering, etc.), specific fMRI approaches (Searchlight (Kriegeskorte et al., 2006), dictionary learning, group spatial ICA, etc.) and plotting of brain images.

[1] http://nilearn.github.io – This package is part of the contributions of my thesis. I am one of the main developpers and contributed 524 commits (23350 lines of code).

2 Perspectives

Cross-dataset reference atlas. In this study, we have shown that increasing the number of subjects in the training set was beneficial to the prediction accuracy. We have also stated that the optimal performance of the prediction pipeline was not reached and that it could benefit from more subjects, for example from the ABIDE 2 initiative. We have also shown that the additional variability induced by multi-site acquisition was not a show-stopper for prediction.

In continuation of the presents study, we should extend our analysis to multiple datasets. Big data initiative are rising in fMRI: ABIDE 2, Human Connectome Project, CORR dataset, Rockland sample, etc. In a first step, I would reproduce out experiment and statistical analysis on several datasets, on autism spectrum disorders but also on diseases more assessed than ASD. I think that we should also take advantage of the huge pool of healthy brains from typical controls to try and extract a *universal* atlas and test it on all possible studies.

New technical challenges may arise from the additional subjects and make TV-MSDL longer to converge. In order to keep atlas extraction doable in a reasonable time, algorithmic ways to optimize TV-MSDL are under consideration

such as random projections – to reduce the dimensionality of the subject-level problem – or bilateral filtering that could impose a regularization close to total-variation while being less computationally costly.

Refining atlas extraction. Despite its good performances, the TV-MSDL model can still be refined. Its multi-level approach allows one to impose sophisticated constraints while keeping a reasonable computation time.

Recent studies highlighted the hierarchical organization of the brain. This hypothesis is consistent with our study in the sense that regions can be seen as sub-divisions of a brain network. I think that enforcing a hierarchical organization in the dictionary learning could be beneficial for some problems.

Finer constraints could also be applied for specific problems. For example, studies of the basal ganglia are based on historical anatomical segmentations. Functional segmentation of this area is difficult because it is very noisy and the basal ganglia are small. A solution could be to rely on the fact that basal ganglia are supposed to be correlated to the activity of the cortex. Using a hierarchical approach, it would be possible to force the coupling of a cortical region with one located in the basal ganglia.

Multiple modalities. Our study only uses rest fMRI as modality. However, structural and functional connectivity are correlated (Honey et al., 2009) and could be used together to get better results. The versatility of our method (MSDL) makes it a good candidate for integration of other modalities – structural connectivity during subject maps estimation for example. We could also use modalities in other steps of the pipelines. Region extraction could also benefit from structural connectivity to delimit the regions. We could also take it into account in the connectome estimation. We did not explore this track because the ABIDE dataset does not offer diffusion tensor imaging modality from which structural connectivity is computed. Also, integrating a new modality may be the subject of a whole thesis.

Statistical validation. As stated in the introduction, validation is hard in rest fMRI. In this study, we propose two validation metrics to evaluate the stability and the ability to model fMRI data of extracted atlases. We validated them using prediction scores of a diagnosis task as surrogate criterion.

Exploration of autism spectrum disorders or other brain disorder. Our study is limited to one dataset and study ASD, a brain disorder that is hard to diagnose. Even though our prediction score (68%) is higher than state of the art results, it is still not enough to use functional connectivity as a diagnosis tool for this task. I think that exploring other prediction tasks on ABIDE, such as trying to find subtypes of patients (functional vs. non-functional patients) or finding neurophenotypes of people responding to a particular treatment,

could be useful. Predicting the severity of the symptoms is also an interesting prediction task.

Finally, to assess the validation of our process, reproducing this study on other diseases would be valuable.

Closing remarks. The capacity of rest fMRI as a diagnosis tool has remained questioned, in particular because of the small size of the datasets. This thesis opens the door of rest fMRI analysis on large heterogeneous datasets. We show that prediction is possible in a clinical setting and provide a pipeline to perform it. Although the prediction accuracy (65%-68%) is not sufficient to perform individual diagnosis, we have shown that our pipeline has not yet reached its maximum accuracy and can be improve by the addition of subjects in the training set. These results open a wide perspective for the future: with the technological breakthrough in acquisition methods and the acquisition of datasets of growing size, rest fMRI may become useful for clinical use.

Bibliography

Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, and Gael Varoquaux. Extracting brain regions from rest fMRI with Total-Variation constrained dictionary learning. In *MICCAI*, page 607. 2013.

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8, 2014.

Kaat Alaerts, Daniel G Woolley, Jean Steyaert, Adriana Di Martino, Stephan P Swinnen, and Nicole Wenderoth. Underconnectivity of the superior temporal sulcus predicts emotion recognition deficits in autism. *Social cognitive and affective neuroscience*, page nst156, 2013.

Andrew L Alexander, Jee Eun Lee, Mariana Lazar, Rebecca Boudos, Molly B DuBray, Terrence R Oakes, Judith N Miller, Jeffrey Lu, Eun-Kee Jeong, William M McMahon, et al. Diffusion tensor imaging of the corpus callosum in Autism. *Neuroimage*, 34(1):61–73, 2007.

David G Amaral, Cynthia Mills Schumann, and Christine Wu Nordahl. Neuroanatomy of autism. *Trends in neurosciences*, 31(3):137–145, 2008.

Jeffrey S Anderson, T Jason Druzgal, Alyson Froehlich, Molly B DuBray, Nicholas Lange, Andrew L Alexander, Tracy Abildskov, Jared A Nielsen, Annahir N Cariello, Jason R Cooperrider, et al. Decreased interhemispheric functional connectivity in autism. *Cerebral cortex*, page bhq190, 2010.

Jeffrey S Anderson, Jared A Nielsen, Michael A Ferguson, Melissa C Burback, Elizabeth T Cox, Li Dai, Guido Gerig, Jamie O Edgin, and Julie R Korenberg. Abnormal Brain Synchrony in Down Syndrome. *NeuroImage: Clinical*, 2013.

Sylvain Arlot, Alain Celisse, et al. A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79, 2010.

American Psychiatric Association et al. Diagnostic and statistical manual of mental disorders (4th ed., text rev.; DSM–IV–TR). *Washington, DC: American psychiatric association*, 2000.

American Psychiatric Association et al. Diagnostic and statistical manual of mental disorders (5th ed.; DSM–V). *Washington, DC: American psychiatric association*, 2013.

Luca Baldassarre, Janaina Mourao-Miranda, and Massimiliano Pontil. Structured Sparsity Models for Brain Decoding from fMRI data. In *PRNI*, pages 5–8, 2012.

Simon Baron-Cohen, Alan M Leslie, and Uta Frith. Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46, 1985.

Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *Trans Image Proc*, 18:2419–2434, 2009.

C. F. Beckmann and S. M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *Trans Med Im*, 23:137–152, 2004.

Y. Behzadi, K. Restom, J. Liau, and T.T. Liu. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, 37: 90, 2007.

Thomas Blumensath and Mike E Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27:265, 2009.

Thomas Blumensath, Timothy Behrens, and Stephen Smith. Resting-State FMRI Single Subject Cortical Parcellation Based on Region Growing. *MICCAI*, pages 188–195, 2012.

Nathalie Boddaert, Nadia Chabane, H Gervais, CD Good, M Bourgeois, MH Plumet, C Barthelemy, MC Mouren, E Artiges, Y Samson, et al. Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *Neuroimage*, 23(1):364–369, 2004.

Nathalie Boddaert, Mônica Zilbovicius, Anne Philipe, Laurence Robel, Marie Bourgeois, Catherine Barthélemy, David Seidenwurm, Isabelle Meresse, Laurence Laurier, Isabelle Desguerre, et al. MRI findings in 77 children with non-syndromic autistic disorder. *PLoS One*, 4(2):e4415, 2009.

Léon Bottou. Stochastic learning. *Lect Notes Artif Int*, pages 146–168, 2004.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press.

K Brodmann. Vergleichende Lokalisationslehre der Groshirnrinde. *Leipzig: Barth*, 1909.

Ariane VS Buescher, Zuleyha Cidav, Martin Knapp, and David S Mandell. Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA pediatrics*, 168(8):721–728, 2014.

Vince D Calhoun, Jing Sui, Kent Kiehl, Jessica Turner, Elena Allen, and Godfrey Pearlson. Exploring the psychosis functional connectome: aberrant intrinsic networks in schizophrenia and bipolar disorder. *Frontiers in psychiatry*, 2, 2011.

Qingjiu Cao, Yufeng Zang, Li Sun, Manqiu Sui, Xiangyu Long, Qihong Zou, and Yufeng Wang. Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. *Neuroreport*, 17(10):1033–1036, 2006.

Joshua Carp. The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage*, 63(1):289–300, 2012.

Manuel F Casanova, Imke AJ van Kooten, Andrew E Switala, Herman van Engeland, Helmut Heinsen, Harry WM Steinbusch, Patrick R Hof, Juan Trippe, Janet Stone, and Christoph Schmitz. Minicolumnar abnormalities in autism. *Acta neuropathologica*, 112(3):287–303, 2006.

Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. In M Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9, pages 263–340. 2010.

Colleen P Chen, Christopher L Keown, Afrooz Jahedi, Aarti Nair, Mark E Pflieger, Barbara A Bailey, and Ralph-Axel Müller. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *NeuroImage: Clinical*, 2015.

Gang Chen, B Douglas Ward, Chunming Xie, Wenjun Li, Zhilin Wu, Jennifer L Jones, Malgorzata Franczak, Piero Antuono, and Shi-Jiang Li. Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. *Radiology*, 259(1):213–221, 2011.

Vladimir L Cherkassky, Rajesh K Kana, Timothy A Keller, and Marcel Adam Just. Functional connectivity in a baseline resting-state network in autism. *Neuroreport*, 17(16):1687–1690, 2006.

John N Constantino and Christian P Gruber. The social responsiveness scale. *Los Angeles: Western Psychological Services*, 2002.

Eric Courchesne, CM Karns, HR Davis, R Ziccardi, RA Carper, ZD Tigue, HJ Chisum, P Moses, K Pierce, C Lord, et al. Unusual brain growth patterns in early life in patients with autistic disorder an MRI study. *Neurology*, 57 (2):245–254, 2001.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer, Xiaoping P Hu, and Helen S Mayberg. A whole brain fMRI atlas generated via spatially

constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

R.C. Craddock, P.E. Holtzheimer III, X.P. Hu, and H.S. Mayberg. Disease state prediction from resting state functional connectivity. *Magnetic resonance in Medicine*, 62:1619, 2009.

William R Crum, Oscar Camara, and Derek LG Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *Medical Imaging, IEEE Transactions on*, 25(11):1451–1461, 2006.

Jessica S Damoiseaux, Katherine E Prater, Bruce L Miller, and Michael D Greicius. Functional connectivity tracks clinical deterioration in Alzheimer's disease. *Neurobiology of aging*, 33(4):828–e19, 2012.

I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'Ardenne, W. Richter, J. D. Cohen, and J. Haxby. Independent component analysis for brain fMRI does not select for independence. *Proc Natl Acad Sci*, 106:10415, 2009.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, 31:968, 2006.

Adriana Di Martino, Clare Kelly, Rebecca Grzadzinski, Xi-Nian Zuo, Maarten Mennes, Maria Angeles Mairena, Catherine Lord, F Xavier Castellanos, and Michael P Milham. Aberrant striatal functional connectivity in children with autism. *Biological psychiatry*, 69(9):847–856, 2011.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Brian Egaas, Eric Courchesne, and Osamu Saitoh. Reduced size of corpus callosum in autism. *Archives of Neurology*, 52(8):794–801, 1995.

A.G. Garrity, G.D. Pearlson, K. McKiernan, D. Lloyd, K.A. Kiehl, and V.D. Calhoun. Aberrant" default mode" functional connectivity in schizophrenia. *Am J Psychiatry*, 164, 2007.

Enrico Glerean, Raj Kumar Pan, Juha Salmi, Rainer Kujala, Juha Lahnakoski, Ulrika Roine, Lauri Nummenmaa, Sami Leppämäki, Taina Nieminen-von Wendt, Pekka Tani, et al. Reorganization of functionally connected brain subnetworks in high-functioning autism. *arXiv preprint arXiv:1503.04851*, 2015.

Cyril Goutte, Peter Toft, Egill Rostrup, Finn Å Nielsen, and Lars Kai Hansen. On clustering fMRI time series. *NeuroImage*, 9(3):298–310, 1999.

Leo Grady. Random walks for image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(11):1768–1783, 2006.

Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Identifying predictive regions from fMRI with TV-L1 prior. In *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*, pages 17–20. IEEE, 2013.

M. Greicius. Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology*, 21:424, 2008.

M.D. Greicius, G. Srivastava, A.L. Reiss, and V. Menon. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proceedings of the National Academy of Sciences*, 101: 4637, 2004.

Shlomi Haar, Sigal Berman, Marlene Behrmann, and Ilan Dinstein. Anatomical abnormalities in autism? *Cerebral Cortex*, page bhu242, 2014.

J Paul Hamilton, Gang Chen, Moriah E Thomason, Mirra E Schwartz, and Ian H Gotlib. Investigating neural primacy in major depressive disorder: multivariate Granger causality analysis of resting-state fMRI time-series data. *Molecular psychiatry*, 16(7):763–772, 2011.

Heather Cody Hazlett, Hongbin Gu, Robert C McKinstry, Dennis WW Shaw, Kelly N Botteron, Stephen R Dager, Martin Styner, Clement Vachet, Guido Gerig, Sarah J Paterson, et al. Brain volume findings in 6-month-old infants at high familial risk for autism. *American Journal of Psychiatry*, 2012.

MR Herbert, DA Ziegler, CK Deutsch, LM O'brien, N Lange, A Bakardjiev, J Hodgson, KT Adrien, S Steele, N Makris, et al. Dissociations of cerebral cortex, subcortical and cerebral white matter volumes in autistic boys. *Brain*, 126(5):1182–1192, 2003.

C Honey, O Sporns, L Cammoun, X Gigandet, J Thiran, R Meuli, and P Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *P Natl Acad Sci Usa*, 106:2035, 2009.

Jane S Howard, Coleen R Sparkman, Howard G Cohen, Gina Green, and Harold Stanislaw. A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in developmental disabilities*, 26(4):359–383, 2005.

Tetsuya Iidaka, Atsushi Matsumoto, Junpei Nogawa, Yukiko Yamamoto, and Norihiro Sadato. Frontoparietal network involved in successful retrieval from episodic memory. Spatial and temporal analyses using fMRI and ERP. *Cerebral Cortex*, 16(9):1349–1360, 2006.

Madiha J Jafri, Godfrey D Pearlson, Michael Stevens, and Vince D Calhoun. A method for functional network connectivity among spatially independent resting-state components in schizophrenia. *Neuroimage*, 39(4):1666–1681, 2008.

Krister Järbrink. The economic consequences of autistic spectrum disorder among children in a Swedish municipality. *Autism*, 11(5):453–463, 2007.

R. Jenatton, G. Obozinski, and F. Bach. Structured sparse principal component analysis. In *Proc. AISTATS*, 2010.

Jutta Joormann and Ian H Gotlib. Updating the contents of working memory in depression: interference from irrelevant negative material. *Journal of abnormal psychology*, 117(1):182, 2008.

Marcel Adam Just, Vladimir L Cherkassky, Timothy A Keller, Rajesh K Kana, and Nancy J Minshew. Functional and anatomical cortical underconnectivity in autism: evidence from an FMRI study of an executive function task and corpus callosum morphometry. *Cerebral cortex*, 17(4):951–961, 2007.

Leo Kanner et al. *Autistic disturbances of affective contact*. publisher not identified, 1943.

Clare Kelly, Xi-Nian Zuo, Kristin Gotimer, Christine L Cox, Lauren Lynch, Dylan Brock, Davide Imperati, Hugh Garavan, John Rotrosen, F Xavier Castellanos, et al. Reduced interhemispheric resting state functional connectivity in cocaine addiction. *Biological psychiatry*, 69(7):684–692, 2011.

Daniel P Kennedy, Elizabeth Redcay, and Eric Courchesne. Failing to deactivate: resting functional abnormalities in autism. *Proceedings of the National Academy of Sciences*, 103(21):8275–8280, 2006.

V. Kiviniemi, T. Starck, J. Remes, et al. Functional segmentation of the brain cortex using high model order group PICA. *Hum Brain Map*, 30:3865–3886, 2009.

Natalia M Kleinhans, Ralph-Axel Müller, David N Cohen, and Eric Courchesne. Atypical functional lateralization of language in autism spectrum disorders. *Brain research*, 1221:115–125, 2008.

Matthieu Kowalski. Sparse regression using mixed norms. *Applied and Computational Harmonic Analysis*, 27:303–324, 2009.

Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3863–3868, 2006.

S. LaConte, J. Anderson, S. Muley, J. Ashe, S. Frutiger, K. Rehm, L.K. Hansen, E. Yacoub, X. Hu, and D. Rottenberg. The evaluation of preprocessing

choices in single-subject BOLD fMRI using NPAIRS performance metrics. *NeuroImage*, 18:10, 2003.

D. Lashkari, E. Vul, N. Kanwisher, and P. Golland. Discovering structure in the space of fMRI selectivity profiles. *NeuroImage*, 50:1085, 2010.

Ann Le Couteur, Michael Rutter, Catherine Lord, Patricia Rios, Sarah Robertson, Mary Holdgrafer, and John McLennan. Autism diagnostic interview: A standardized investigator-based instrument. *Journal of autism and developmental disorders*, 19(3):363–387, 1989.

O Ledoit and M Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.*, 88:365, 2004.

D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

Aline Lefebvre, Anita Beggiato, Thomas Bourgeron, and Roberto Toro. Neuroanatomical diversity of corpus callosum and brain volume in autism: meta-analysis, analysis of the Autism Brain Imaging Data Exchange project, and simulation. *Biological psychiatry*, 2015.

Haihong Liu, Zhening Liu, Meng Liang, Yihui Hao, Lihua Tan, Fan Kuang, Yanhong Yi, Lin Xu, and Tianzi Jiang. Decreased regional homogeneity in schizophrenia: a resting state functional magnetic resonance imaging study. *Neuroreport*, 17(1):19–22, 2006.

Yong Liu, Kun Wang, YU Chunshui, Yong He, Yuan Zhou, Meng Liang, Liang Wang, and Tianzi Jiang. Regional homogeneity, functional connectivity and imaging markers of Alzheimer's disease: a review of resting-state fMRI studies. *Neuropsychologia*, 46(6):1648–1656, 2008.

Catherine Lord. Follow-up of two-year-olds referred for possible autism. *Child Psychology & Psychiatry & Allied Disciplines*, 1995.

Catherine Lord. Unweaving the autism spectrum. *Cell*, 147(1):24–25, 2011.

Catherine Lord, Michael Rutter, Susan Goode, Jacquelyn Heemsbergen, Heather Jordan, Lynn Mawhood, and Eric Schopler. Austism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of autism and developmental disorders*, 19(2):185–212, 1989.

Catherine Lord, Edwin H Cook, Bennett L Leventhal, and David G Amaral. Autism spectrum disorders. *Neuron*, 28(2):355–363, 2000.

Linda J Lotspeich, Hower Kwon, Cynthia M Schumann, Susanna L Fryer, Beth L Goodlin-Jones, Michael H Buonocore, Cathy R Lammers, David G Amaral, and Allan L Reiss. Investigation of Neuroanatomical Differences

Between Autism and AspergerSyndrome. *Archives of general psychiatry*, 61 (3):291–298, 2004.

Per Martinsson, Vladimir Rokhlin, and Mark Tygert. A randomized algorithm for the decomposition of matrices. *Applied and Computational Harmonic Analysis*, 30:47, 2011.

V. Michel, A. Gramfort, G. Varoquaux, E. Eger, and B. Thirion. Total Variation Regularization for fMRI-Based Prediction of Behavior. *Trans Med Imag*, 30: 1328–1340, 2011.

Christopher S Monk, Scott J Peltier, Jillian Lee Wiggins, Shih-Jen Weng, Melisa Carrasco, Susan Risi, and Catherine Lord. Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *Neuroimage*, 47(2):764–772, 2009.

Mary Beth Nebel, Ani Eloyan, Anita D Barber, and Stewart H Mostofsky. Precentral gyrus functional connectivity signatures of autism. *Frontiers in systems neuroscience*, 8, 2014.

Bernard Ng, Martin Dressler, Gaël Varoquaux, Jean Baptiste Poline, Michael Greicius, and Bertrand Thirion. Transport on Riemannian Manifold for Functional Connectivity-based Classification. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 405–412. Springer International Publishing, 2014.

Jared A Nielsen, Brandon A Zielinski, P Thomas Fletcher, Andrew L Alexander, Nicholas Lange, Erin D Bigler, Janet E Lainhart, and Jeffrey S Anderson. Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in human neuroscience*, 7, 2013.

Alfonso Nieto-Castanon, Satrajit S Ghosh, Jason A Tourville, and Frank H Guenther. Region of interest based analysis of functional imaging data. *Neuroimage*, 19(4):1303–1316, 2003.

Jyri-Johan Paakki, Jukka Rahko, Xiangyu Long, Irma Moilanen, Osmo Tervonen, Juha Nikkinen, Tuomo Starck, Jukka Remes, Tuula Hurtig, Helena Haapsamo, et al. Alterations in regional homogeneity of resting-state brain activity in autism spectrum disorders. *Brain research*, 1321:169–179, 2010.

Saskia JMC Palmen, Hilleke E Hulshoff Pol, Chantal Kemner, Hugo G Schnack, Sarah Durston, Bertine E Lahuis, René S Kahn, and Herman Van Engeland. Increased gray-matter volume in medication-naive high-functioning children with autism spectrum disorder. *Psychological medicine*, 35(04):561–570, 2005.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825, 2011.

Darius Pfitzner, Richard Leibbrandt, and David Powers. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19(3):361–394, 2009.

Joseph Piven, James Bailey, Bonnie J Ranson, and Stephan Arndt. An MRI study of the corpus callosum in autism. *American Journal of Psychiatry*, 154 (8):1051–1056, 1997.

Mark Plitt, Kelly Anne Barnes, and Alex Martin. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage: Clinical*, 2014.

Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.

Jonathan D Power, Kelly A Barnes, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage*, 59(3): 2142–2154, 2012.

M.E. Raichle. Two views of brain function. *Trends in cognitive sciences*, 14:180, 2010.

Siddharth Ray, Meghan Miller, Sarah Karalunas, Charles Robertson, David S Grayson, Robert P Cary, Elizabeth Hawkey, Julia G Painter, Daniel Kriz, Eric Fombonne, et al. Structural and functional connectivity of the human brain in autism spectrum disorders and attention-deficit/hyperactivity disorder: A rich club-organization study. *Human brain mapping*, 35(12):6032–6048, 2014.

J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville. Decoding brain states from fMRI connectivity graphs. *NeuroImage*, 2010.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1): 259–268, 1992.

W.W. Seeley, R.K. Crawford, J. Zhou, B.L. Miller, and M.D. Greicius. Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron*, 62:42, 2009.

William R Shirer, Heidi Jiang, Collin M Price, Bernard Ng, and Michael D Greicius. Optimization of rs-fMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination. *NeuroImage*, 2015.

WR Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and MD Greicius. Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*, 22:158, 2012.

S.M. Smith, P.T. Fox, K.L. Miller, D.C. Glahn, P.M. Fox, C.E. Mackay, et al. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci*, 106:13040, 2009.

S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, T.E. Nichols, J.D. Ramsey, and M.W. Woolrich. Network modelling methods for fMRI. *Neuroimage*, 54:875, 2011.

Sara S Sparrow, Domenic V Cicchetti, and David A Balla. The vineland adaptive behavior scales. *Major psychological assessment instruments*, 2:199–231, 1989.

Wendy L Stone, Evon B Lee, Linda Ashford, Jane Brissie, Susan L Hepburn, Elaine E Coonrod, and Bahr H Weiss. Can autism be diagnosed accurately in children under 3 years? *Journal of Child Psychology and Psychiatry*, 40 (02):219–226, 1999.

S.C. Strother, J. Anderson, L.K. Hansen, U. Kjems, R. Kustra, J. Sidtis, S. Frutiger, S. Muley, S. LaConte, and D. Rottenberg. The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage*, 15:747, 2002.

Kaustubh Supekar, Lucina Q Uddin, Amirah Khouzam, Jennifer Phillips, William D Gaillard, Lauren E Kenworthy, Benjamin E Yerys, Chandan J Vaidya, and Vinod Menon. Brain hyperconnectivity in children with autism and its links to social deficits. *Cell reports*, 5(3):738–747, 2013.

Bertrand Thirion, Gaël Varoquaux, Elvis Dohmatob, and Jean-Baptiste Poline. Which fMRI clustering gives good brain parcellations? *Frontiers in neuroscience*, 8, 2014.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15:273, 2002.

K.R.A. Van Dijk, M.R. Sabuncu, and R.L. Buckner. The influence of head motion on intrinsic functional connectivity MRI. *Neuroimage*, 59:431, 2012.

G. Varoquaux and C. Craddock. Learning and comparing functional connectomes across subjects. *NeuroImage*, 80:405, 2013.

G. Varoquaux, F. Baronnet, A. Kleinschmidt, P. Fillard, and B. Thirion. Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *MICCAI*, pages 200–208. 2010a.

G. Varoquaux, M. Keller, J.B. Poline, P. Ciuciu, and B. Thirion. ICA-based sparse features recovery from fMRI datasets. In *ISBI*, page 1177, 2010b.

G. Varoquaux, S. Sadaghiani, P. Pinel, A. Kleinschmidt, J. B. Poline, and B. Thirion. A group model for stable multi-subject ICA on fMRI datasets. *NeuroImage*, 51:288, 2010c.

G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Inf Proc Med Imag*, pages 562–573, 2011.

Gael Varoquaux and Bertrand Thirion. How machine learning is shaping cognitive neuroimaging. *GigaScience*, 3(1):28, 2014.

Judith S Verhoeven, Nathalie Rommel, Elena Prodi, Alexander Leemans, Inge Zink, Ellen Vandewalle, Ilse Noens, Johan Wagemans, Jean Steyaert, Bart Boets, et al. Is there a common neuroanatomical substrate of language deficit between autism spectrum disorder and specific language impairment? *Cerebral Cortex*, 22(10):2263–2271, 2012.

N.X. Vinh, J. Epps, and J. Bailey. Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *Journal of Machine Learning Research*, 11:2837–2854, 2010.

David Wechsler. Wechsler intelligence scale for children. 1949.

Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity*, 2(3):125–141, 2012.

Lorna Wing and Judith Gould. Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of autism and developmental disorders*, 9(1):11–29, 1979.

Chao-Gan Yan, R Cameron Craddock, Xi-Nian Zuo, Yu-Feng Zang, and Michael P Milham. Standardizing the intrinsic brain: towards robust measurement of inter-individual variation in 1000 functional connectomes. *Neuroimage*, 80:246–262, 2013.

Zhijian Yao, Li Wang, Qing Lu, Haiyan Liu, and Gaojun Teng. Regional homogeneity in depression and its relationship with separate depressive symptom clusters: a resting-state fMRI study. *Journal of affective disorders*, 115(3):430–438, 2009.

B.T.T. Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysio*, 106:1125, 2011.

Yufeng Zang, Tianzi Jiang, Yingli Lu, Yong He, and Lixia Tian. Regional homogeneity approach to fMRI data analysis. *Neuroimage*, 22(1):394–400, 2004.

Yuan Zhou, Meng Liang, Lixia Tian, Kun Wang, Yihui Hao, Haihong Liu, Zhening Liu, and Tianzi Jiang. Functional disintegration in paranoid schizophrenia using resting-state fMRI. *Schizophrenia research*, 97(1):194–205, 2007.

Xi-Nian Zuo, Clare Kelly, Adriana Di Martino, Maarten Mennes, Daniel S Margulies, Saroja Bangaru, Rebecca Grzadzinski, Alan C Evans, Yu-Feng Zang, F Xavier Castellanos, et al. Growing together and growing apart: regional and sex differences in the lifespan developmental trajectories of functional homotopy. *The Journal of neuroscience*, 30(45):15034–15043, 2010.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Apprentissage d'atlas fonctionnels du cerveau modélisant la variabilité inter-individuelle
Mots-clés: IRMf au repos, apprentissage de dictionnaire, connectivité fonctionnelle, biomarqueurs

Résumé : L'Imagerie par Résonance Magnétique fonctionnelle (IRMf) est une source prometteuse de biomarqueurs permettant le diagnostic de troubles neuropsychiatriques sur des sujets non coopératifs. Un connectome est communément estimé en établissant un atlas de régions cérébrales représentatif de l'organisation fonctionnelle puis en étudiants la corrélation entre leurs signaux.

La première étape, la définition de régions, est réalisée à l'aide d'une analyse statistique de l'activité cérébrale spontanée au repos. Pour les extraire, nous utilisons une approche d'apprentissage de dictionnaire multi-sujets intégrant une pénalité imposant compacité spatiale et parcimonie. Nous extrayons les unités fonctionnelles de base des réseaux fonctionnels extraits à l'aide de techniques de segmentation inspirées du domaine de la vision. Nous montons à l'échelle sur de gros jeux de données en utilisant une stratégie d'optimisation stochastique.

L'absence de vérité de terrain est un frein à l'évaluation de la pertinence de nos modèles. Pour pallier ce problème, nous introduisons des métriques évalouant la stabilité et la fidélité des modèles générés. Appliquées à plusieurs sous-ensembles d'un grand jeu de données, nous montrons que nos méthodes de définition de régions expliquent mieux les données et sont plus stables à travers les sujets que les méthodes de référence.

Nous intégrons ensuite notre méthode de définition de régions dans un pipeline entièrement automatisé afin d'estimer un connectome et de l'utiliser dans des tâches de prédiction. Nous étudions sa pertinence empirique sur une tâche clinique de diagnostic des troubles autistiques et montrons la première prédiction de trouble psychiatrique à travers différents sites d'acquisition et sur des sous-ensembles d'homogénéité variable. Nos résultats de prédiction sont supérieurs à l'état de l'art. Nous démontrons que la prédiction sur des sites inconnus est aussi efficace que la prédictions sur site connus à condition d'avoir suffisamment de sujets. Les neurophénotypes extraits sont compatibles avec la littérature.

Enfin, par une analyse post-hoc des résultats, nous montrons que la définition de région est l'étape la plus importante du pipeline et que l'approche que nous proposons obtient les meilleurs résultats. Nous fournissons également des recommandations sur les méthodes les plus performantes pour les autres étapes du pipeline.

Title: Learning functional brain atlases modeling inter-subject variability
Keywords: resting state fMRI, dictionary learning, functional connectivity, biomarkers

Abstract: Resting-state functional Magnetic Resonance Imaging (fMRI) holds the promises to reveal functional biomarkers for neuropsychiatric disorders applicable to non-cooperative patients. To extract such biomarkers, the standard approach is, first, to establish an atlas of the cerebral areas defining the functional organization of the brain, and then to study the correlation of their brain signals. They form the functional connectome.

The first step of building a connectome requires defining the regions that constitute brain functional units. This is done by performing a statistical analysis of the dynamics of spontaneous brain activity during resting-state. We introduce a method combining spatially-structured and sparsity-inducing penalties in a multi-subject dictionary learning approach to automatically extract brain networks from rest-fMRI. A stochastic optimization strategy enables scaling to big datasets. We then show how computer vision inspired segmentation techniques can be used to threshold automatically and break down these networks into connected functional units.

To compensate for the lack of ground truth, we introduce two metrics that aim at scoring the stability and the data fidelity of the generated models. Using these metrics, we show that our methods better explain the data and are more stable across subjects than reference decomposition or clustering methods.

We then integrate this region-definition method in a fully-automatic prediction pipeline, to build connectomes from the data and use them in classification tasks. We study its empirical relevance on the clinical task of predicting autism spectrum disorders. We demonstrate the first prediction of pyschiatric condition across different scanning sites and apply it on subsets of participants of variable homogeneity. We exhibit prediction scores higher than state of the art and show that, given a sufficient number of individuals in the training set, prediction across sites is as efficient as traditional prediction. We also extract autism neurophenotypes compatible with the litterature.

Finally, we show that region definition is the most important step of the pipeline and that our approach is the best performer. We also explore the other steps of the pipeline and give recommandations on how to choose a prediction pipeline.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France