Structural and probabilistic methods for group analysis in functional neuroimaging

Bertrand Thirion

Mémoire d'Habilitation à Diriger des Recherches (HDR) de Bertrand Thirion:
Méthodes structurales et probabilistes pour l'analyse de groupe en neuroimagerie fonctionnelle.

Bertrand Thirion

May 22, 2009

INRIA Saclay-Île-de-France
Neurospin, CEA Saclay, Bâtiment 145,
91191 Gif-sur-Yvette Cedex
http://collaborations.saclay.inria.fr/parietal/
E-mail: bertrand.thirion@inria.fr
Tel : (+33) 1 69 08 79 92
Fax: (+33) 1 69 08 79 80

Parc Orsay Université,
4 rue J. Monod,
91893 Orsay Cedex France.
Tel. : (+33) 1 72 92 59 00

Ecode doctorale: CMLA, ENS Cachan

0

Contents

Part I

Notice individuelle

Chapter 1

Notice individuelle

Bertrand Thirion est un ingénieur de l'équipe PARIETAL de l'INRIA Saclay-Île-de-France spécialisé dans le traitement de données en neuroimagerie. Après une thèse de modélisation de données d'IRM fonctionnelle soutenue en 2003 à l'INRIA Sophia-Antipolis et un post-doc réalisé au Service Hospitalier Frédéric Joliot (CEA Saclay), il est entré à l'INRIA Futurs/Saclay comme Ingénieur spécialiste, où il est en train de créer l'équipe PARIETAL.
Ses intérêts de recherche concenrent la modélisation du cerveau et de sa variabilité macroscopique à partir de données d'IRM anatomiques et fonctionnelles (MRI/fMRI), la fouille de données (réduction de dimension, clustering), la modélisation probabiliste (modèles graphiques), les systèmes dynamiques. Il encadre actuellement trois étudiants en thèse.

1.1 Curriculum Vitae

Current situation

Engineer in the Corps des Télécommunications (https://www.corpstelecom.org/).
Employed at INRIA Saclay-Île-de-France as a specialist engineer

Education

2000-2003 PhD thesis, Ecole Nationale Supérieure des Télécommunications (Paris), image and signal processing department, performed at INRIA Sophia-Antipolis, Robotvis/Odyssée project, on the analysis of functional MRI data.
1998-2000 Ecole Nationale Supérieure des Télécommunications (Paris), general cursus in informatics, electronics, numeric communication systems, artificial intelligence, cryptography, specialization in image and signal processing.
1999-2000 Master in theoretical Mathematics, Paris 7, Riemannian and Hyperkählerian Geometry, Cauchy rigidity theory.

1998-1999 Master Mathematics-Vision-Perception at E.N.S. Cachan, Advanced mathematical techniques in signal and image processing: wavelets, PDEs, 3D scene modeling, variational methods, probabilistic methods.
1995-1998 Master degree at Ecole Polytechnique, general scientific cursus, specialization in mathematics: geometry, dynamical systems, applications to quantum physics and relativity.
1993-1995 Lycée Sainte-Geneviève (Versailles): Mathématiques Supérieures et Spéciales M'.

Main Research interests

- Studying the statistical and dynamical structure of biological signals.
- Study of the brain anatomical and functional variability.
- Statistical learning theory and its applications.
- Graph theory, graphical models and Bayesian inference.
- Geometrical data representation, dimension reduction, manifold theory.
- Information theory and its links with statistics.

Professional experience

2005-now INRIA Futurs/Saclay, specialist engineer: Creation of an INRIA team (PARIETAL) at Neurospin.
Supervision of three PhD students about *i)* the joint variability of anatomical and functional structures as seen in MRI *ii)* classification based on neuroimaging data and *iii)* brain decoding techniques.
Works on brain functional landmarks, anatomo-functional parcellations of the cortex, the reproducibility of group studies in neuroimaging and inverse inference in neuroimaging.
2003-2005 CEA, Service Hospitalier Frédéric Joliot
Post-doc about the modelling of inter-subject variability in fMRI: Setting and analysis of a functional localizer experiment on a large cohort of subjects, analysis of the variability through inter-subject parcellation techniques. In parallel, conception, setting and analysis of an inverse inference (brain reading) experiment in collaboration with S. Dehaene.
2000-2003 INRIA Sophia-Antipolis, Robotvis/Odyssée team
PhD thesis on the modelling of fMRI datasets : Introduction of non-parametric statistical models (information theory and dynamical models)
Setting and analysis of a retinotopy experiment in fMRI: automatic delineation of the visual areas given functional maps of the cortex.
April-October 1999 Siemens Corporate Research (Princeton, NJ, USA)
Internship in Image processing department: recognition of objects in industrial environments through color-based, lighting-based and Markov Random Field segmentation techniques.

April-July 1998 LITEF (Freiburg, Germany), maker of navigation sensors, R&D department: Implementation of new plane trajectory model in the simulation software.

1.2 Liste des publications et communications scientifiques

1.2.1 Journal publications

- [212] Bertrand Thirion and Olivier Faugeras. Dynamical components analysis of fMRI data through kernel PCA. *Neuroimage*, 20(1):34–49, Sep 2003

- [213] Bertrand Thirion and Olivier Faugeras. Feature characterization in fMRI data: the information bottleneck approach. *Medical Image Analysis*, 8(4):403–419, December 2004

- [199] B. Thirion, S. Dodel, and J.-B. Poline. Detection of signal synchronizations in resting-state fMRI datasets. *Neuroimage*, 29(1):321–327, January 2006

- [203] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Human Brain Mapping*, 27(8):678–693, August 2006

- [211] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, J.-B. Poline, D. Le Bihan, and S. Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4):1104–1116, December 2006

- [141] S. Mériaux, A. Roche, G. Dehaene-Lambertz, B. Thirion, and J.-B. Poline. Combined permutation test and mixed-effect model for group average analysis in fMRI. *Human Brain Mapping*, 27(5):402–410, May 2006

- [204] B. Thirion, P. Pinel, S. Mériaux, A. Roche, S. Dehaene, and J.-B. Poline. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35(1):105–120, 2007

- [206] B. Thirion, P. Pinel, A. Tucholka, A. Roche, P. Ciuciu, J.-F. Mangin, and J.-B. Poline. Structural analysis of fMRI data revisited: Improving the sensitivity and reliability of fMRI group studies. *IEEE Transactions on Medical Imaging*, 26(9):1256–1269, September 2007

- [167] P. Pinel, B. Thirion, S. Mériaux, A. Jobert, J. Serres, D. Le Bihan, J.-B. Poline, and S. Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci*, 8(1):91, Oct 2007

- [175] A. Roche, S. Mériaux, M. Keller, and B. Thirion. Mixed-effects statistics for group analysis in fMRI: A nonparametric maximum likelihood approach. *Neuroimage*, 38:501–510, 2007

- [127] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969, July 2008

- [163] M. Perrin, Y. Cointepas, A. Cachia, C. Poupon, B. Thirion, D. Rivière, P. Cathier, V. El Kouby, A. Constantinesco, D. Le Bihan, and J.-F. Mangin. Connectivity-based parcellation of the cortical mantle using q-ball diffusion imaging. *Int J Biomed Imaging*, 2008

1.2.2 conferences

- [198] B. Thirion, B. Bascle, V. Ramesh, and N. Navab. Fusion of color, shading and boundary information for factory pipe segmentation. In B. Bascle, editor, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 349–356 vol.2, 2000

- [208] B. Thirion, B. Thirion, and O. Faugeras. Revisiting non-parametric activation detection on fmri time series. In O. Faugeras, editor, *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis MMBIA 2001*, pages 121–128, 2001

- [209] B. Thirion, B. Thirion, and O. Faugeras. Dynamical components analysis of fMRI data. In O. Faugeras, editor, *Proc. IEEE International Symposium on Biomedical Imaging*, pages 915–918, 2002

- [214] Bertrand Thirion and Olivier D. Faugeras. Dynamical components analysis of fmri data: A second order solution. In *EUROCAST*, pages 552–563, 2003

- [201] B. Thirion and O. Faugeras. Feature detection in fMRI data: The information bottleneck approach. In *Proceedings 6th International Conference on Medical Image Computing and Computer Assised Intervention*, pages 83–91, Montreal Canada, November 2003

- [202] B. Thirion and O. Faugeras. Nonlinear dimension reduction of fMRI data: the Laplacian embedding approach. In *Proceedings 2st International Symposium on Biomedical Imaging*, pages 372–375, Arlington, VA, April 2004

- [205] B. Thirion, P. Pinel, and J.-B. Poline. Finding landmarks in the functional brain: detection and use for group characterization. In *MICCAI'05*, volume 8, pages 476–483, 2005

- [200] B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, and S. Dehaene. Reading the brain visual system as an inverse problem. In *Proceedings 3th International Symposium on Biomedical Imaging*, Washington, USA, April6-9 2006

- [207] B. Thirion, A. Roche, P. Ciuciu, and J.-B. Poline. Improving sensitivity and reliability of fMRI group studies through high level combination of individual subjects results. In *Proceedings MMBIA2006*, New York, USA, June17-18 2006

- [142] S. Mériaux, A. Roche, B. Thirion, and G. Dehaene-Lambertz. Robust statistics for nonparametric group analysis in fMRI. In *Proceedings 3th International Symposium on Biomedical Imaging*, pages 936–939, Arlington, VA, April 2006

- [217] B. Thyreau, B. Thirion, G. Flandin, and J.-B. Poline. Anatomo-functional description of the brain: a probabilistic approach. In *Proceedings 31th Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, volume V, pages 1109–1112, Toulouse, France, May 2006

- [210] B. Thirion, A. Tucholka, M. Keller, P. Pinel, A. Roche, J.-F. Mangin, and J.-B. Poline. High level group analysis of FMRI data based on Dirichlet process mixture models. In *International Conference on Information Processing in Medical Imaging*, volume 4584 of *LNCS*, pages 482–494, 2007

- [80] Alexandre Gramfort, Benoit Cottereau, Maureen Clerc, Bertrand Thirion, and Sylvain Baillet. Challenging the estimation of cortical activity from MEG with simulatedfmri-constrained retinotopic maps. In *Proc. of the 29th IEEE EMBS Annual international conference*, volume 1, pages 4945–4948, 2007

- [195] Z. Y. Sun, D. Rivière, , E. Duchesnay, B. Thirion, F. Poupon, and J.-F. Mangin. Defining cortical sulcus patterns using partial clustering based on bootstrap and bagging. In *5th International Symposium on Biomedical Imaging*, pages 1629–1632, Paris, France, May 2008

- [225] A. Tucholka, B. Thirion, P. Pinel, J.-B. Poline, and J.-F. Mangin. Triangulating cortical functional networks with anatomical landmarks. In *5th International Symposium on Biomedical Imaging*, pages 612–615, Paris, France, May 2008

- [232] T. Vincent, P. Ciuciu, and B. Thirion. Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI. In *5th International Symposium on Biomedical Imaging*, pages 568–571, Paris, France, May 2008

- [135] V. Michel, C. Damon, and B. Thirion. Mutual information-based feature selection enhances fMRI brain activity classification. In *5th International Symposium on Biomedical Imaging*, pages 592–595, Paris, France, May 2008

- [224] A. Tucholka, B. Thirion, M. Perrot, P. Pinel, J.-F. Mangin, and J.-B. Poline. Probabilistic anatomo-functional parcellation of the cortex: how many regions? In *11thProceedings MICCAI, LNCS Springer Verlag*, New-York, USA, 2008

- [33] C. Damon, P. Pinel, M. Perrot, V. Michel, E. Duchesnay, J.-B. Poline, and B. Thirion. Discriminating between populations of subjects based on FMRI data using sparse features selection and SRDA classifier. In *MICCAI 2008 Workshop on Analysis of Functional Medical Images*, New York, Sept 2008

1.3 Liste des encadrements scientifiques

- Nicolas Wotawa at INRIA Sophia/Antipolis, Odyssée team, 2002-2003 (Msc and beginning of PhD): Study of the retinotopic structure of the visual areas based on the analysis of anatomo-functional data.
- Benjamin Thyreau at CEA/SHFJ, 2005, (Msc): Probabilistic modelling of anatomo-functional parcellations ; application to the optimization of the number of parcels.
- Alan Tucholka at INRIA Futurs/SHFJ, 2006 (Msc): Study of the co-variability of anatomical and functional structures observed with MRI ; surface-based approach.
- Alan Tucholka au CEA/Neurospin, 2006- (PhD): Robust and non-parametric detection of brain signals (co-supervision with J.B. Poline).
- Cécilia Damon at INRIA Futurs/Neurospin, 2006- (PhD): use of functional neuroimaging to characterize and classify subject populations (co-supervision with J.B. Poline).
- Vincent Michel at INRIA Saclay/Neurospin, 2007 (Msc): Classification of mental states based on neuroimaging data.
- Vincent Michel at INRIA Saclay/Neurospin, 2007- (PhD): Understanding the visual cortex by using classification techniques (co-supervision with G. Celeux).

1.4 Participation à la vie scientifique

Invited presentations and course

- Société d'anatomie fonctionnelle cérébrale (SAFC): June 2003.
- Preparation of course and applied sessions for the national MRI course (Marseille, November 2004; Paris, September 2006; Marseille, May 2009)
- Laboratoire de neurosciences cognitives et d'Imagerie Cérébrale (la Pitié-Salpêtrière): November 2005, February 2007
- Unité de Neurosciences Intégratives et Computationnelles du CNRS: January 2006
- Human Brain Mapping Organization: Morning workshop, June 2006
- FMRIB laboratory (Oxford): April 2007
- NIH-INRIA symposium: April 2007, June 2009
- NIH-ICT symposium: October 2007

- Alpine Brain Imaging Meeting: January 2008
- FENS forum: July 2008
- NIPS workshop on fMRI data analysis: December 2008.
- MIT CSAIL (Boston): May 2009.
- Organizer of the MICCAI workshop on statistical methods for fMRI data analysis: Sept. 2009.

Reviewer activities :
Journals: NeuroImage, Human Brain Mapping, Medical Image Analysis and IEEE Transactions on Medical Imaging, International Journal of Biomedical Imaging.
Conferences: MICCAI, IEEE EMBC, ICCV.
ANR: Syscomm call (2008)

Software projects:

- Participation to the nipy (neuroimaging in python) project, to a high-level library dedicated to functional MRI, in C/C++/Python, http://neuroimaging.scipy.org/, registered at the french software foundation (APP)
- Responsibility of the building of the functional data analysis toolbox in Brainvisa software (http://brainvisa.info/), which is currently developed by an INRIA associate engineer, Lise Favre.

Work groups and grants :

- Participation to the BrainVAR collaborative research action (Xavier Pennec 2007-2008), dedicated to the study of cortical variability.
- Creation of a research group with Select team(INRIA Saclay/Paris 11) dedicated to group analysis and population classification in neuroimaging (2007-).
- B. Thirion is Member of Vimagine, an accepted ANR blanc project (2008-2012), dedicated to the retinotopic organization of the visual cortex, based on MEG and MRI; this project should open the way to understanding the dynamics of brain processes for low-level vision, with an emphasis on neuropathologies.
- B. Thirion is responsible of a joint INRIA-INSERM research project (2008-2010) dedicated to the study of the connectivity of the brain during resting state based on fMRI data (funding for one post-doc, Gaaroquaux 24 months, started on 1/10/2008).

- B. Thirion is indirectly involved in WP7 (bio-statistics) of IMAGEN European project http://www.imagen-europe.com/, 2007-2012) that aims at building a database with multi-modal MRI, behavioural and genetics data from 2000 adolescents.
- Member of the research program committee at Neurospin: choice of the main research projects, coordination and prospection.

Part II

Travaux de recherche

Chapter 2

Introduction

This chapter is not an original contribution, but an introduction to our work on group analysis of functional Magnetic Resonance Imaging (fMRI) datasets; it defines some of the concepts that will be used in the following sections. This chapter presents a synthetic view of the principles, goals and techniques of intra-subject or first-level fMRI data analysis, with a short glimpse of anatomical MRI data processing. Our aims here is not to be exhaustive but proposes a specific point of view on this domain of research. In particular, we prepare the discussion for multi-subject or second-level fMRI data analysis, which will be the main focus of this document.

Before starting, let us simply remind that, as the first functional Magnetic Resonance Images were acquired in the early nineties, the domain is still young and the current techniques are evolving quickly.

2.1 Understanding brain structure and function

In this section we start by reminding general facts on *human brain mapping* considered as a domain of research. We start by describing the main challenge in human brain mapping, which is merely brain complexity. Then we recall a few general facts on neuroscience and functional neuroimaging.

Brain complexity: Let us start with a review of some aspects of the brain complexity. Brain complexity is first *combinatorial*, since the brain comprises about 10^{11} neurons and 10^{15} synapses. Producing a realistic model or a simulation of such a huge amount of entities is thus a challenge that will need decades to be addressed (see e.g. the blue brain project http://bluebrain.epfl.ch/). For this reason, neuroscience can only proceed by revealing some snapshots of brain structure and function, depending on the modality of observation. This yields a second aspect of brain complexity: Several levels of organization at *different scales* must be considered, e.g. (in the case of the cortex) neurons, cortical columns, cortical maps, brain systems and lobes Brain modelling is thus addressed at these different levels, but it is yet not clear how to integrate all these levels in a common framework. A third aspect of brain complexity will be discussed further in this document: *inter-subject variability* (see Chapter 3). It means that there is no one-to-one

mapping between any two different brains, so that it is difficult to identify and describe brain structures at the population level.

Observing the brain: The modalities for exploring brain structure and function are numerous, and they evolve quite rapidly according to technological breakthroughs that allow the observation of new phenomena. In this section, we give only a superficial overview of these modalities, with an emphasis on human studies.
Historically, many insights on the functional architecture of the brain have been gained through electrophysiology studies in monkeys: Electrical recordings provided local information on brain function (e.g. firing neurons) which is the basis for the characterization of functionally specific brain areas. On the other hand, this technique does not provide a global view on brain activity, is expensive (due to the acquisition setup and monkey training), and generalization to the human brain raises cross-species issues [221]. More anecdotally, e.g. in the case of implanted epileptic patients, direct electrophysiological measurements are also performed on humans.
A second source of information, historically the first one, comes from lesion studies: it consists in describing the impact of local lesions on brain function. This description is of course very macroscopic. Brain architecture has been explored through post-mortem analyses (histology), so that the cell-level organization (cyto-architecture) has become one of the major ways to describe different brain regions (Brodmann areas). This description is limited, however, by the large inter-subject variability in brain structure.

Today, the main source of information is neuroimaging: Magnetic Resonance Imaging (MRI), Positon Emission Tomography (PET, SPECT), Optical Imaging (OI) and Magneto- and Electro-encephalography (MEEG).

- MRI has provided, for a long time now, anatomical images (description of the brain in terms of tissues and macroscopic model of brain shape). More recently, it has also been used to acquire functional images (metabolic activity) [151]. It is now the main tool for structural and functional brain investigation. Finally, diffusion MRI provides a unique picture on brain anatomical connectivity [169].

- Before functional MRI became the standard tool for producing brain map, PET was used to map metabolic activity of the brain. However this modality is invasive, slow, and does not yield a very high spatial resolution. PET is now used for pharmacological studies, due to its ability to trace specific chemical components.

- OI is probably the most recent modality for brain investigation. It can provide images of brain metabolism or intrinsic activity. To date, it does not allow a full coverage of the brain volume.

- MEEG, unlike the other modalities, performs some measurement of the intrinsic electro-magnetic activity of the brain. But it is recorded on the scalp, with a rather coarse resolution. The localization of the activity is then quite challenging.

These modalities are represented, with their resolution and invasivity in Fig. 2.1.

Let us now quickly outline the general philosophy and current trends in human brain mapping.

Modern human brain mapping: The main aim of modern human brain mapping is to build a cartography of human brain functions (see e.g. [64]). Functionally, this is based on a subtractive reasoning principle: given activity maps related to two experimental conditions A and B, the difference map is expected to represent the spatial substrate of the cognitive processes involved in condition A and not in condition B. Localizing the corresponding regions in a population of subjects is a fundamental issue, which is not reduced to the construction of brain maps, but also involves the understanding of the underlying brain connectivity (see e.g. [97, 8]). More recently, there has also been a trend toward getting more insight on the mechanisms of brain activity, i.e. to understand and model what effects regions exert on each other in a certain experimental context [165, 9].

Human brain mapping is certainly an important project as such, but it also has some impact on the study of brain disease. Currently, one of the main hopes is that imaging might be used as an objective diagnosis tool and therapy follow-up for several diseases. In particular, one of the aims of anatomical and functional imaging is to finding some *biomarkers* for psychiatric diseases by comparing different population of patients: this is the case for autism [12], schizophrenia [18], Alzheimer's disease [215], or various cognitive and sensori-motor impairments [42]. The main anatomical biomarkers can be the cortical thickness [53, 124], the cortical shape [46], differences of diffusion anisotropy in the white matter [42], structure of fiber tracts [45]. Functional biomarkers are simply the amount of activity in certain regions which may characterize certain disabilities, or the amount of correlation of the brain signals across different regions (functional connectivity patterns, see e.g. [99]).
However, diagnosis is not the only question of interest, and in many cases, brain imaging may help to understand or provide some mechanisms for some cognitive diseases, e.g. dyslexia [93] or dyscalculia [138]. This is a major challenge for the next years.

2.2 Background on fMRI data analysis

2.2.1 Principle of an activation studies

The principle of an activation studies with fMRI is simple. It consists in placing a subject or a patient in the scanner and acquiring data that reflect the hemodynamic activity of the brain. While the acquisition occurs, subjects are submitted to various stimuli or asked to perform certain tasks. These will influence the regional neural activity of the brain which in turn will modulate locally the hemodynamic (blood flow, blood volume and oxygenation) of some brain regions. The data acquired is a series of slices, each of which can be recorded typically in less than 80ms, with an in plane resolution and thickness of the order of the mm (1-3 mm). Therefore, the whole brain volume can be covered in a few seconds with current scanners but equipment is improving fast and these numbers will get down. Data acquisition is continuous for 5 to 20 minutes, and is called a *run*, an MR session usually consists of several runs for a subject, together with the acquisition of other kind of images reflecting the anatomy of the brain (e.g.

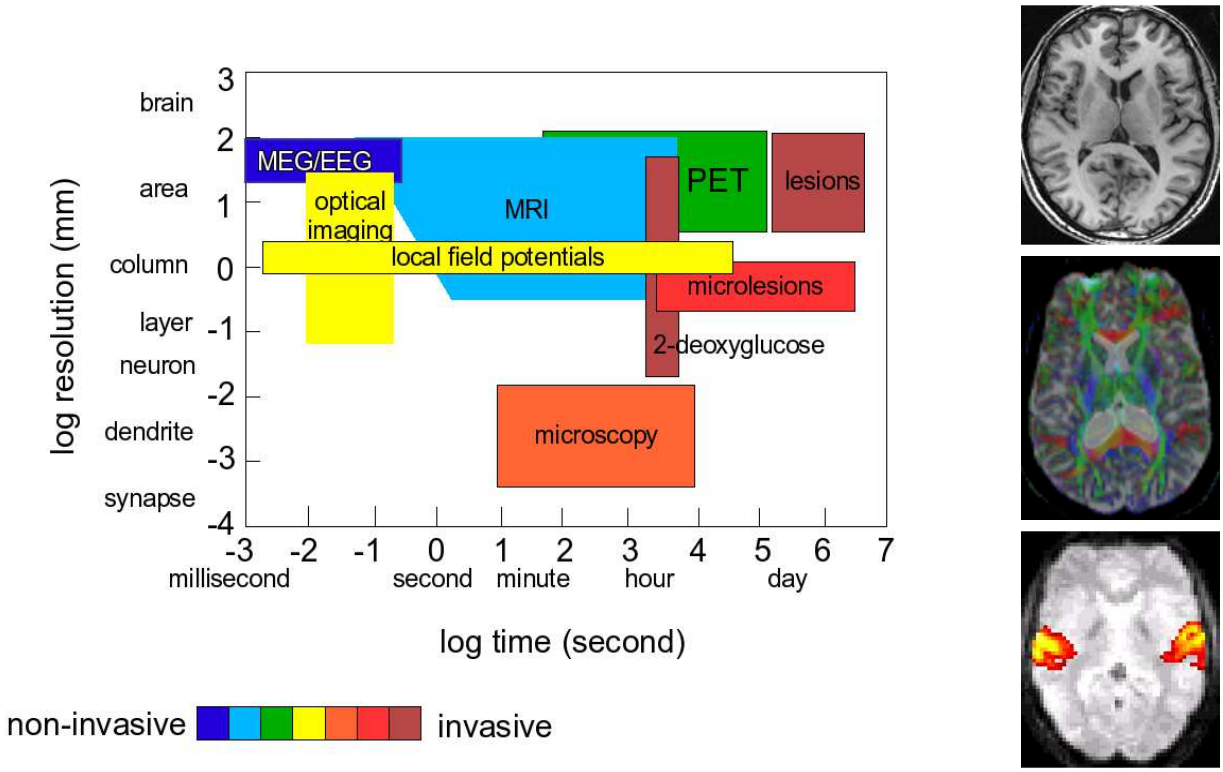


Figure 2.1: (left) The different imaging modalities for brain mapping. MRI and functional MRI have the unique property to yield high-resolution information while being minimally invasive. Unlike other modalities, MRI allows whole brain imaging. (right) Typical example of T1 anatomical MRI (top), pre-processed Diffusion-Weighted (DW) MRI (middle) and fMRI (bottom) images, presented in axial views. These images are from the Neurospin 3T scanner. For the DW-MRI image (courtesy from P.Fillard), the main direction of water diffusion is color-coded: green for antero-posterior diffusion, red for lateral diffusion, blue for vertical diffusion. The functional image has been analyzed to yield the regions activated in an auditory task.

T1 images with grey white matter contrast, diffusion images from which fibre bundles connecting brain regions can be reconstructed). The data acquired for one subject is of the order of the Giga byte, and an actual experiment usually involves 15 to 40 subjects (for instance using different groups of subjects such as normal versus patients). During a run, an experimental paradigm designed to elicit a certain component of perception or cognition is proposed to the subject. It consists in several tasks or stimuli (the *conditions*) and fMRI data is commonly used to detect brain regions whose hemodynamic signal varies between conditions. Other data such as reaction time, eye movement, cardiac or breathing rhythms can be conjointly acquired. A comparison of two or more experimental conditions is called a *contrast*.

2.2.2 A short perspective on functional neuroimaging.

As previously described, fMRI data are often used for detecting brain regions whose hemodynamic signal varies across experimental conditions. Applications in human studies can generally be considered as addressing one of the two aspects: group versus single subject studies and normal versus patient studies. On the latter axis, the understanding of human brain functions is opposed to dysfunctions in psychiatric of neurological diseases. On the former axis, the specific information obtained from a particular subject is contrasted to the description of the information obtained at the population level. Using groups of normal subjects, cognitive neuroscientists that use fMRI to probe the brain functions are seriously challenged by philosophers or psychologists such as J. Fodor [62] who claim that localizing brain regions that respond to certain stimuli does not help to understand how the brain functions. Fodor uses the mechanistic analogy of a car engine, and asks how the knowledge of the localisation of pieces such as the piston or carburettor helps to understand the function of theses pieces in the engine. It does not, unless one is interested in mending some parts, as the neurosurgeon might be for pathologies involving brain surgery. While the argument is potent, it does not account for the numerous occasions where the spatial organisation is a reflection of the functioning, such as the retinotopic organisation of the early visual cortex, and that brain region characterisation and localisation might be a necessary first step in the process of defining models of brain functioning.

2.2.3 What we need to know about the data

As described previously, the origin of the data is hemodynamic. Following an increase of synaptic and spiking activity, neurons require more energy conveyed through the blood in the form of oxygen and glucose [136]. Through a mechanism not fully known, the result is first an increase of oxygen extraction followed by an increase of blood flow that over compensates the loss of oxy-hemoglobin. The Blood Oxygen Level Dependant (BOLD) contrast can be measured with MRI because the oxy-hemoglobin is diamagnetic while the deoxy-hemoglobin is paramagnetic, such that the MR signal increases with the ratio oxy- over deoxy-hemoglobin [151]. The timing of this effect is relatively slow and picks around four to six seconds after a short stimulation and returns to baseline in about 25s. This behaviour is usually modeled by a hemodynamic filter h. It is interesting to note that even if the absolute timing may be considered as poor, the differential timing between two conditions can be precise to the order of 100 ms.

The BOLD contrast does not show many anatomical details, and is sensitive to artifacts due to magnetic field variation around air/tissue interfaces inducing loss of signal or volume deformation in these areas.

2.3 fMRI data analysis: State of the art

The most common approach that has dominated the past decade can be decomposed in the following steps. It success is linked to freely available tools such as SPM (see www.fil.ucl.ac.uk/SPM).

2.3.1 Data pre-processing: temporal and spatial realignments.

First, some distortions are specific to echo-planar images (EPI) commonly used in fMRI; in order to unbias the localization of brain regions, it is thus important to correct these distortions [96]. However, part of the signal cannot be recovered. Fortunately, the use of modern multi-channel antennas currently limits these distortions.

Second, subjects are never completely still in the scanner and movement needs to be corrected with a strong impact on the signal obtained. Movements correlated to the experimental paradigm are particularly difficult to correct if not impossible. Most current techniques assume a rigid body movement between two brain scans. Temporally, the slices of one brain volume are not acquired at the same time, and all voxel-based time series are usually interpolated to impose a unique timing for all voxel of one brain volume.

Spatial coregistration to a template is usually considered as a pre-processing, although it might also be done after the first level analysis. We defer the description of that processing step to section 2.5.1 since we consider it as crucial for group analyses.

2.3.2 The general linear model (GLM)

The BOLD response This step consists in modelling the BOLD signal and constructing statistical maps. As a first approximation the BOLD hemodynamic response function (HRF) can be considered as a linear phenomenon with respect to stimulation. In practice, the following model [77] is used:

$$h(t) = \mathbb{I}_{t>0} \left[\left(\frac{t}{a_1 b_1} \right)^{a_1} \exp\left(b_1 - \frac{t}{a_1} \right) - c \left(\frac{t}{a_2 b_2} \right)^{a_2} \exp\left(b_2 - \frac{t}{a_2} \right) \right], \tag{2.1}$$

where $a_1 = 6$, $a_2 = 12$, $b_1 = 0.9s$, $b_2 = 0.9s$ and $c = 0.35$. However, the hypothesis that a certain filter h can be used to fit the temporal responses in a linear way, whatever the experimental paradigm, in any region of the brain and in any individual, is not fulfilled in general [1, 2, 44], and important work has been performed to better model these response [24, 231, 25, 127].

More fundamentally, there is a need to model the set of phenomena that occur during the BOLD response. To date, the most satisfactory generative model of the BOLD is the so-called balloon model [16], which has been further discussed and assessed in a series of works [74, 172, 37, 192, 94]. However, it is important to recall that this model is over-parametrised, hence can hardly be reliably tested based

on the noisy and short times series available in fMRI datasets. To give an idea of the state of the art, we simply provide in Fig. 2.2 a description of the historical model of [16] completed by [74].

Let u be the neuronal activity at a certain cortical location, (f, v) be the blood flow and the blood venous volume at that location, q the vein deoxyhemoglobin content (all these variables are normalized to be 1 at rest), and y be the bold signal. The balloon model is then

$$\begin{aligned}
\ddot{f} &= \varepsilon u - \kappa_s \dot{f} - \kappa_f (f-1) \\
\dot{v} &= \frac{1}{\tau}(f - v^{1/\alpha}) \\
\dot{q} &= \frac{1}{\tau}\left(f \frac{1-(1-E_0)^{1/f}}{E_0} - v^{1/\alpha - 1} q \right) \\
y &= V_0(a_1(1-q) - a_2(1-v))
\end{aligned}$$

Here ε is the variable of interest, i.e. the neural efficacy. The constants $(\kappa_s, \kappa_f, \tau, \alpha, E_0, V_0)$ have to be estimated, since they may vary from locations to locations or from subjects to subjects. (a_1, a_2) are usually assumed to be known.

Figure 2.2: Formulation of the Balloon model according to [74].

To conclude this overview on BOLD modelling, the standard model based on the hemodynamic filter in Eq. 2.1 works astonishingly well in many contexts, and has never been seriously challenged by concurrent approaches for group studies, which is our perspective here.

The design matrix. A linear model is constructed that includes all experimental factors which are believed to have an impact on the BOLD signal. Each regressor is constructed as the convolution of a standard HRF as in Eq. (2.1) with a time series representing the occurrence and duration of the experimental condition (stimuli or tasks). Given a set of time courses describing the occurrence of experimental condition $(p_1(t), .., p_C(t))_{t=1..N_t}$, where N_t is the number of scans, the fMRI design matrix is thus obtained by

$$X = \begin{bmatrix} (h * p_1)(1) & \dots & (h * p_C)(1) \\ \vdots & & \\ (h * p_1)(N_t) & \dots & (h * p_C)(N_t) \end{bmatrix} \tag{2.2}$$

Note that in practice, the design matrix has to be computed at higher temporal resolution and then down-sampled at the fMRI acquisition resolution. The intensity, or more generally parameters (loudness, speed, angle etc.) of the stimuli may vary in some experimental designs; in such cases, the time courses $p_c(t)$ take scalar values corresponding to the encoded quantities. Moreover, the temporal derivative of

the C regressors $\left(\frac{d(h*p_c)(t)}{dt} = \frac{dh}{dt} * p_c\right)_{c=1..C}$ may be included in order to avoid that temporal misfit of the provided model inflates the residual variance.

X is further completed by adding a constant regressor, low frequency drifts models and possibly other known confounds (motion estimates, cardiac and respiratory measurements). It is thus an $N_t \times R$ matrix, N_t being the number of scans and $R > C$ being the number of model regressors. The remaining noise is modelled as a Gaussian autoregressive process. For simplicity, we assume that X is a full rank matrix, so that $rank(X) = R$. The generative model of the data $y(t)$ at a given site is:

$$y(t) = \sum_{r=1}^{R} X_{t,r}\beta_r + \epsilon(t), \epsilon \sim \mathcal{N}(0, \sigma^2 V) \tag{2.3}$$

where β is a set of parameters that relates the observed signal to the model, σ^2 the noise variance and V is the normalized noise covariance matrix; $\mathbb{E}(\epsilon^T\epsilon) = \sigma^2 V$.

Let Y be an fMRI dataset, written as a $(N_t \times N_v)$ matrix (scans×voxel), and let X be the design matrix that describes effects of interest and confounds; the GLM proceeds by estimating the effect vectors β such that

$$Y = X\beta + \epsilon, \epsilon \sim \bigodot_{v=1}^{N_v} \mathcal{N}(0, \sigma^2(v)V(v)) \tag{2.4}$$

where ϵ represents the residual matrix. Importantly, the noise magnitude and covariance structure are voxel-dependent.

Estimation of the GLM parameters. Assuming that $V(v)$ is known in each voxel, the model parameters are estimated using the best linear unbiased estimate:

$$\hat{\beta}(v) = (X^T V(v)^{-1} X)^{-1} X^T V(v)^{-1} Y(v) \tag{2.5}$$

$$\hat{\sigma}(v)^2 = \frac{1}{N_t - R}\left(Y^T(v)V(v)^{-1}Y(v) - Y^T(v)V(v)^{-1}X(X^TV(v)^{-1}X)^{-1}X^TV(v)^{-1}Y(v)\right) \tag{2.6}$$

However, the voxel-based noise covariance matrix $V(v)$ is unknown. In general, the noise is assumed to be an AR(1) process [241, 243, 5]: assuming such a noise structure, the best linear unbiased estimator of β is obtained as follows: let $\epsilon = Y - X(X^TX)^{-1}X^TY$ be the ordinary least squares residual, the order 1 auto-correlation ρ is estimated in each voxel, possibly with some regularization [241], yielding an estimated covariance matrix $\hat{V}$ and the associated whitening operator $\hat{V}^{-\frac{1}{2}}$ is applied to both the data Y and the design matrix X, thus yielding their filtered counterpart $\tilde{Y}, \tilde{X}$. The effects, variance and degrees of freedom are then estimated using ordinary least squares on the whitened data

$$\hat{\beta}(v) = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y}(v) \tag{2.7}$$

$$\hat{\sigma}(v)^2 = \frac{1}{N_t - R}\left(\tilde{Y}(v)^T\tilde{Y}(v) - \tilde{Y}(v)^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y}(v)\right) \tag{2.8}$$

$$\hat{\nu} = N_t - R \tag{2.9}$$

Contrasts and tests. Let c be the linear combination of the experimental conditions that is of particular interest; c is also called a functional contrast. The next part consists in making a statistical decision on the presence of an activation at any voxel v of the dataset (the dependence on v is implicit in the following equations). In terms of hypothesis testing, the alternatives are typically:

$$(H_0) \quad c^T\beta = 0 \tag{2.10}$$

$$(H_1) \quad c^T\beta > 0 \tag{2.11}$$

According to Neyman-Pearson lemma, the uniformly most powerful test to decide which hypothesis is true is the Likelihood ratio test: Let $\mathcal{L}(Y)$ be the likelihood of the data,

$$\Lambda = \frac{sup_{c^T\beta=0}\mathcal{L}(Y|\beta)}{sup_{c^T\beta>0}\mathcal{L}(Y|\beta)} \tag{2.12}$$

Moreover, assuming that the noise covariance is known, Λ is a monotonous function of the exhaustive statistic t (Student, for a signed test) or F (Fisher, which is unsigned and can be defined for multidimensional contrasts); see Appendix B.1 for a sketch of the proof

$$t(v) = \frac{c^T\hat{\beta}(v)}{\hat{\sigma}(v)\sqrt{c^T(X^TVX)^{-1}c}} \tag{2.13}$$

$$F(v) = \frac{Tr(c^T\hat{\beta}(v)\hat{\beta}(v)^Tc)}{\hat{\sigma}(v)^2Tr(c^T(X^TVX)^{-1}c)} = \frac{Y(v)^T(P_X - P_R)Y(v)}{Y(v)^T(I - P_X)Y(v)}\frac{N_t - R}{R - rank(c)} \tag{2.14}$$

where $P_X = X(X^TX)^{-1}X^T$, and $P_R = Xc(c^TX^TXc)^{-1}c^TX^T$. Note that we write t and F as functions.

Note that in general, the equivalence between the F test and the likelihood ratio test does not hold when the noise covariance is unknown [36]. However, the t or F tests are preferred in general in the neuroimaging community due to their simplicity, and because their significance can be readily obtained by comparing their values to Student/Fisher distributions, while the likelihood ratio test cannot be assessed without (unfulfilled) asymptotic assumptions. In the remainder of this document, we will refer to the first level decision statistic as ϕ.

Although we did not write it explicitly, the model can easily deal with multiple session in this formalism: the data are concatenated along the temporal dimension, while the design matrix has to be written as a Kronecker product of the mono-session design matrices; the only minor difference is that the residual parameters have to be estimated for each session.

2.3.3 Thresholding

Once a statistical map is constructed at the individual or group level, it remains to be decided which brain regions are worth reporting, hence to derive a threshold defining brain regions with significant activity. The principle on which this threshold is chosen varies across studies. Three types of statistical

threshold are used. Uncorrected thresholds ϕ_c are defined to control for the risk of false positive at the voxel level.

$$p(\phi > \phi_c | H_0) < \alpha \tag{2.15}$$

If only one region or voxel is tested, this is a valid approach. Note that when the GLM is written in a Bayesian framework as in appendix B.2, the test is performed on posterior probability maps

$$p(c^T \beta > \delta | Y) > 1 - \alpha, \tag{2.16}$$

where δ is some reference value (0 typically). When uninformative priors are used, and when $\delta = 0$ the tests corresponding to 2.15 and 2.16 are equivalent. We do not further consider Bayesian tests in this section.

Often however, a priori localisation is not known, and because a brain volume contains several tens of thousands of voxels, a large number of regions are likely to be reported by chance with this kind of threshold. To prevent false positive results, several techniques have been derived for controlling the risk of error family wise (across voxels or regions). The most straightforward way consists in correcting the threshold ϕ_c for the number of tests performed -i.e. the number N_v of voxels- assuming independence. This is known as the Bonferroni correction. ϕ_c has to adjusted so that

$$p(\phi > \phi_c | H_0) < 1 - (1 - \alpha)^{1/N_v} \simeq \frac{\alpha}{N_v}. \tag{2.17}$$

This approach is correct in general cases [242], but, since *ı)* the independence hypothesis is trivially violated and *ii)* the spatial structure of the data is not accounted for, this approach is not very sensitive.

More sophisticated approaches have based this thresholding on the notion of the expected number of above threshold regions approximated by the Euler characteristic of random fields. These approximations have been extended to a number of statistical fields (t, F, χ^2) and geometry (any volume, surface, 4D data) [243]. For instance, the expected Euler characteristic χ_c of a Gaussian Random Field ϕ thresholded of dimension D (2 or 3) in a volume V, at level ϕ_c is

$$\mathbb{E}(\chi_c) = \Omega |\Lambda|^{\frac{1}{2}} (2\pi)^{-\frac{D+1}{2}} H_D(\phi_c) e^{-\frac{\phi_c^2}{2}} \tag{2.18}$$

where Λ is the covariance matrix of the field (basically, its smoothness or inverse point spread function) and H_D the Hermite polynomial of degree D. Now, for a high value of z, one has

$$P(\phi_{max} \geq \phi_c) \approx P(\chi_c > 1) \approx \mathbb{E}(\chi_c) \tag{2.19}$$

Equations (2.18) and (2.19) together give a new way of setting a threshold on a smooth Gaussian random field. A concept of interest is the -dimensionless- notion of the number of resolution elements ($RESELS$): This number is given by

$$RESELS = \frac{\Omega}{\prod_{i=1}^{D} FWHM_i}, \tag{2.20}$$

where $FWHM_i$ represents the full width at half maximum of the spatial filter associated with the map, in direction i, and related to the smoothness of the field by

$$\Omega|\Lambda|^{\frac{1}{2}} = RESELS(4\log(2))^{\frac{D}{2}} \tag{2.21}$$

This method has been further improved with the introduction of the spatial extent of the activated areas [71], joint test on the height and size of supra-threshold clusters [168], and the robust estimation of smoothness in presence of activations [106]. Note that this approach is compatible with spatial smoothing of the data (more precisely, this model encourages smoothing, to fulfill the hypotheses of the model [88].

These tests are computationally efficient but also rely on several assumptions, including a heavy smoothing of the volume to be thresholded which contradicts the MRI physicist efforts to improve image resolution. A few approaches have been proposed to use permutation-based non-parametric framework, but they face several difficulties due to the non-whiteness of the residuals of the GLM [67].

The third approach uses the false detection rate (FDR) threshold that controls for the false positive rate as a function of the number of reported voxels. The threshold ϕ_c corresponds to a FDR value of α is

$$p(H_0|\phi > \phi_c) = \frac{p(\phi > \phi_c|H_0)p(H_0)}{p(\phi > \phi_c)} < \alpha \tag{2.22}$$

Following [76], the estimation of the best threshold given α is usually carried out based on the Benjamini-Horchberg procedure; some more robust alternatives have been proposed recently [180]. Due to its definition, the FDR yields typically a less severe threshold than the family-wise error control procedures (2.17-2.18).

In general, thresholding is a trade-off between sensitivity and specificity, which may depend on the application purpose although this is rarely acknowledged in the literature. The defined regions are then reported using their local maxima in the standard space.

Note that cluster-level thresholding can be more sensitive, given that the truly active voxels are generally spatially clustered. Among the previously discussed techniques, only the Gaussian Random Field Model can be used in this perspective. Several alternatives have been proposed, such as cluster mass thresholding [190], but the calibration of this statistic in the intra-subject framework is still problematic.

2.3.4 Detecting modules versus establishing functional connections.

While the techniques described previously aim at localizing the activity in the brain, and therefore defining spatially defined functional modules, an increasingly large part of the literature is now devoted to establishing the functional connections between brain regions. The original observation in [11] shows that even during no motor activity (resting state or other such as visual stimulation), BOLD signal of a series or region that respond to motor tasks are correlated. Since then, two main approaches are concurrently explored. The first one tries to extract networks of correlated activity (or sharing some information) and techniques such as principal component analysis, independent component analysis (ICA) [133], probabilistic ICA [7], partial least square (PLS)[131, 105], various clustering techniques, self

organizing maps [159]. Escaping from the linear decomposition framework, kernel- and manifold learning-based techniques have also been proposed and can certainly be more sensitive [212, 199]. Such methods aim at defining the various functional networks that underlie brain activity and their relation to external tasks or stimulation. These exploratory approaches often suffer from a lack of interpretability.

The alternative approach consists in defining a specific network, a graphical model, choosing a priori the nodes and the structure of the graph, and to estimate the functional links given the experimental paradigm. This led to the development of structural equation models [132] and dynamical causal models [73, 160, 23]. By definition, the results of such procedures depend heavily on the regions introduces as nodes of the graphical model, and introducing many regions may yield ambiguous results [15].

Furthermore, the steady states used in many functional connectivity studies are not always well-defined states, and hence it would be of interest to extend the notion of functional connectivity to states that are controlled to a larger extent by the experimenter and that follow a predefined dynamic. Finally, signal similarity does not only come about by functional interaction, but can be influenced by confounding physiological effects of no interest like heart beat or artifacts like subject motion [41].

2.4 Spatial models for fMRI data

Several alternatives spatial models have been considered for modelling the spatial correlations fMRI data modelling.

- The first one consists in using the theory of Gaussian Random Field (GRF) to obtain estimates of the significance of activity peaks in fMRI maps
- The second one consists in using a Markov Random Field (MRF) model to model the spatial smoothness of the *true* activity maps and/or of the background noise and detect the active regions more sensitively.
- The third one consists in using spatial wavelets to decompose the fMRI signal.

Given our own interest on spatial models, we give a quick overview of these approaches. It should be reminded that all these developments have been performed in the framework of intra-subject analyses.

2.4.1 Gaussian Random Fields

Modelling the noise as a GRF: In many softwares used for activation detection, the thresholding procedure is based on a Gaussian Random Field (GRF) model, where the reference GRF has the smoothness of the data and is assumed to be centered. As detailed in section 2.3.3, the Euler characteristic of the excursion sets of such GRFs is estimated analytically, and this readily provides a significance of activation values for high threshold. The problem is that these thresholds are correct only for very smooth data, and for high threshold values.

An interesting alternative has been proposed in [242], where the probability p_{dlm} of detecting local maxima above some threshold ϕ_c is controlled, instead of the probability of detecting any voxel above such a threshold:

$$p_{dlm}(\phi_c) = \sum_{v=1}^{N_v} p((\phi(v) > \phi_c) \text{ and } \phi(v) > \phi(w), \forall w \sim v, \tag{2.23}$$

where $v \sim w$ denotes neighboring relationship in the spatial domain. This model further requires that the spatial covariance function between two sites x_1 and x_2 is of the form $\exp\left(-(x_1 - x_2)^T \Sigma^{-1}(x_1 - x_2)\right)$.

The derivation has been performed for Gaussian fields, hence it requires a conversion from standard Student or Fisher fields to Gaussian fields. Let $\mathcal{N}$ be the standard Gaussian density, and $\mathcal{M}(z) = \int_z^\infty \mathcal{N}(u)du$; let ρ be the spatial autocorrelation, $\alpha = \arcsin\left(\sqrt{\frac{1-\rho^2}{2}}\right)$, $h = \sqrt{\frac{1-\rho}{1+\rho}}$ and

$$q(\rho, z) = \frac{1}{\pi} \int_0^\alpha \exp\left(-\frac{h^2 z^2}{2(\sin\theta)^2}\right) d\theta$$

Then, if we define $\mathcal{Q}(\rho, z) = 1 - 2\mathcal{M}(hz^+) + q(\rho, z)$, where $z^+ = z.(z > 0)$, p_{dlm} is defined as:

$$p_{dlm}(\phi_c) = \sum_v \int_{\phi_c}^\infty \left(\prod_{d=1}^D \mathcal{Q}(\rho_d, z)\right) \mathcal{N}(z)dz, \tag{2.24}$$

where the central product is over the volume dimensions. This procedure has been shown in [242] to provide more sensitive tests than the standard Gaussian random field theory to assess the significance of local maxima.

Modelling the data as a GRF: Another perspective has been proposed in [85], in which the noise covariance process is assumed to be (spatially) white, while the model parameters β are given a smooth spatial prior

$$p(\beta) = \mathcal{N}\left(0, \lambda \exp\left(-L(\beta)\tau\right)\right), \tag{2.25}$$

where $(\lambda > 0, \tau > 0)$ are the scalar parameters of the model (to be estimated), and L is some operator that represents the spatial structure of the image, and possibly depends on β itself (to model the image anisotropy). The estimation of the parameters follows through an EM procedure; the posterior of the parameters $p(\beta|Y)$ can finally be derived: the marginals at each voxel yield the probability that the given voxel is activated. With respect to traditional univariate estimates, this Gaussian process approach allows some (possibly anisotropic) spatial regularization; in [85], this is at the expense of the unrealistic assumption of spatially white noise. The merit of this approach has not been established yet in practical situations.

2.4.2 Markov Random Fields

Markov Random Fields (MRFs) have been introduced in fMRI thresholding procedures to introduce the prior knowledge that realistic brain activations are spatially extended, hence smooth fields, while part of the noise (measurement error) is spatially white; under such hypotheses the compromise between specificity and sensitivity can be improved by imposing a Markov Random Field prior on the voxel labels: let Z be the label map, i.e. the set of values $(Z(v))_{v=1..N_v}$; it is associated with the prior $p(Z)$; let ζ be a particular configuration for Z:

$$p(Z=\zeta) \propto f(\xi) \exp\left(-\frac{\xi}{4}\sum_{v=1}^{N_v}\sum_{w\sim v}\mathbb{I}(\zeta(v)\neq\zeta_w)\right), \tag{2.26}$$

where $f(\xi)$ is the normalization constant of the MRF, which cannot be computed explicitly, and the relation $w \sim v$ denotes the fact that v and w are neighbours. It is then standard to assume conditional independence of the likelihood:

$$p(Y|Z,\theta) = \prod_{v=1}^{N_v} p(Y(v)|Z(v),\theta), \tag{2.27}$$

where θ is a set of parameters of the likelihood. The MRF definition requires a parameter ξ that controls the smoothness of the MRF, i.e. the spatial regularization of the labels implied by the model. The posterior distribution of the unknown parameters given the observed data Y is then

$$p(Z=\zeta,\theta,\xi|Y) \propto p(Z=\zeta|\xi)p(\xi)p(\theta)\prod_{v=1}^{N_v} p(Y(v)|Z(v),\theta) \tag{2.28}$$

The likelihood model is generally fairly simple, each components being modeled by a Gaussian. Note that this model corresponds to a kind of spatially constrained mixture model. The estimation of θ is thus no more difficult than estimating the parameters of a Gaussian Mixture Model (GMM). A two-classes or three classes model is usually chosen, in order to model the three competing hypotheses:

$$\begin{aligned} H_0 &: c^T\beta = 0, \\ H_1 &: c^T\beta > 0, \\ H_{-1} &: c^T\beta < 0, \end{aligned}$$

where the relation of the hypotheses with the observed data conforms to the model detailed in Section 2.3.2. Assuming that ξ is known, inference of θ and Z can be carried out either through MCMC sampling or Variational Bayes inference (which is only approximate, but in general relatively accurate [238, 155]), or by using the efficient minimum cut approach [98]; the latter provides the optimal solution for a two-classes problem (only), but is still much more expensive than Variational Bayes [155].

The main problem with this model is that ξ is unknown a priori. One solution consists in estimating the model with different values of ξ, and to choose one a posteriori, but it is even not clear how to choose

the best solution (moreover, the requires a partition of the data into training, validation and test set). A second possibility is to use a full Bayesian approach, where all the quantities are obtained through MCMC sampling, but this solution is particularly inefficient [239]. A third solution consists in replacing the discrete label model with a continuous model, which is reminiscent of the Gaussian Random Field approach described in Section 2.4.1; inference on this continuous model can be much more efficient than on the discrete model. An elegant solution based on Variational Bayes approximation can be found in [238].

An extension of the basic fMRI data-based MRF model has been proposed to account for anatomical information, such as the probability of one voxel being in grey matter, white matter or cerebro-spinal fluid. The generative model consists in introducing a second observation w, namely the anatomical voxel type in the model. The set of states Z becomes richer with a number of (activation status $\times$ tissue type) combinations (9 in the present case). Eq (2.28) becomes

$$p(Z = \zeta, \theta, \xi | Y, W) \propto p(Z = \zeta|\xi)p(\xi)p(\theta) \prod_{i=1}^{N_v} p(Y(v)|Z(v), \theta)p(W(v)|Z(v)) \tag{2.29}$$

This model has been shown in [155] to provide sensible constraints in the detection of active regions.

In spite of this, Markov Random Fields have received little interest in the neuroimaging community. The shortcoming of this model is that the topography of the activation maps is relatively poorly modeled with the generative model $p(Y|Z, \theta)$; while this model is optimal for segmenting *flat* activation patterns corrupted by white noise, they do not carry striking improvements in real cases [238, 155]. Finally, the evaluation of these approaches in the context of group analyses still has to be done.

2.4.3 Spatial wavelets

Finally, another generic alternative to model the spatial distribution fMRI activity consists in using spatial wavelet bases. The approach is simply to decompose the signal onto the basis, and then to use a wavelet-based shrinkage to filter the activation and noise parameters i.e. the wavelets coefficients of the parameters maps associated with low ϕ values are discarded [227]. The reconstruction of the data putatively provides denoised activation estimates. However, this model is not generative, and should be regarded as a post-processing trick, which is not free of artifacts. More complex strategies may be necessary may then be necessary to reduce the bias in the resulting maps [229].

A more sophisticated framework has been presented in [61], in which the parameters are given a prior

$$\beta = \gamma W + E, \tag{2.30}$$

where W represents a set of spatial basis function (typically wavelet-based). At each resolution level, the wavelet coefficient are then modeled as a mixture of two Gaussian densities to separate the noise coefficients from the signal coefficients. The parameters and hyperparameters of the model are then estimated through a Variational Bayes approach, thus providing posterior estimates for $p(\beta|Y)$. The resulting posterior probability maps for the contrasts $p(c^T\beta|Y)$ can then be used for assessing the presence

of effects. The whole procedure is finally shown in [61] to better preserve the shape of activations than Gaussian Markov Random Fields or smoothing.

Still, such a procedure is blind to the underlying anatomy, and the ultimate benefit in the case of group studies has not been proved.

2.5 Dealing with the absence of a brain template

In this section, we describe a problem that underpins neuroimaging group analysis: namely that statistical inference methods require that the brain locations or structures have to be matched across subjects. We first describe the standard framework, which is based on a so-called spatial normalization procedure, then discuss its shortcomings.

2.5.1 Spatial normalization

There is a consensus in the neuroimaging community on the way to perform multi-subject analyses: data are first coregistered to a reference template and re-sampled (stereotactic normalization, see e.g. [72]). Activation is detected in a mass univariate framework using a General Linear Model at the individual level, then at the group level through the use of mixed-effects models.

While there are many different techniques to perform the spatial coregistration, the most usual procedure is to first realign the functional volumes to the anatomical volume acquired in the same scanning session, and use this more detailed image to derive a deformation field that warps the subject brain anatomy to a standard template (generally the so called ICBM152 volume image which represents the average of 152 healthy T1 brain images). This template corresponds (but only approximately) to a brain neurosurgical atlas, the Talairach and Tournoux atlas [196].

The coregistration of the functional data onto the anatomy is relatively simple, because an affine deformation model is usually sufficient (if we ignore the local effect of residual EPI distortions, which is hard to correct for anyway [96]). The standard matching criterion is normalized mutual information, given that there no simple statistical relationship between the T1 and the T2 contrast.

The coregistration of anatomical image to the template is more involved [4]. The matching criterion is usually the sum of squared differences after histogram equalization of the smoothed input and target images. More recent approaches work on pre-processed anatomical images, where the tissue type probability maps have been extracted [6].

An affine coregistration is first performed to roughly align the images. Then, in [4] a set of deformation parameters are computed in a basis of smooth spatial deformations, typically the low frequency part of a discrete cosine transform of the ambient (3D) space. The number of degrees of freedom in the deformation can be controlled through the number of basis deformation used. The bending energy is also included in the criterion to minimize the risk of unrealistic deformation patterns. There is no guarantee that the deformation is invertible.

More recent approaches use diffeomorphic registration algorithms. Although he conceptual framework goes back to [22], the application to neuroimaging has become important only since [6]. In this framework,

the deformation is defined within some pre-defined spatial basis, and the smoothness is also enforced by placing shrinking priors on the deformation coefficients. The diffeomorphic framework with constant vector field can then be handled as traditional small deformation problems using the scaling and squaring procedure. The efficient implementation described in [6] makes it usable in large scale datasets. With respect to previous alternatives, this algorithm improves the coregistration of many brain structures (see [109]. for a comparison of many coregistration approaches). Finally, it should be pointed that brain registration can be surface-based instead of merely anatomy-based. We discuss this in Section 2.6.

2.5.2 Shortcomings of this procedure

The SPM procedure within the stereotactic normalization framework is challenged by:

- The residual inter-subject variability that is known to remain after normalization. Indeed, the algorithms that warp a subject anatomy to a template do not and cannot perform a perfect match. The information used for the warping is the main anatomical contrasts (deep sulci, ventricles) but the variation of the anatomy between subjects is such that there is no obvious point to point correspondences between subjects. Based on anatomical [91] or functional landmarks [193] this variability has been evaluated as about 1cm in many cortical regions. This residual mismatch should not only be seen as a shortcoming of normalization procedures that can be improved, but rather as an effect of the intrinsic variability of cortical shape, folding patterns [173], cytoarchitecture [56] and functional organization [236].

- Indeed, even if the anatomy could be matched perfectly, this would not necessarily be the case for functional regions, given that the function/structure matching implicit in brain mapping is only approximately true. For instance, the Fusiform Face Area (responding more to face than to objects) may be localised more anteriorly in the fusiform gyrus in one subject compared to another.

- The difficulty of assessing the functional specialization of brain regions, given that regions of interest are not entirely correctly characterized by their coordinates in the template space or with respect to sulco-gyral anatomy, but also by their relative position with respect to anatomical landmarks, their extent or their connections to other regions [14].

Several approaches are currently used to limit the impact of inter-subject variability on functional MRI (fMRI) group analyses. First, spatial smoothing is frequently applied to the data, thus sacrificing resolution for sensitivity in voxel- and cluster-level detection. Another approach consists in defining functional Regions of Interests (ROIs) using a *localizer* experiment, and then to restrict the group analysis to these particular regions [179]. This approach is particularly well suited to study some fine-grained functionally specialized regions [179]. While it is often more sensitive, the investigation is restricted to the functional ROI defined and is also questionable from the perspective of the reliability of region definition [70]. Anatomical atlases can be used to define ROIs in the common space (see e.g. [226, 49, 184], but this approach precludes any analysis of inter-subject anatomical variability.

2.6 A quick overview of anatomical data processing

In this section, we give a quick overview on how to process anatomical data in the perspective of functional data analysis and of assessing anatomo-functional structure correspondences. It is important to notice that this approach complements the iconic normalization described in the previous section, in the sense that it aims at recognizing different brain structures instead of standardizing brain shapes. This approach specifically aims at preserving the idiosyncratic structure of each individual dataset. We focus on the cortical parts, given that sub-cortical structures are usually handled separately, based on dedicated methods (see e.g. [170]). Among several alternatives [54, 52], we take as a reference the Brainvisa pipeline (among others, see [173, 130]). Anatomical data typically includes the following steps:

- An approximate pre-registration of the data to the Talairach template based on the anterior commissure/posterior commissure definition.
- A correction of the bias in the anatomical image, which is important to improve the detection of grey/white matter [128].
- The detection of the grey/white matter and cerebro-spinal fluid (csf) based on the analysis of the histogram of the data. This may or may not include spatial constraints.
- Segmentation of the grey matter under the constraint of making it have a spherical topology. This provides two interfaces, the grey/white matter interface and the grey matter/csf interface.
- Segmentation of the brain sulci, i.e. the inward folds of the cortical surface. The difficulty here is to decide when sulci end and starts, and what the meaningful portions are
- Sulci recognition follows yes, using a basis of 20 manually annotated brain database, based on pattern recognition technique [173, 164].
- Projection of the sulci fundus onto the cortical surface. The difficulty here is to defined reliably the ends of the sulcal lines.
- Derivation of a surface-based coordinate system. This may take into account identified sulci [27] or simply the surface geometry [54]. In general a spherical coordinate is used; since the spherical coordinate system can certainly not be isometric, it has to decide which property has to be preserved. For instance, conformal mapping preserve the angles, possibly at the expense of large areal deformation [234].
- Segmentation of the cortical surface into gyri; this operation can be merely coordinate-based: the definition of cortical gyri in the reference sphere is simply pulled back onto the individual brain meshes using the previously defined coordinate system [26].

The whole procedure can be considered as performing a registration of cortical surfaces; as such it may be relatively efficient [56]. However, it is also rooted in deeper considerations on the structure and

the evolution of brain folding patterns [222, 27]. Alternative procedures can also take place to combine 2D and 3D information in normalization, but they are not a standard yet.

Recently, it has been proposed to direct align cortical meshes by using an iterative search for optimal vertex correspondences based on feature fields (that basically represents the distance of the nodes to the outer envelope of the cortex) and a regularization to preserve local topology [125]. The procedure is applied in a multi-scale fashion with alternate matching/template estimation steps. The meshes are then re-parametrized, and cortical averages can be constructed, which preserve some salient features of cortical geometry (although some regions are more problematic). It is shown in [125] that about 50 subjects are necessary to yield a bias free template. Still folding variants exist and it is suggested to have several population averages to represent all the folding variants observed in the different subgroups of subjects. The problem of unbiased anatomy representation is and will probably remain open for several years.

Chapter 3

Statistical inference on multi-subject fMRI data

In this chapter, we develop some of our work of group analyses for fMRI data. This is organized in three contributions:

- In Section 3.1, we introduce the main univariate models used for group analysis, and introduce the main definitions that are necessary for discussing group-level analysis methodology. We give an overview of some contributions published in [141, 175, 102].

- In Section 3.2, we detail our work on the reproducibility of group-level brain activation maps, which has been published in [204].

- We end up with an original work on multi-subject parcellation for group inference; this is an updated version of the work published in [203].

Here we concentrate on the problem of one-sample inference, in which it has to be decided where in a certain common brain space there is a positive effect across subjects. It is assumed that a one-dimensional contrast c of experimental conditions has been defined in a coherent manner across subjects; for simplicity, this contrast-related effect $c^T\beta$ will be denoted β.

3.1 Classical statistical models

Classical statistical models refer to mass-univariate models, in which the functional images are assumed to be coregistered. According to the first-level analysis described in Sec. 2.3.2, the data consists in vectors $\hat{\beta}_s = (\hat{\beta}_s(v))_{v\in\{1..N_v\}}$ where N_v denotes the number of spatial sites (voxels), and $s \in \{1, .., S\}$ denotes the subject. These vectors represent the estimated effect for a certain contrast. Moreover, the

variance of the estimator of β_s is also estimated, and provided as a corresponding set of scalars $\hat{\sigma}_s^2$ [1]. This provides variance values for each site and subject. The degrees of freedom $(\nu_s)_{s\in 1,..,S}$ are assumed to be known, and constant across voxels, but not necessarily across subjects.

3.1.1 Fixed effects and mixed effects models

The simplest problem consists in deciding, for a given position v, whether there is a positive effect at the population level, given the observations $(\hat{\beta}_s(v))_{s\in\{1,..,S\}}$ and the associated variances. This simple question can receive multiple answers, depending on whether the null hypothesis is that no subject activated (fixed-effect model, FFX) or that the average in the population is not different from zero (mixed-effect model, MFX). Usually, only the second question is considered, since the result of the fixed effect inference is valid only on the cohort at hand, and not on the whole population from which the cohort has been sampled [68]. Note that other question could be asked, such as whether at least half (or any other proportion) of the subjects showed a positive effect (see [90]).

Within the mixed-effects model, it is often possible to neglect the first-level variance $\hat{\sigma}_s^2(v)$, given that this variance source is in general largely dominated by the second level variance of $(\beta_s(v))_{s\in 1,..,S}$. In that case, the model is traditionally referred to as *random effects* (RFX), under the convention that the subject is the random factor in the model. However, it may be important to take the first-level variance into account [175]. In fact, the so-called RFX model can be seen as an approximation of the correct MFX model.

The MFX model can be written formally as (here and in the sequel the dependence on v will be omitted, although all the quantities are in fact voxel-dependent):

$$\hat{\beta}_s = \beta_s + \varepsilon_s, \varepsilon_s \propto \mathcal{N}(0, \sigma_s^2) \quad (3.1)$$

$$\beta_s = \bar{\beta} + \eta_s, \quad (3.2)$$

where $\bar{\beta}$ denotes the parameter of interest, i.e. the group average response. The first equation is simply a restatement of the intra-subject General Linear Model, and the second one is the group model.

3.1.2 Classical hypothesis testing

In order to perform hypothesis testing, the usual approach consists in assuming a certain distribution for the effects $(\beta_s)_{s\in 1,..,S}$, which will be parametrized by $\bar{\beta}$; this distribution will thus be denoted $p(\beta|\bar{\beta})$. This distribution is usually chosen as normal (we discuss the validity of this hypothesis in section 3.2)

The following likelihood ratio (LR) test can thus be defined:

$$\Lambda = \frac{\sup_{\bar{\beta}=0} \mathcal{L}(\beta|\bar{\beta})}{\sup_{\bar{\beta}\in\mathbb{R}^+} \mathcal{L}(\beta|\bar{\beta})}, \quad (3.3)$$

[1] One should take care that this is not the same term as in Eq. (2.6), where σ represents the variance of the residuals of the model. Here this term represents in fact the contrast related variance $\sigma_c^2 = c^T(X^TVX)^{-1}c$. But, Given that c is implicit in this chapter, no confusion is possible, and we retain the notation σ.

We denote the set of possible values for $\bar{\beta}$ as Θ (unconstrained case, here $\mathbb{R}^+$ or $\mathbb{R}^\star$ for a two-sided test), and Θ_0 (constrained case, here $\{0\}$). The likelihood ratio test rejects the null hypothesis H_0 when its inverse exceeds a critical value λ_α. That is, the decision rule has the form:

- if $\Lambda^{-1} \geq \lambda_\alpha$ reject H_0.
- if $\Lambda^{-1} < \lambda_\alpha$ accept (or do not reject) H_0.

The critical value λ_α is usually chosen to obtain a specified significance level α, through the relation: $P_0(\Lambda^{-1} \geq \lambda_\alpha) = \alpha$. The Neyman-Pearson lemma states that this likelihood ratio test is the most powerful among all level-α tests for this problem. Under the normal hypothesis, and assuming that $\forall s \in \{1, .., S\}, \sigma_s^2 = 0$, the likelihood test simply boils down to the standard t-test,

$$t = \frac{mean_s(\hat{\beta}(s))\sqrt{S}}{\sqrt{var_s\hat{\beta}(s)}} \tag{3.4}$$

as shown in Section B.1.

In more general cases, however, the exact distribution of the likelihood ratio corresponding to specific hypotheses is very difficult to determine. A convenient result, though, says that as the sample size n approaches ∞, the test statistic $-2\log(\Lambda)$ will be asymptotically χ^2 distributed with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0, which is 1 here. This readily provides a statistical inference procedure, but the latter is only asymptotically correct. This is an issue here, given that the number S of observations is low (typically 10 to 20). Non-parametric procedures can be used as an alternative (see below).

Before analyzing the Gaussian MFX model in more detail, it should be noted that different alternatives can be used as statistics: for instance, the sign statistic, the Wilcoxon signed rank statistic or the empirical likelihood ratio. Mixed-effects model extensions (i.e. models that incorporate the information of first-level variance) have been introduced and discussed in [175]. Let us just recall the definition of Wilcoxon's signed rank statistic (WKX) [92], which sorts the absolute effects in ascending order, then sums up the ranks modulated by the corresponding effect's sign:

$$W = \sum_{s=1}^{S} sign(\hat{\beta}_s) rank(\hat{\beta}_s) \tag{3.5}$$

Interestingly, the calibration of this statistic is straightforward and voxel-independent; it depends only on S.

3.1.3 The Gaussian mixed effect model

Given its popularity in the literature [243, 64, 240, 175, 204], it is important to emphasize the detail of the Gaussian mixed effects model. This model writes:

$$\hat{\beta}_s = \beta_s + \varepsilon_s, \varepsilon_s \propto \mathcal{N}(0, \sigma_s^2) \quad (3.6)$$
$$\beta_s = \bar{\beta} + \eta_s, \eta_s \propto \mathcal{N}(0, v_g) \quad (3.7)$$

Let us recall that $(\hat{\beta}_s)$ and σ_s^2 are assumed to be known. This relies on the approximation $\sigma_s^2 = \hat{\sigma}_s^2$, which is usually accepted, assuming that the first-level model has been correctly specified, and given the number of degrees of freedom ν_s is large enough (typically ≥ 100) to provide an accurate estimate of σ_s^2. Given that both v_g and $\bar{\beta}$ are unknown, there is no closed-form formula for estimating the model. An Expectation Maximization algorithm has been proposed in [175], restricted maximum likelihood has been used in [243], Markov Chain Monte Carlo methods in [240]. In fact, in this simple model, the maximum likelihood solution can be reached very efficiently (see Fig. 3.1.3)

The joint estimation of the group effect and the group variance proceeds from Eqs. (3.6-3.7). At a voxel v, S values of estimated effects $\hat{\beta}$ are available, together with S estimates of the associated variances $(\hat{\sigma}_s^2), s \in \{1..S\}$ (we drop the voxel index v for simplicity). We also assume that the estimated variance is correct, so that $\sigma_s^2 = \hat{\sigma}_s^2$, $\forall s \in \{1..S\}$ (note that the estimator generally relies on $\nu > 100$ degrees of freedom).
For this model, the log-likelihood of the data is written as:

$$\mathcal{L}((\hat{\beta})|\bar{\beta}, v_g) = \text{cst} - \frac{1}{2}\left(\sum_{s=1}^{S} \log(\sigma_s^2 + v_g) + \sum_{s=1}^{S} \frac{(\bar{\beta} - \hat{\beta}_s)^2}{\sigma_s^2 + v_g}\right) \quad (3.8)$$

maximizing $\mathcal{L}$ with respect to $\bar{\beta}$ while keeping v_g fixed yields:

$$\bar{\beta} = \sum_{s=1}^{S} \frac{\hat{\beta}_s}{\sigma_s^2 + v_g} \left(\sum_{s=1}^{S} \frac{1}{\sigma_s^2 + v_g}\right)^{-1} \quad (3.9)$$

while the minimization of $\mathcal{L}$ with respect to v_g, while $\bar{\beta}$ is fixed yields

$$\sum_{s=1}^{S} \frac{(\bar{\beta} - \hat{\beta}_s)^2}{(\sigma_s^2 + v_g)^2} = \sum_{s=1}^{S} \frac{1}{\sigma_s^2 + v_g} \quad (3.10)$$

Let $L(v_g)$ and $R(v_g)$ be the left and right hand side terms in Eq. (3.10). We solve it by iterating the solution of $L(v_g) = R$ in under the constraint $v_g > 0$ using Newton's method, then updating the right hand side term. This procedure always converges in a few iterations.

Finally, the following decision statistic can be formed:

$$\phi = \sum_{s=1}^{S} \frac{\hat{\beta}_s}{\hat{\sigma}_s^2 + v_g} \left(\sum_{s=1}^{S} \frac{1}{\hat{\sigma}_s^2 + v_g} \right)^{-\frac{1}{2}} \tag{3.11}$$

This decision statistic can be simply interpreted as a Bayesian test where the average size of the group effect is compared to the uncertainty of this effect. Note that it is not in general equivalent to the LR test - which remains a plausible alternative [175].

The next problem is to assess the significance of this statistic (LR or (3.11)): although μ is formally constructed as a Student statistic, it cannot be identified as such because of the heterogeneous variance terms. One conservative solution consists in using a Student distribution with $S-1$ degrees of freedom, see e.g. [240], but this solution is not very powerful. In that case, the significance of the statistic should rather be assessed in a non-parametric framework.

3.1.4 Non-parametric assessment of the statistics

Although the use of analytical bounds for statistical tests is both efficient and elegant, this can be performed in only very restricted cases; in the case of the Gaussian mixed-effect model (see above), there is no such solution. Thus we need to resort to non-parametric techniques to estimate the reference pdf $p_0 = p(\phi|H_0)$ for any statistic ϕ.

Resampling techniques are are usually used for that purpose [175]. In particular, in the case of group studies in neuroimaging, the sign permutation approach consists in tabulating p_0 by resampling the estimated effects across all possible flips of sign, the number of which is 2^S. Such a procedure is intended to test a simple null hypothesis $H_0 : \bar{\beta} = 0$ about the parameter of interest under the restrictive hypothesis that p_0 is symmetric.

In the present case, two additional conditions must be fulfilled [141]: *i)* the subjects are drawn independently *ii)* first-level estimators are location equivariant and scale-invariant. These conditions imply that the distribution of standard test statistics are stochastically increasing w.r.t. the population mean effect, which validates the use of sign permutations to perform one-sided tests, i.e. testing the composite hypothesis $H_0 : \bar{\beta} \leq 0$. Note that these hypotheses are weaker than the standard normality assumption.

The test is exact conditional on the effects values and first level variances up to the discretization induced by the finite number of permutations. The accuracy of the test is traditionally controlled with binomial statistics. The main challenge with this kind of procedure is of course computational, given the number of voxels and permutation to be considered.

3.1.5 Making inference on images

So far, we have considered only voxel-based inference. However, neuroimaging data consists of images, which implies that many tests are performed simultaneously or that inference is performed on more

complex objects. A known issue is that the expected number of false detections for a test performed on N_v independent sites at significance α is $N_v\alpha$. Controlling this later number amounts to applying a Bonferroni correction (see Sec. 2.3.3), which is correct for independent voxels, and typically conservative for positively correlated data as in neuroimaging (see e.g. [242] for a discussion on that point). In the case of non-parametric assessment procedure, it is however much more useful to control the family-wise error rate by tabulating the maximum statistic at the map level under the null hypothesis [175].

However, this kind of procedure rests only on signal amplitude, and is not optimal for spatially extended but weaker activations. For this kind of case -which is especially relevant in group studies, given the shortcomings of normalization discussed in section 2.5.1- a dual procedure might be preferred: *i)* in a first step, the map is thresholded at a relatively arbitrary threshold (corresponding to $p < 0.01$, uncorrected, typically); *ii)* in a second step the maximal cluster size is tabulated under using the non-parametric framework, i.e. sign flips of the individual data; under the null hypothesis, the resulting distribution represents the distribution of the maximal supra-threshold cluster size *iii)* finally the clusters whose size is greater than a proportion $(1-\alpha)$ of the tabulated distribution are reported as active. Here, a *cluster* refers to a set of spatially connected voxels assuming a certain model of spatial contiguity (6, 18 or 26 nearest neighbors typically; we use 18 in our implementation).

There exists an interpretation difficulty with this procedure: the fact that a cluster is significantly greater than expected by chance does not mean that all the voxels within this cluster are active, but that there is at least one voxel within the cluster for which the null hypothesis can be rejected; in statistical terms, this approach provides a weak control of false detections. Assuming that the clusters are not too large, this problem is not fundamental, given that voxel-level information has little relevance in group studies.

3.1.6 A Bayesian perspective

It is perfectly possible to rephrase the previous analysis based using Bayesian concepts [240]: Assuming a weakly informative prior distribution on $\bar{\beta}$ and v_g, posterior distribution can be derived for these parameters. Note that in that particular case, non-informative priors as in [240] yield an improper posterior, hence a non-identifiable model [144]. An interesting feature of Bayesian assessment is that the H_1 hypothesis can also be rejected, showing that a region is indeed inactive, but this possibility has rarely been used in the neuroimaging literature.

A recent contribution [102] has tried to include the cross-subject spatial variability in data modelling, as if the unknown displacement variable could be marginalized out for the estimation of the template map. But this turns out to be impossible, and the proposed model is not a global generative model of the data. Still, a statistical test can be defined, and the procedure proposed in [102] is shown to provide less biased group-level activation maps than standard univariate detection procedures (the final test is a classical test, not a Bayesian test).

In fact, the Bayesian point of view on inference has not been successful so far, for several reasons:

- Weakly informative priors have, by definition, little or no effect on the posterior, which thus essentially represents the likelihood term.

- Alternatively, non-trivial priors, e.g. spatial priors under a Markov Random Field can be used (see Section 2.4), but this boils down to a post-processing of the map, which is furthermore costly. There is still little evidence in favor of its usefulness in group studies ([78, 239]).
- More generally, Bayesian approach do not solve multiple comparison issues and do not fit well with cluster-level assessment of the activated regions. For all these aspects, the classical non-parametric procedures provide reliable p-values which are still preferred for interpretation purpose (and for publication in neuroimaging journals).

3.2 Analysing the reproducibility of group studies

3.2.1 Specificity, sensitivity, reproducibility

Let ϕ be a statistic that measures the positivity of a group-level parameter as defined in model (3.1-3.2); usually ϕ is defined at the voxel level in the template (normalized) space. The quality of the ensuing inference can be assessed by different criteria: specificity, sensitivity, and reproducibility.

Specificity is the number or rate of detections that are allowed by a certain detection procedure, and corresponds to assessing how frequently ϕ is above some threshold ϕ_α under the null hypothesis:

$$p(\phi > \phi_\alpha | H_0) < \alpha \tag{3.12}$$

This is the most important criterion, since it governs the probability of making false inference from a dataset, which is the most prominent pitfall for neuroscientists. Thus statistical approaches have been designed to control specificity at the map level, in particular with the use of Gaussian Random Field Theory [243], or non-parametric procedures (see Sec. 3.1.4).

Sensitivity or power is the capacity of the analysis to detect truly activated regions.

$$p(\phi > \phi_\beta | H_1) \geq \beta \tag{3.13}$$

Although this is of primary interest to neuroscientists, a lack of sensitivity is not a fundamental scientific issue because the failure to detect an effect does not allow one ton conclude that the effect is indeed absent: the (H_1) hypothesis is not rejected if $\phi \leq \phi_\alpha$, see Eq. (3.12). Sensitivity is clearly limited by the small number $S \sim 10 - 20$ of subjects in the experiments and by the lack of spatial correspondence between activated areas across subjects: in univariate procedure, the lack of spatial correspondence results in an inflated variance in the activated regions, hence a lower sensitivity of the test statistic [204]. Moreover, the data may be non-normal, which generally reduces the sensitivity of the standard statistical tests.

Reproducibility has rarely been considered so far, but it is also a fundamental aspect for neuroimaging inference: the results should be stable e.g. with respect to the cohort under study.

$$p(\phi > \phi_\gamma | \phi' > \phi_\gamma) \geq \gamma, \tag{3.14}$$

where ϕ and ϕ' represent the same statistic obtained from two different samples of subjects. In realistic conditions ($S \leq 20$ subjects), reproducibility is clearly limited by the number of false negatives rather

than the presence of false positives. This means that the requirement of increasing the reproducibility of the results in neuroimaging studies is largely consistent with the need to increase the sensitivity.

We now describe some of our contributions to quantify reproducibility in a useful manner.

3.2.2 Quantifying the reproducibility in group studies

We have proposed two measures to assess the reproducibility of the activation maps derived from group analysis. The first, based on a mixture of binomial distributions, characterizes the stability of the status (active/inactive) of each voxel of the dataset. The second measures how frequently clusters of voxels are found at similar locations in the normalized MNI/Talairach space across groups of subjects. We use these measures in a jackknife subsampling procedure that enables us to characterize the reproducibility of activation maps obtained at the group level.

Reproducibility measure at the voxel level. In order to estimate the reproducibility of a statistical model, we need a method to compare statistical maps issuing from the same technique, but sampled from different groups of subjects. We use the reliability indexes elaborated in [75, 119, 120]. Assume that a statistical procedure (e.g., thresholding) yields binary maps $g_1, .., g_R$ for different groups of subjects. At each voxel v, an R-dimensional binary vector $[g_1(v), .., g_R(v)]$ is thus obtained. At the image level, the distribution of $G(v) = \sum_{r=1}^{R} g_r(v)$ is modelled by a mixture of two binomial distributions, one for the null hypothesis, one for the converse hypothesis: Let π_A^1 be the probability that a truly active voxel is declared active, $\pi_A^0 = 1 - \pi_A^1$ the probability that a truly active voxel is declared inactive, π_I^1, the probability that a truly inactive voxel is declared active, $\pi_I^0 = 1 - \pi_I^1$ the probability that an truly inactive voxel is declared inactive, and λ the proportion of truly activated voxels. Then, using a spatial independence assumption, the log-likelihood of the data is written as

$$\log(p(G)|\lambda, \pi_A^0, \pi_I^0) = \text{cst} + \sum_{v=1}^{N_v} \log\left(\lambda(\pi_A^0)^{R-G(v)}(\pi_A^1)^{G(v)} + (1-\lambda)(\pi_I^0)^{R-G(v)}(\pi_I^1)^{G(v)}\right) \tag{3.15}$$

Assuming $R \geq 3$ the three free parameters, $\pi_A^0, \pi_I^0, \lambda$ can be estimated using an EM algorithm. Note that optimizing the model over its different parameters sequentially, and using an adequate initialization, we could run the model for $R = 2$, though with higher variability in the estimation. An example of mixture of binomial distributions is given in Fig 3.1.

Given these estimates, the coherence index κ, known as Cohen's kappa is computed to measure the concordance of the different observations with the mixture model. Let $p_0 = \lambda\pi_A^1 + (1-\lambda)\pi_I^0$ be the fraction of voxels that are correctly classified by the mixture model. p_0 should be compared to the fraction of correct classifications that occur by chance $p_C = \lambda\pi^0 + (1-\lambda)(1-\pi^0)$, where $\pi^0 = \lambda\pi_A^0 + (1-\lambda)\pi_I^0$ is the proportion of voxels declared inactive. The fraction of correct classifications corrected for chance is thus

$$\kappa = \frac{p_0 - p_C}{1 - p_C} \tag{3.16}$$

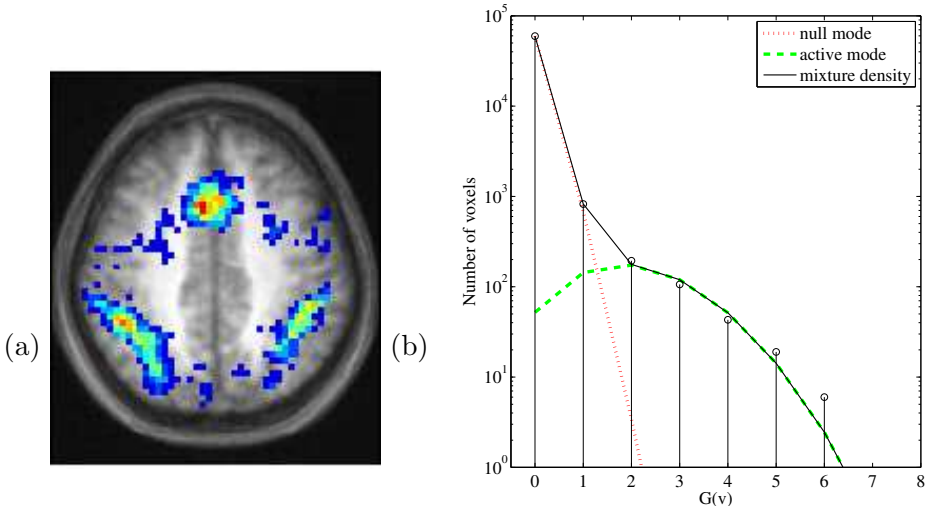


Figure 3.1: (a) Example of a reproducibility map G: the color codes for the number of times a given voxel has been declared active in a R-split of the population: from 0 (no color) to $R = 8$ (red). (b) Fitting the histogram of G with a mixture of binomial distribution. The empirical histogram of $G(v)$ is modelled by the model in Eq. (3.15), with $R = 8$. The Y axis is in log-coordinates for the sake of clarity.

In this setting, $0 \leq \kappa \leq 1$ measures the fit of the mixture model to the data, which in turns reflects the concordance of the binary maps given as input to the model (3.15). If there is very little agreement on which voxels are active, the two components of the mixtures overlap, and κ is close to 0, whereas the separation between the components of the mixture increases and κ is close to 1 if there is a good agreement between binary maps. λ can also be retained as an index of the test sensitivity.

Note that more complex -and realistic- models have been proposed in the literature [126], in which the parameter λ is allowed to vary spatially. However, our main purpose is not activation detection, but obtaining a global reproducibility measurement; for this reason, we keep the basic setting.

Reproducibility measure at the cluster level Another way to assess the reproducibility of the results is to compare the position of the clusters of supra-threshold voxels that arise through any group analysis. Assuming that the binary maps $g_1, .., g_R$ are obtained from different groups of subjects through a thresholding procedure, one can post-process them in order to yield connected components. The connected components with a size greater than a given threshold η are then retained, and their centre of mass (cm) is computed: let $x_i^r, i = 1..I(r)$ be the spatial coordinates of the cms derived from map g_r, we

propose the following average distance between any two maps:

$$\Phi = \frac{1}{R(R-1)} \sum_{r=1}^{R} \sum_{s \in \{1,..,R\}-\{r\}} \frac{1}{I(r)} \sum_{i=1}^{I(r)} min_{j \in \{1,..,I(s)\}} \phi(\|x_i^r - x_j^s\|), \tag{3.17}$$

where $\phi(x) = 1 - \exp\left(-\frac{x^2}{2\delta^2}\right)$ is a penalty function that is close to zero when the cluster centroids are properly matched and close to 1 otherwise. Φ represents the average mismatch between any *cm* of a supra-threshold component in a given map and the closest *cm* of any supra-threshold cluster obtained from another map. Appropriate penalty terms are used to handle the case $I(r) = 0$. We have performed some experiments using $\eta = 10$ voxels or $\eta = 30$ voxels, and use $\delta = 6mm$.

Procedure for the assessment of reproducibility The procedure consists in dividing the population of 81 subjects in $R = 2, 3, 4, 5, 6$ or 8 disjoint groups of $\mathcal{S} = 40, 27, 20, 16, 13$ and 10 subjects respectively. The computation of different statistics, the derivation of an adequate threshold and the thresholding are performed in the different subgroups, and global reproducibility measures are derived from the ensuing binary maps. This procedure is repeated 100 times for each instance, yielding a distribution of the indexes κ, λ and Φ for each possible technique/parameter.

3.2.3 Observations from a large cohort

The dataset used in this experiment is the subset of the first $S = 81$ subjects from the localizer database (see Sec. B.4.1).

First, we choose the traditional RFX analysis procedure [see Eq. (3.4)], thresholded at $p < 0.001$ uncorrected, using an analytical threshold and evaluate the distribution of the different indexes for three contrasts of interest. This is important to understand how well the indexes are characteristic of the amount, the spread and the variability of supra-threshold activity. In particular, it is important that the estimated reproducibility indexes are less variable for a given contrast than across contrasts.

Second, we evaluate the choice of the threshold on the different indexes, in the case of the voxel-based t-test. While the sensitivity index certainly decreases while the threshold increases, the behaviour of the reproducibility may be more complex, due to the trade-off between false positive and false negative rates (non-standard behaviours due to extremely low or high thresholds are not considered here).

Third, we study the behavior of the different measurements when the number of subjects in the group varies; while it is obvious that reproducibility increases with the group size, it is not clear whether there exists a plateau and at which level. Previous studies [38, 139] suggest also a steady increase of sensitivity with the group size.

Finally, we choose the following statistics: RFX, RFX on smoothed (12mm Full Width at Half Maximum [FWHM] instead of 5mm) effect maps (SRFX), MFX, Wilcoxon(WKX), Cluster-level RFX (CRFX), Parcel-based RFX (PRFX) and ΨFX. ΨFX is formally defined as mixed effect statistic in Eq. (3.11), in which the group variance v_g has been set to 0. This is also the statistic defined in [146]. However,

we calibrate it with the non-parametric framework, making it a correct group inference statistic. RFX, SRFX, MFX, ΨFX and PRFX maps are thresholded at the $p < 0.001$ level, uncorrected for multiple comparisons. CRFX is thresholded at $p < 0.01$, uncorrected level at the voxel level, then at $p < 0.01$, corrected, at the cluster level. Note that these choices are made in order to roughly balance the specificity of the methods, while using them in a standard way.

PRFX maps (see Sec. 3.3) are computed for $K = 500$ parcels. Since the parcel centres are defined at the group level in Talairach space, the voxels in the group result map are assigned to the parcel with the closest center in Talairach space. This results in a piecewise constant map, the pieces resulting from a Voronoi parcellation of the group mask into parcels. Note that in our bootstrap procedure, such boundaries are defined independently in each subgroup of subject. For parcellation, we use the hierarchical procedure presented in [203].

Reproducibility measurements for different cognitive contrasts. We computed the random effects z-variate for different cognitive contrasts, using $R = 5$ groups of $\mathcal{S} = 16$ and a threshold $\theta = 3.1$ corresponding to $p < 0.001$ uncorrected for the contrasts *left-right button press*, *audio instructions-video instructions* and *computation-reading*. The reliability index κ, the proportion of putative true positives λ, and the inter-cluster distance penalty Φ are given in Fig. 3.2. It shows that κ and λ have different behaviours and are strongly dependent on the cognitive contrast under study. For instance, the left motor contrast activates relatively small regions with a relatively low reproducibility; the auditory-selective contrast activates larger regions with high reproducibility; the computation-selective contrast activates larger regions, but with low reproducibility. The inter-cluster distance penalty Φ does not discriminate between the different contrasts as strongly as κ. As could have been expected, it has the opposite behaviour (maximal for the computation contrast, minimal for the auditory contrast).

How the threshold affects the reproducibility of the analysis ? Here we study the behaviour of our reproducibility measures when applied to a thresholded RFX map, when we let the threshold vary. The reproducibility measure is computed for 100 different splits of the population of subjects into $R = 5$ groups of $\mathcal{S} = 16$ subjects, in the case of the *left-right button press* contrast. The threshold (in z-variate scale) varies from $\theta = 2.2$ ($p < 0.015$, uncorrected) to $\theta = 4.0$ ($p < 3.2.10^{-5}$, uncorrected) in steps of 0.2.

As expected, the sensitivity parameter λ decreases when θ increases (see Fig. 3.3(b)). More interestingly, κ reaches a maximum for $\theta^\star \sim 2.7$, but the index remains close at least for $\theta < 3.5$ as can be seen in seen in Fig. 3.3(a). Accordingly, the inter-cluster distance penalty Φ is minimized for a threshold $\theta^\star \sim 3$. The correspondence of these results is interesting, given that these two similarity measures are obtained independently, and based on different considerations. Note that we have obtained similar results when studying the other contrasts with slightly higher (auditory contrast) or lower (computation contrast) threshold values. Thereafter, we retain the threshold $\theta = 3.1$ ($p < 0.001$, uncorrected for multiple comparisons) for random effects z-statistics.

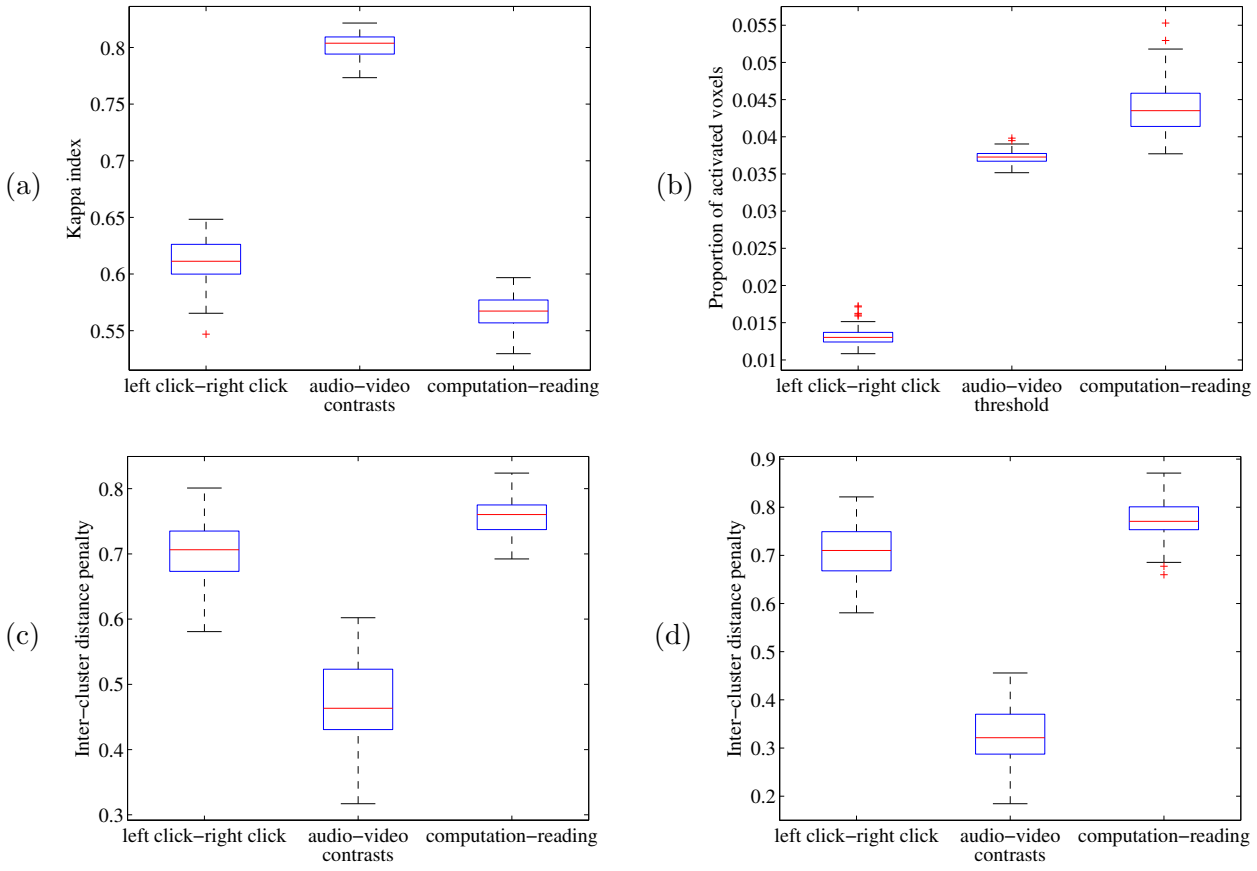


Figure 3.2: Dependence of the reproducibility and of the sensitivity of the random effects analysis on the functional contrast under consideration. These results are obtained by drawing 5 disjoint groups of $\mathcal{S} = 16$ subjects in the population of 81 subjects, and applying the whole analysis procedure. The threshold is $\theta = 3.1$ ($p < 0.001$) (a) Over 100 replications, the reproducibility index is higher for the *audio instructions-video instructions* contrast than for the *left-right button press* and *computation-reading* contrast. (b) However, the size of the putatively activated areas is greater for the contrast that shows regions involved in computation, and smaller for the contrast that shows the regions involved in motor activity. (c-d) The cluster variability penalty Φ is presented for clusters of more than $\eta = 10$ (c) or $\eta = 30$ (d) voxels (the lower the better). The behaviour is as expected, with the smallest value for the auditory-specific contrast.

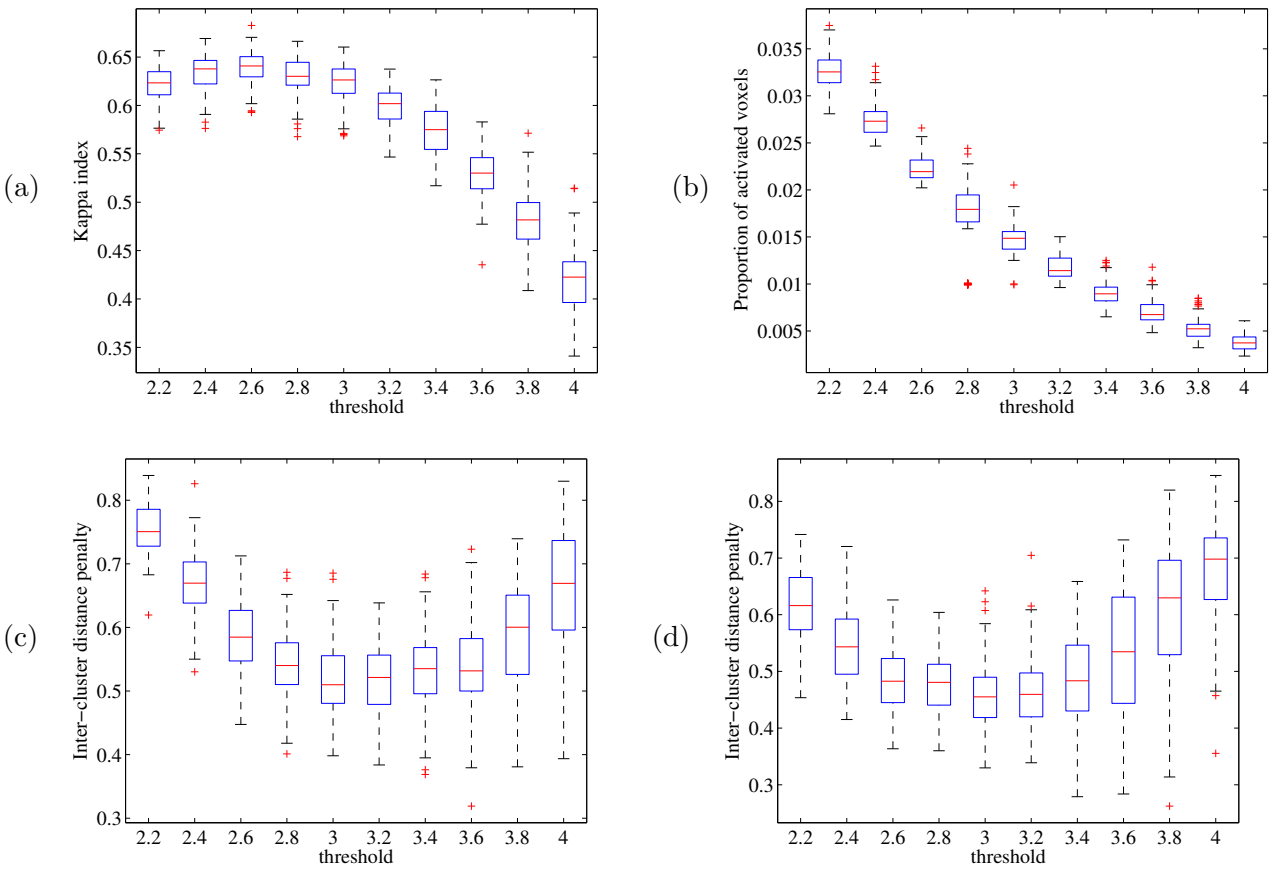


Figure 3.3: Dependence of the reproducibility, the sensitivity, and the distance between supra-threshold clusters of the group random effects analysis on the threshold chosen to binarize the statistic maps. These results are obtained by drawing 5 disjoint groups of $\mathcal{S} = 16$ subjects in the population of 81 subjects, and applying the procedure described in section 3.2.2. This is performed on the images of the *left-right button press* contrast, with 100 resamplings. (a) The reproducibility index κ shows is maximized for $\theta \sim 2.7$. (b) The sensitivity decreases when θ increases. (c,d) The average distance between supra-threshold clusters of more than 10(c) or 30(d) voxels across groups has a minimum around $\theta \sim 3$.

How many subjects are necessary to obtain a reliable group map ? We study the dependence of κ, λ and Φ when we let the size $\mathcal{S}$ of the group vary. We base our investigation on the *left-right button press* contrast, with group maps thresholded at the $\theta = 3.1$ ($p < 0.001$, uncorrected) level. The results are presented in Fig. 3.4. It shows that the reproducibility increases with the group size, which was expected. The sensitivity also increases with the group size. Interestingly, the reproducibility reaches a plateau only for $\mathcal{S} \approx 25$. The inter-cluster distance penalty Φ has a similar behaviour, with a plateau for $\mathcal{S} = 27$ subjects when $\eta = 10$, while lower values are reached when using $\eta = 30$.

Comparison of different group analysis methods. Now we study how the reproducibility index behaves for different statistical methods: The t statistic [RFX, see Eq. (3.4)], the same test after 12 mm smoothing of the data -instead of 5mm- (SRFX), the mixed effects statistic, controlled by permutation [MFX, see Eq. (3.11)], the parcel-based RFX test (PRFX), the t-statistic thresholded at the cluster-level (CRFX), the Wilcoxon test (WKX), and the pseudo-MFX test ΨFX. RFX, SRFX, MFX, WKX, PRFX and ΨFX maps are thresholded at the $p < 0.001$ level, uncorrected for multiple comparisons. The CRFX map is first thresholded at the $p < 0.01$, uncorrected level, then at the $p < 0.01$, corrected, cluster-level. The results are obtained by bootstrapping in $R = 8$ groups of size $\mathcal{S} = 10$. The results are presented in Fig. 3.5.

From the point of view of reproducibility, the WKX and RFX tests have the worst performance overall, while the SRFX performs slightly better. CRFX, PRFX and MFX techniques yield higher reproducibility, but ΨFX yield the highest scores. The results are more variable with PRFX than with other techniques; this reflects the fact that PRFX is based on a smaller number of volume elements, so that statistical tests have a less stable behaviour purely due to fewer number of parcels compared to voxels.

CRFX, MFX, and to a lesser extent, PRFX tests are more sensitive, i.e. have a larger fraction of generally activated voxels, than voxel-based tests. Note however that the specificity control of CRFX matches the other approaches only approximately.

Finally, the average supra-threshold cluster distance Φ is minimal for ΨFX, and relatively low for MFX. It is similar for the other techniques. Before discussing these results, let us first introduce some descriptive statistics to enlighten and complete the analysis of this dataset.

3.2.4 Limitations of the standard Gaussian mixed-effects model

Are the effects Gaussian distributed ? An important question that may arise is whether the effects $\hat{\beta}_s(v)$ are normally distributed or not, since this is a key assumption in standard (random effects) group analysis. We have used the D'Agostino-Pearson test (see Section B.3 and [246] for details), based on the computation of the skewness and the kurtosis (third and forth order cumulants) of the values $\{\hat{\beta}_s(v)\}, s = 1..S$ in each voxel v. This provides the p-value of the D'Agostino-Pearson statistic under the null (normal) hypothesis. For the sake of visualization, we convert the p-value into a z-value. We have then repeated the procedure based on the normalized effects $\{\underline{\beta}_s(v) = \frac{\hat{\beta}_s(v)}{\hat{\sigma}_s(v)}\}, s \in \{1, .., S\}, v \in$

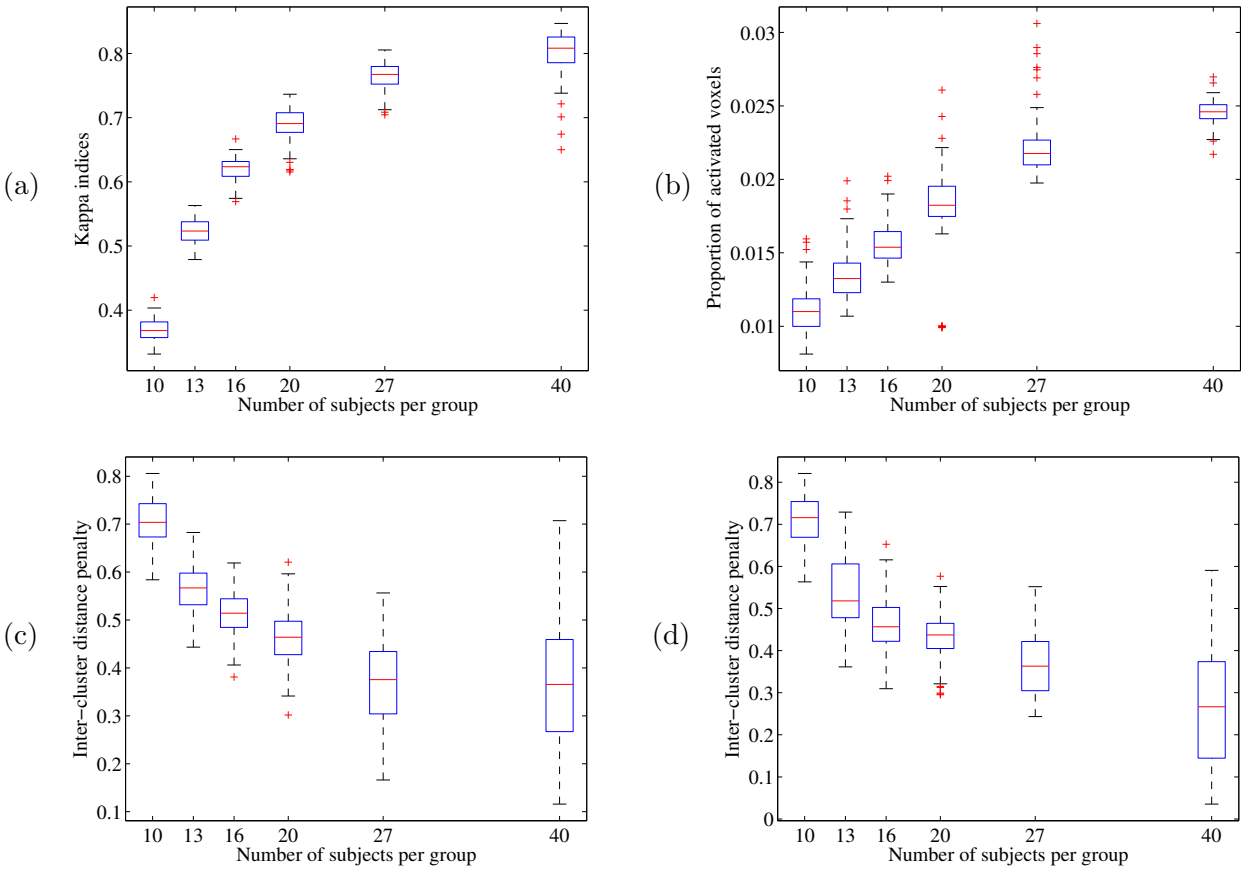


Figure 3.4: Effect of the RFX group size on reproducibility κ (a), sensitivity λ (b) and the average distance between supra-threshold cluster centroids Φ (c-d). The reproducibility is assessed considering disjoint groups of size $\mathcal{S} = 10, 13, 16, 20, 27, 40$ within the population of 81 subjects. This is performed on the images of the *left-right button press* contrast, with 100 resamplings. (a) The reproducibility index increases with $\mathcal{S}$ and reaches a plateau for $\mathcal{S} > 20$. (b) The size of putatively activated areas steadily increases with $\mathcal{S}$. (c-d) The average intra-cluster distance decreases with $\mathcal{S}$; it reaches a plateau for $\mathcal{S} > 20$ when $\eta = 10$ (c), whereas it further decreases when $\eta = 30$ (d).

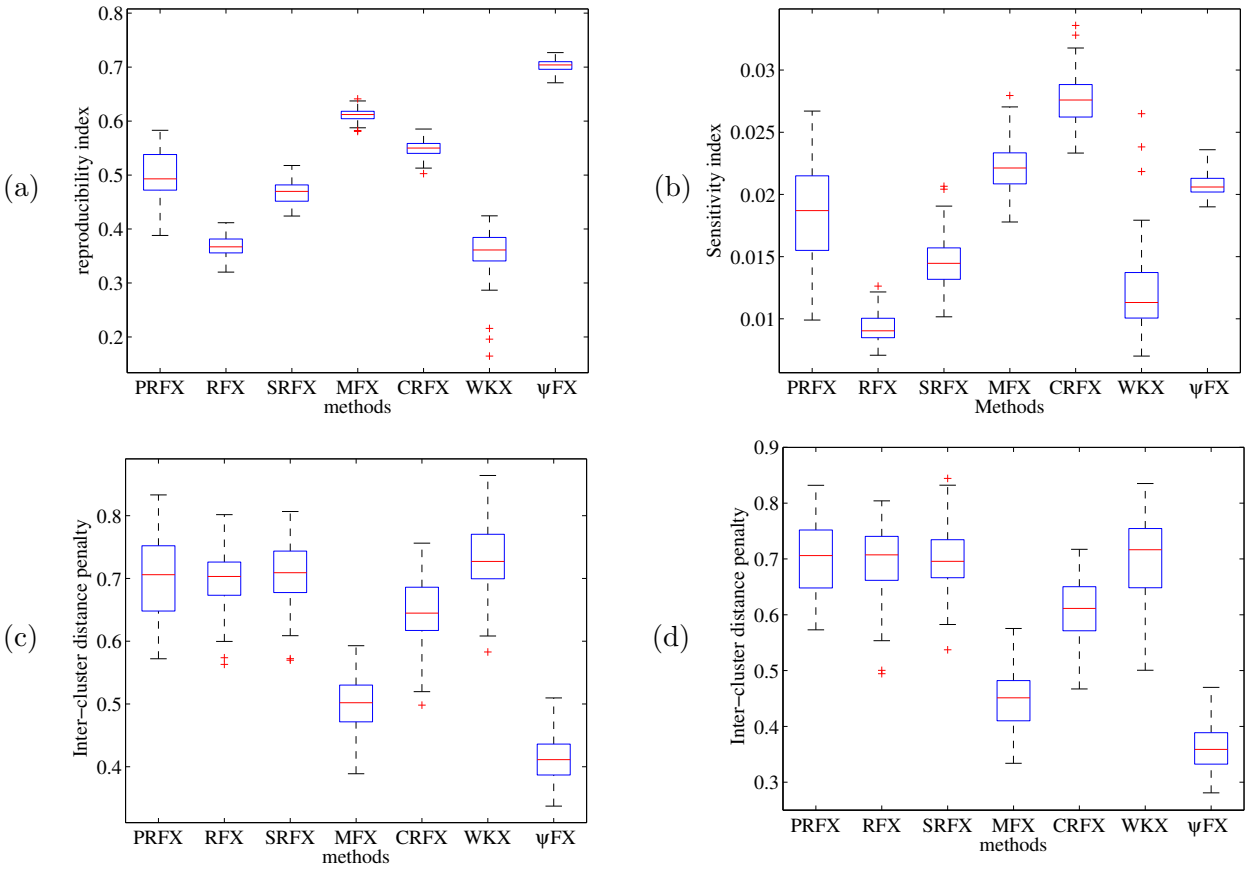


Figure 3.5: Dependence of the reliability κ (a), the sensitivity λ (b), inter-supra-threshold cluster distance penalty Φ (c-d) of the statistical analysis on the group statistic used. Φ is based clusters of size greater than $\eta = 10$ (c) or $\eta = 30$ (d). These quantities assessed considering $R = 8$ disjoint groups of size $\mathcal{S} = 10$ within the population of 81 subjects, using the *left-right button press* contrast, and 100 resamplings.

$\{1, .., N_v\}$ which removes a potential variability in signal scaling across the population. At the group level, the normalization through the residual magnitude has a much greater impact than the deviation from normality on the resulting tests due to the fact that $\hat{\sigma}_s(v)$ is estimated with a finite ($\nu = 100$) number of degrees of freedom.

Analysis of the second-level variance We performed the D'Agostino-Pearson test on the effects $\hat{\beta}(v)$ of all the voxels, as well as the normalized effects $\underline{\beta}(v)$, which yields two maps for each contrast. We present them for *left-right button press*, *audio instructions-video instructions* and *computation-reading*, thresholded at the $p < 0.001$ uncorrected level. We also present the inter-subject variance maps $v_g(v)$ computed in a mixed-effect model (see Sec. 3.1.3). We present these maps together with the RFX map (converted to a z-variate) based on 81 subjects in Figs. 3.6-3.8. Note that other contrasts, e.g. *horizontal-vertical checkerboards*, *sentence reading-low-level vision*, *cognitive trials-motor trials*, and the opposite ones, not presented here due to space limitations, yield qualitatively similar results.

In each case, the regions with highest group variance are also found in the regions with highest random effects statistics in absolute values; some of them are absent in the maps 3.6-3.8, where signed statistics are presented.

Inspection of these maps suggests that

- Areas of high variance tend to co-localize with the activated areas. This implies that the parameters $v_g(v)$ and $\bar{\beta}(v)$ are not independent and that statistics that are penalized by the group variance may not be very efficient in general.
- Non-normality is very significant in wide regions of the brain: deviation from normality of $\hat{\beta}$ across subjects concerns 22% of the brain voxels at ($p < 0.001$, uncorrected) for the *computation-reading* contrast, 27% for the *left-right button press* contrast and 30 % for the *audio instructions-video instructions* contrast.
- Deviation from the normality hypothesis is much lower for the normalized effects $\underline{\beta} = \frac{\hat{\beta}}{\hat{\sigma}}$ than for the raw effects $\hat{\beta}$. For instance, the rate of voxels with normality rejected at ($p < 0.001$,uncorrected) drops from 22% to 9.2 % for the *computation-reading* contrast, from 27% to 2.9% for the *left-right button press* contrast and from 30 % to 10% for the *audio instructions-video instructions* contrast. This means that dimensionless first-level statistics yield more homogeneous quantities across subjects than effects expressed in percents of baseline signal increase.
- Deviation from normality of the effects does not specifically co-localize with activated areas, but, in several cases it coincides with the boundaries of activated areas.

From Figures 3.6-3.8, one of the most striking effects is the co-localization of high second-level variance areas with large random effects areas. Such an effect is not expected since the RFX is defined as the quotient of the estimated mean effect by the standard deviation of this estimate.

The interpretation could be that 1) the contrast-to-noise ratio (CNR) of the BOLD effect is highly variable across subjects, and by definition this effect does not appear in non-activated areas and/or 2)

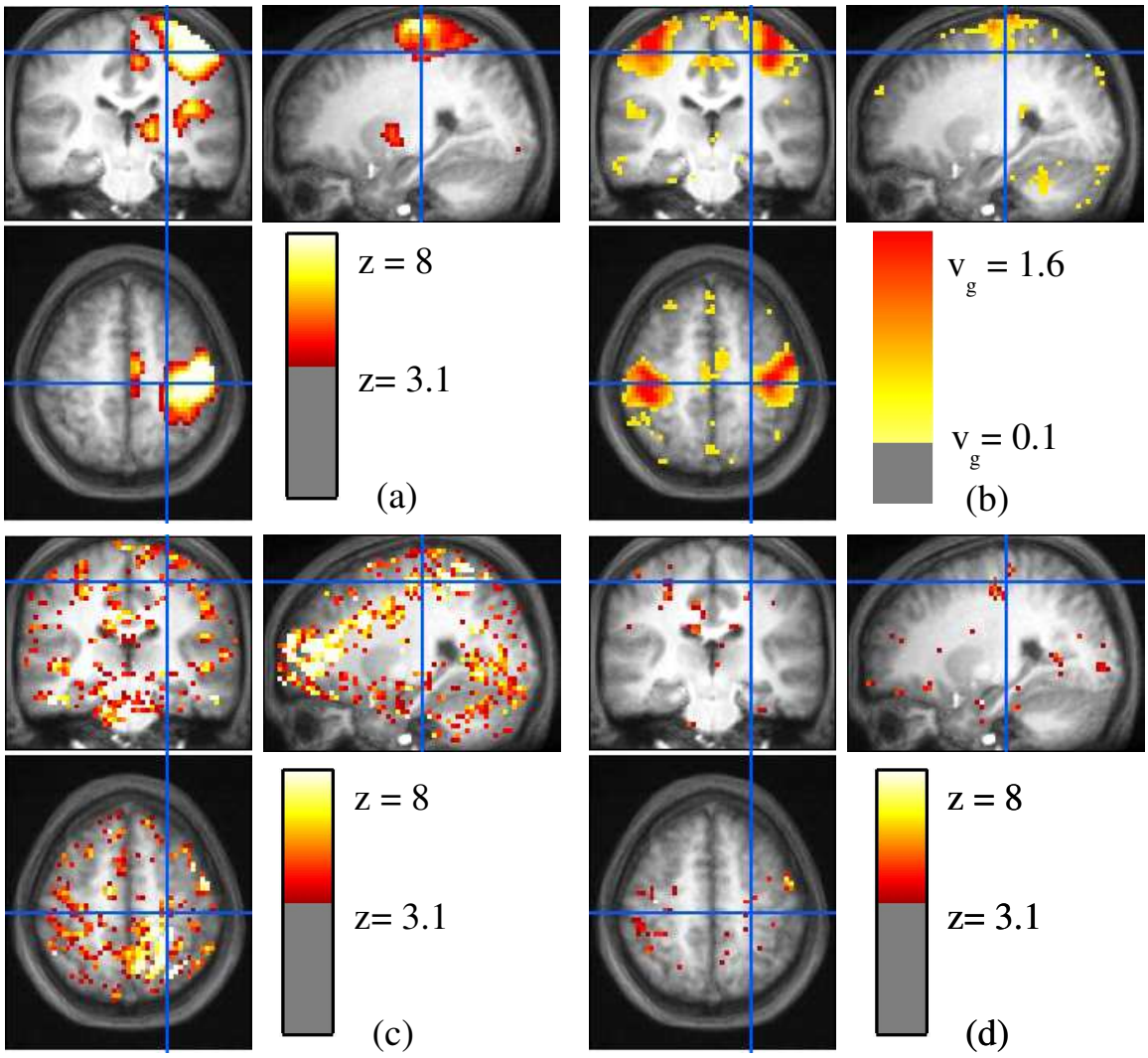


Figure 3.6: Statistical model of the effects for the *left-right button press* contrast, on $S = 81$ subjects. (a) z-value associated with the RFX test; (b) group variance estimate; (c) z-value of the D'Agostino-Pearson test for normality of the effects $\hat{\beta}$; (d) z-value of the D'Agostino-Pearson test applied to the normalized effects $\underline{\beta}$. Note that all the z values are limited to the $[-8, 8]$ range. The color scale of the variance image has been chosen arbitrarily in order to have supra-threshold areas that are comparable with the other maps. The variance is expressed in squared percentage of the BOLD mean signal. Cross position: (-23,-28,56)mm in the MNI space.

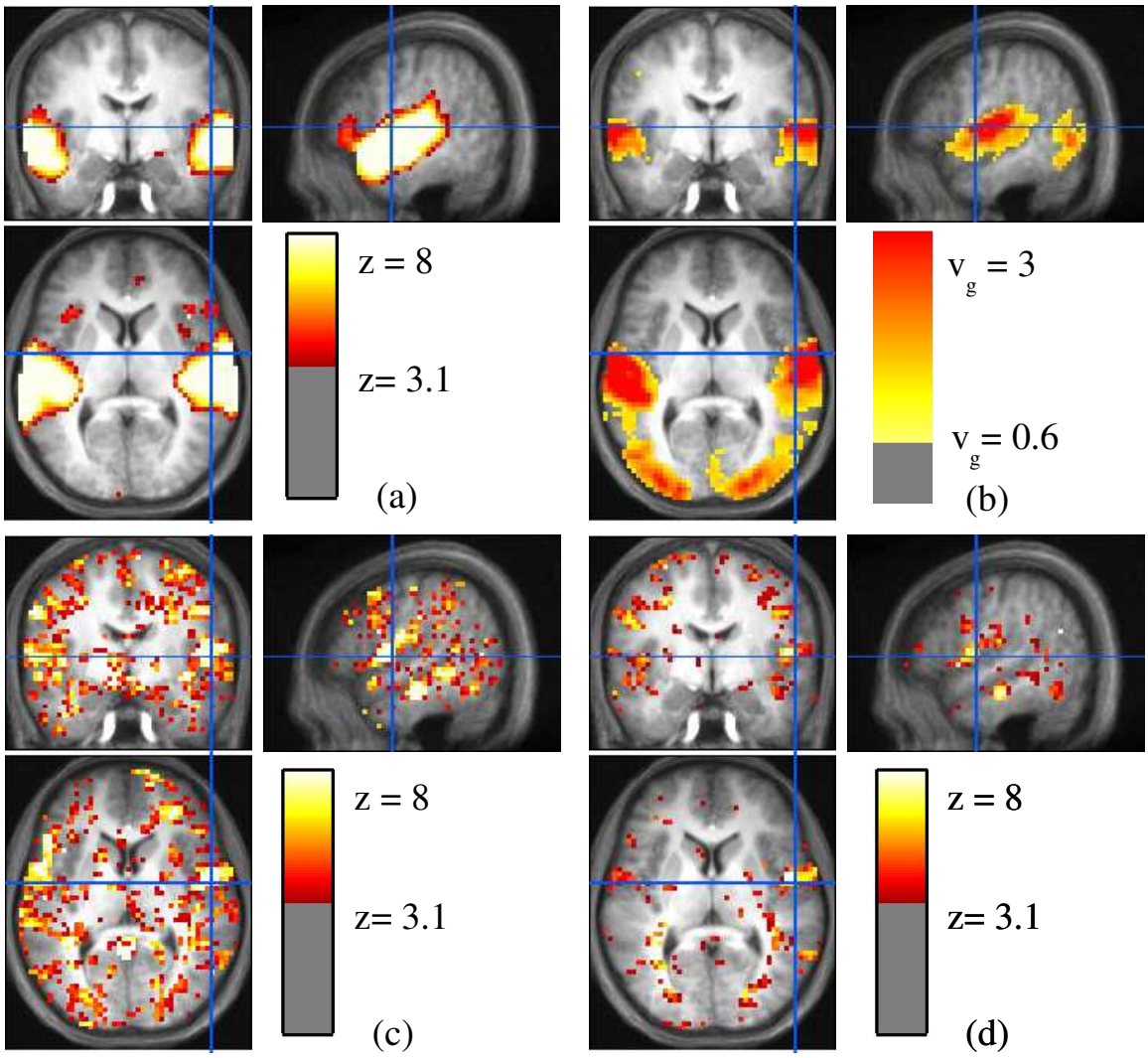


Figure 3.7: Statistical model of the effects for the *audio instructions-video instructions* contrast, on $S = 81$ subjects. (a) z-value associated withe the RFX test; (b) group variance estimate; (c) z-value of the D'Agostino-Pearson test for normality of the effects $\hat{\beta}$; (d) z-value of the D'Agostino-Pearson test applied to the normalized effects $\underline{\beta}$. Note that all the z values are limited to the $[-8, 8]$ range. The color scale of the variance image has been chosen arbitrarily in order to have supra-threshold areas that are comparable with the other maps. The variance is expressed in squared percentage of the BOLD mean signal. Cross position: (-54,-6,8)mm in the MNI space.

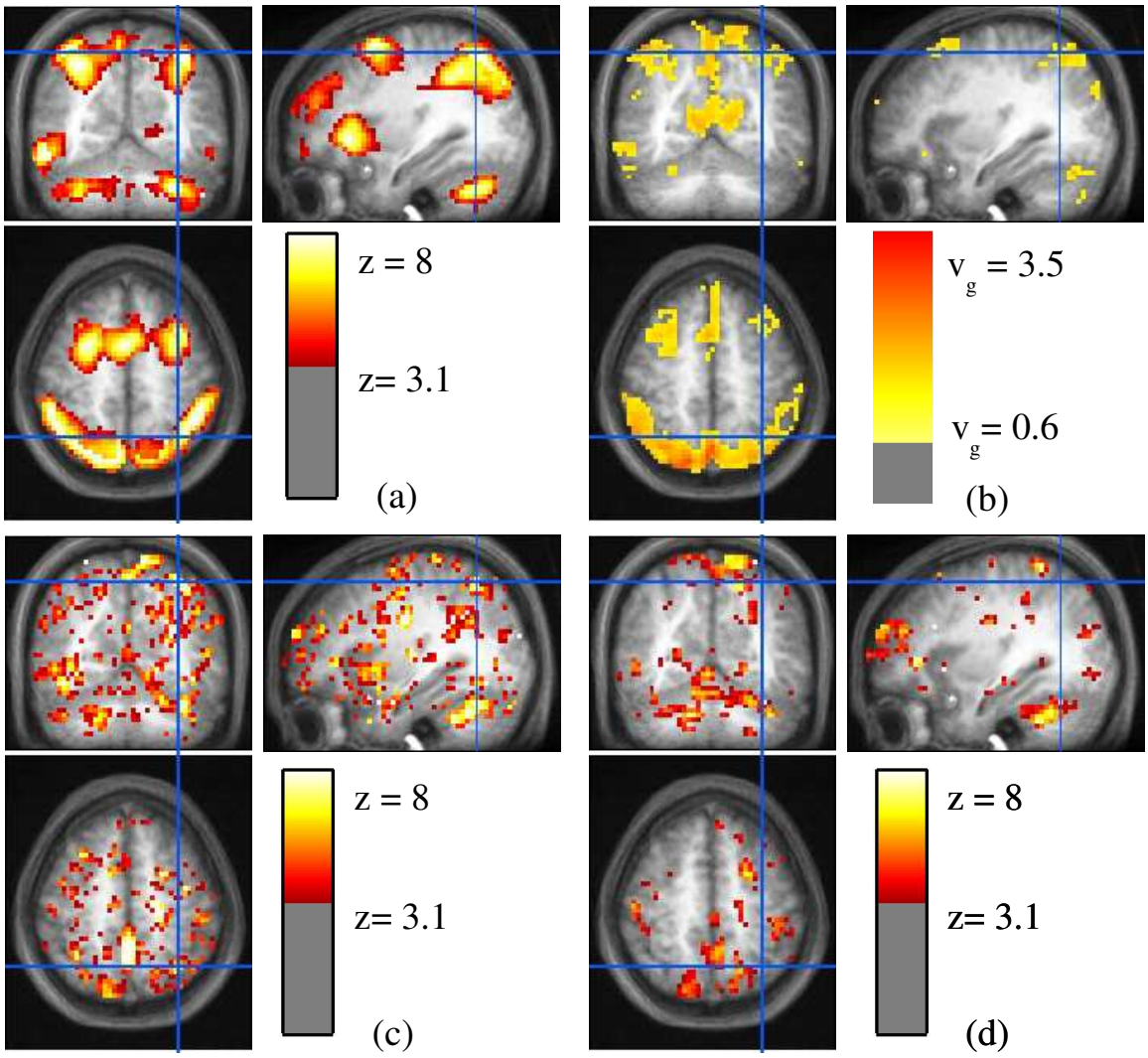


Figure 3.8: Statistical model of the effects for the *computation-reading* contrast, on $S = 81$ subjects. (a) z-value associated withe the RFX test; (b) group variance estimate; (c) z-value of the D'Agostino-Pearson test for normality of the effects $\hat{\beta}$; (d) z-value of the D'Agostino-Pearson test applied to the normalized effects $\underline{\beta}$. Note that all the z values are limited to the $[-8, 8]$ range. The color scale of the variance image has been chosen arbitrarily in order to have supra-threshold areas that are comparable with the other maps. The variance is expressed in squared percentage of the BOLD mean signal. Cross position: (-33,-60,56)mm in the MNI space.

spatial mis-registration[2] implies that at a given voxel, i.e. a given position in MNI space, some subjects have activity while other subjects do not, thus spatially widening the signal distribution. For simple contrasts such as those used (left or right button press, sentence listening), different cognitive strategies should be ruled out.

This inflated variance effect certainly deserves more investigation, given its prominent effect on statistics (sensitivity and reproducibility): for instance, the ΨFX statistic -that does not take into account the group variance, hence is simply a weighted average of the subject-based effects- seems more reliable than the MFX statistic, which is itself much more reliable than the RFX statistic (see Fig. 3.5). The effect of group variance is also an argument in favour of Bayesian analysis of fMRI data, if the reference signal level is not 0 [69].

Non-normality is another important factor. To our knowledge, this has not been investigated before, since it requires a high number of subjects. Interestingly, the importance of non-normality is reduced when considering normalized effects $\underline{\beta}_s(v)$ instead of raw effects $\hat{\beta}_s(v)$. This shows that information from the first-level analysis that are used in group statistics have to be chosen carefully. In particular, the difference observed between the normality of $\hat{\beta}$ and $\underline{\beta}$ maps possibly indicates that the current way of normalizing signal magnitude with respect to the mean signal may not be optimal for inter-subject comparison (this is also an open question for inter-session variability). However, the normalization with respect to first-level variance might not be satisfactory, since it could in turn be highly dependent on acquisition artifacts, motion and physiology, whether these are modeled of not. We are not aware of any successful signal calibration strategy, but mixed-effects model may solve part of the problem. Interestingly, several areas with significant non-normality are found at the periphery of activation maxima, confirming the impact of spatial shifts on group statistics. Once again, further investigations on non-normality may be performed, e.g. searching different groups of subjects in the population or outlier subjects (see [104]). Robust statistics might also be used for inference [233], but at the risk of a weaker control on specificity. Moreover, such inference schemes raise the difficult question of the generalizability of group results to other groups of subjects (given that the concept of *outlier* is ill-defined when considering a small group). In general, it is advisable to use non-parametric assessment to obtain reliable thresholds [140]. However, the choice of robust statistics (statistics that adapt to non-normal data) is not necessarily advantageous: for instance, the Wilcoxon statistic did not perform better than other statistics in our experiments (see Fig. 3.5).

Practical conclusion : reliability of the different statistical tests One of the most important practical questions is to describe or design the most efficient ways to perform group studies in neuroimaging. Based on this first study we can suggest some guidelines.

First of all, given the results on normality tests, non-parametric assessment of functional activity should be preferred to analytical tests, which may rely on incorrect hypotheses. This can be done using adapted toolboxes e.g. SnPM [88] or Distance [140]). It is worthwhile to note that C implementation of the tests reduces computation time to a reasonable level (e.g. cluster-level p-values can be computed in

[2]Spatial mis-registration may be artefactual (incorrect normalization) or not (intrinsically different functional anatomy.)

less than one minute on a ten subjects dataset). Non-parametric estimation of the significance improves both the study sensitivity and reproducibility.

Second, mixed-effects models should systematically be preferred to mere random effects analyses: there is some information in the first level of the data that improves the estimation of the group effects/variance and statistic.

Third, cluster- and parcel-based inference should be preferred to voxel-based thresholding. Cluster-level inference is of frequent use, which benefits the sensitivity and the reliability of group analyses. However, it is based on the assumption that activated regions are large, which is not necessarily true. Parcel-based inference may thus be a interesting alternative, since it further allows some spatial relaxation in the subject-to-subject correspondence. The price to pay is a larger variability of the results due to a less stable decision function (activated vs non activated). We recommend the combination of one of these techniques together with MFX. By contrast, stronger smoothing (12mm FWHM instead of 5mm) did not increase significantly the results reliability.

Fourth, to our surprise, ΨFX was found to be the most reliable technique. Although the statistic function does not take into account the group variance - as argued earlier, this is probably the reason for its higher performance - its distribution under the null hypothesis is tabulated by random flips of the effects signs, so that it is indeed a valid group inference technique. However, it should be used with care because first the thresholds have to be computed voxel per voxel, (i.e. are not spatially stationary), and second the statistic value itself has no obvious interpretation, in contrast to the RFX and MFX statistics.

3.3 Spatially relaxed inference: parcel-based RFX

3.3.1 Parcellation for group inference: definition and constraints

Definition of the problem As discussed in Sec. 3.2.1, the main limit of standard inference procedures is their weak sensitivity. This lack of sensitivity is, in turn largely related to the choice of mass-univariate analysis, which entails a severe multiple comparison problem. As noticed in Section 3.1.5, cluster-level inference is in general much more sensitive than voxel-level inference. Moreover, given the shortcomings of spatial normalization discussed in section 2.5.1, there is little interest in inferring the presence of activation at the voxel level: a regional or super-voxel level of inference is clearly more adapted.

Recently, [60] proposed to group voxels into anatomically and functionally homogeneous parcels across subjects. This latter approach is appealing for several reasons: *i)* the voxels, which are the spatial units of functional images, define an arbitrary spatial resolution; small regions modelled as groups of voxels may better reflect true regions of activity, as they can be characterized in fMRI; *ii)* this makes the analysis less sensitive to -artefactual or intrinsic- mis-registrations in multi-subject studies; *iii)* technically, this under-sampling of the brain volume or surface reduces the well-known problem of multiple comparisons, allowing for less conservative Bonferroni corrections; *iv)* in general, it may also better take into account the true between subjects anatomical variability such as cyto-architectonic [176] or sulco-gyral [173] structure of the cortex. The work proposed here follows this parcellation idea.

Previous work related to parcellation Although the idea of constructing automatically anatomo-functional parcels is relatively recent in neuroimaging, some works are clearly related to this line of thought and we briefly review some of those attempts.

First, brain parcellations have been proposed in a purely anatomical context. Existing methods are based either on prior knowledge of anatomy and connectivity [134], on sulcal geometry [129, 216, 122, 197], on sulcal identification [17], or on probabilistic atlases [178, 57]. We note that these parcellations can be used in a functional context, after co-registration of anatomical and functional images. This may be done semi-automatically [149], or using atlases [226]. Note that this procedure might suffer from poor anatomo-functional correspondence related to EPI distortions; more generally the functional homogeneity of the resulting parcels should be checked for inference purposes. We come back to this question in Chapter 5.

By contrast, parcellation based on functional information is a relatively novel approach, with specific challenges. Clustering of similar time series yields homogeneous functional regions [187], but purely functional clusters are well-defined only locally [81], and specific methods have to be designed for the parcellation of the entire brain. For instance, [161] have proposed an EM algorithm to jointly model the spatial location of activation, together with the activation amplitude at each cluster. This joint modeling is reminiscent of [59]. However, this model is rather adapted to encode sparse activation patterns, while we aim at a parcellation of the entire brain volume or large brain regions, consistently across subjects.

Last, in [58, 59, 60] Flandin et al. launched fMRI-tailored parcellation approaches; they have proposed a clustering approach based on Gaussian Mixture Models (GMMs) that groups voxels from multiple datasets according to a spatio-functional criterion, blindly to the subject. This is a very effective approach, since it is based on fast algorithms. In this framework, the anatomy of the subject can be adequately introduced in the core algorithms and the technique naturally yields results such that inter subject analysis assumptions are naturally enforced. To our knowledge, this is a first approach to deal with anatomical and functional variability for multi-subject analysis. However, it has a few limitations. First, it does not guarantee the spatial connectivity of the parcels. Secondly, it does not necessarily produce multi-subject parcels (or cliques) in which each subject of the group is present, which should be a desirable feature[3]. This may or may not be a limitation depending on whether subjects are functionally homogeneous or not, but some application may require that corresponding parcels can be found in all or most of the subjects. Last, the definition of a spatial/functional criterion may be cumbersome, since it contains non-homogeneous terms and the relative weighting of the functional and the spatial information is still an open question.

Particular constraints when using parcellations for group-level inference Thereafter, it is understood that a parcellation can be performed on volumetric data (images), or surface-based data (meshes); it is only required that a coordinate system defines the position of any object in either case. Let us briefly describe the desired parcel characteristics:

Let $(\beta_s)_{s=1..S}$ be a set of fMRI data consisting in one or several contrast maps sampled from S different subjects (note that each β_s can be multi-dimensional). Basically, the information available on each site

[3]This may be called the *onto property* of the parcellation.

(voxel in 3D space or vertex of a mesh) consists of a position information τ and a feature information β. A parcellation of the dataset is a subdivision of each individual dataset into K regions which is consistent across subjects. Let $(\lambda_s)_{s=1..S}$ be the associated labelling functions -on in each subject [4]. Consistency refers to the following criteria

1. *Onto property*: Each subject should be present in each individual data
2. *Intra-subject connectivity:* Each parcel in each subject should be spatially connected.
3. *Intra subject homogeneity:* Each parcel in each subject should represent a functionally homogeneous region, which means that within parcel variance should be as small as possible in each subject.
4. *Inter-subject spatial coherence:* the relative position of parcels should be the same across individuals. This means that the spatial transformation implicit in the parcellation should be regular in some sense (e.g. diffeomorphism), i.e. a smooth function of space.
5. *Inter-subject homogeneity:* the parcels should match across subjects concerning some feature information (position, functional activity, possibly anatomical or functional connectivity)

Before proposing a solution, it is important to notice that the problem is relatively easily solved in the intra-subject case with an adapted clustering algorithm. In particular, conditions 4 and 5 disappear, while condition 1) is almost naturally enforced. Thus the problem boils down to clustering the 3D set in order to minimize the within-parcel variance, while enforcing the spatial connectivity of the parcels. We present in Fig. 3.9 a solution based on a modification of Ward's algorithm that includes spatial connectivity constraints.

3.3.2 Solution: parcellation algorithm

Unfortunately, solving the problem in the intra subject case does not help much for solving it in the inter-subject case: While it is a priori possible to define parcels independently in each dataset and then to match them in a second step, the matching complex is at the same time complex and ill-posed, given that there may exist no natural one-to-one mapping between the individual structures.

A more adapted solution consists in building a population-level model of the parcellation, which reflects the average of individual information, and defines the topological constraints and the spatial and functional information related to the parcellation. This model is called *template*, and its counterpart in each subject will be called *prototype*. This template/prototype/data hierarchy can be loosely interpreted as a generative model, as shown in figure 3.10(left). The functional information/position of the template, prototypes, and voxels will be denoted $\bar{\beta}(k)/\bar{\tau}(k)$ for $k \in 1,..,K$, $\tilde{\beta}_s(k)/\tilde{\tau}_s(k)$ for $k \in 1,..,K, s \in 1..S$ and $\beta_s(v)/\tau_s(v)$ for $v \in 1,..,N_v, s \in 1..S$. Recall that the assignment from voxels to parcels in subject s is denoted λ_s.

We can now rephrase the problem Find $(\lambda_s)_{s\in1,..,S}$ such that

[4] Note that this is a hard labelling: each site belongs to a single parcel

Input : Neighborhood relationship $\sim$ between the N_v sites
F: feature matrix of shape (N_v, dim)
K: number of desired clusters

Initialize all the clusters
for $k = 1 : N - K$
find (i, j) such that $(i, j) = \text{argmin}_{i' \sim j'} var_{q \in i' \cup j'}(F(q))$
assign i and j to cluster $(k + N)$
let $F[k + N] = F[i] \cup F[j]$
endow $(k + N)$ with the neighbors of i and j
delete i and j
end for
Output: the K remaining components $k \in \{1, 2N - K\}$

Figure 3.9: Pseudo-code of an algorithm that performs a variance-minimizing and topologically correct parcellation of an individual dataset. This can be seen as a variant of Ward's algorithm.

1. $\forall s \in [1, .., S], \lambda_s ([1, .., N_v] \rightarrow [1, .., K])$ is onto,
2. $\forall k \in [1, .., K], (\lambda_s)^{-1}(k)$ is connected in 3D,
3. $\forall s \in [1, .., S], \sum_{k=1}^{K} var(\beta_{(\lambda_s)^{-1}(k)})$ is minimized, i.e. $J_{intra}(s) = \sum_{v=1}^{N_v} \|\tilde{\beta}_s(\lambda_s(v)) - \beta_s(v)\|^2$ is minimized, assuming that $\tilde{\beta}_s(k) = mean_{\lambda_s(v)=k} \beta_s(v)$.
4. $\forall (s, \sigma) \in [1, .., S]^2$, $(\tilde{\tau}_s) \rightarrow (\tilde{\tau}_\sigma)$ is smooth. Here we weaken this condition to: $\forall s \in [1, .., S]$, $(\bar{\tau}) \rightarrow (\tilde{\tau}_s)$ is smooth, in a sense which is made precise below.
5. $\sum_{k=1}^{K} var_s(\tilde{\beta}_s(k))$ is minimized, i.e. $J_{inter} = \sum_{k=1}^{K} \|\tilde{\beta}_s(k) - \bar{\beta}(k)\|^2$ is minimized, assuming that $\bar{\beta}(k) = mean_{s \in 1,..,S} \tilde{\beta}_s(k)$

The estimation model essentially minimizes the criteria 3) and 5), while criteria 1) and 2) are taken as hard constraints. Note that, assuming that the number of voxels per parcel is constant across subjects, by Huygens formula, minimizing $J = J_{inter} + \sum_{s=1}^{S} J_{intra}(s)$ simply amounts to minimizing

$$J_T = J_{inter} + \sum_{s=1}^{S} J_{intra}(s) = \sum_{s=1}^{S} \sum_{v=1}^{N_v} \|\beta_s(v) - \bar{\beta}(\lambda_s(v))\|^2, \tag{3.18}$$

meaning that our procedure simply minimized the global inertia of the data clustering some hard constraints. Criterion 4) is handled in the following way: the group-level coordinates $(\bar{\tau}(k)_{k=1..K})$ are used to build a topological model of the parcellation which is simply the Delaunay triangulation of the ambient

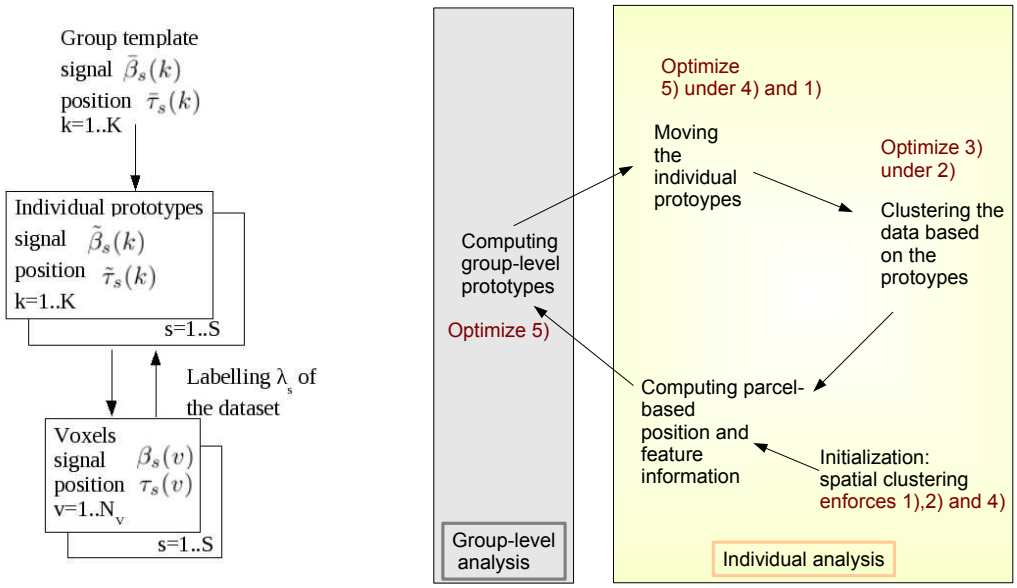


Figure 3.10: (left) Graphical interpretation of the parcellation model the descending arrows represent the *generative model*. (right) Scheme of the solution for the estimation of parcels for group inference.

space. The resulting Delaunay graph defines the neighboring system on the template. $\forall s \in [1, .., S]$, the Jacobian of the mappings $(\bar{\tau}) \rightarrow (\tilde{\tau}_s)$ w.r.t. the graph topology can then be computed and only assignments that have a positive Jacobian $\forall k \in [1, .., K]$ are accepted.

In practice, the model is estimated in a step-wise manner, as illustrated in Fig. 3.10(right). Assuming that an initial labellings λ_s^0 are available, and that these labellings enforce conditions 1),2) and 4), the following steps are alternated:

- Computation of parcel-based position $\tilde{\tau}$ in the anatomical space and feature-based information $\tilde{\beta}$, for each parcel $k \in 1, .., K$ and any subject $s \in 1, .., S$.

$$\tilde{\beta}_s(k) = mean_{\lambda_s(v)=k}\ \beta_s(v) \quad (3.19)$$
$$\tilde{\tau}_s(k) = mean_{\lambda_s(v)=k}\ \tau_s(v) \quad (3.20)$$

  Note that these average values make sense only if the parcels are compact enough in both cases, which corresponds to the criterion 3) above, and requires that the number K of parcels is not too large.

- Computing the group-level prototypes or template, which means that population-average values

are computed for both the position and functional information.

$$\bar{\beta}(k) = mean_{s\in 1,..,S}\ \tilde{\beta}_s(k) \quad (3.21)$$
$$\bar{\tau}(k) = mean_{s\in 1,..,S}\ \tilde{\tau}_s(k) \quad (3.22)$$

- In each subject $s \in 1,..,S$, recompute prototypes $(\tilde{\beta}, \tilde{\tau})$ that minimize criterion 5) given the constraints of 4). Here it is understood that $\forall k \in 1,..,K$, the searched prototype is a particular voxel v in subject s that minimizes $j(v) = \|\beta_s(v) - \bar{\beta}(k)\|^2$ under the constraint that the mappings $(\bar{\tau}) \rightarrow (\tilde{\tau}_s)$ have a positive Jacobian at each node of the template graph. In practice, the voxel v is thus the minimizer of j within a domain which is the intersection of the half spaces supported by its neighbours.

- The following step simply consists in computing the mapping λ^s in order to minimize $\|\tilde{\beta}_s(\lambda_s(v)) - \beta_s(v)\|^2$. In a spirit similar to the intra-subject case, a greedy procedure is used to group each voxel to the nearest prototype in functional space, but where assignments are allowed only when connectivity is ensured. This is done using a greedy procedure similar to Dijkstra's algorithm.

Convergence The convergence of the algorithm is checked by computing the criterion J. In practice, convergence proceeds in 5 steps.

Initialization The initialization is performed by drawing random positions $\bar{\tau}(k), k = 1,..,K$ in the common space and then by deriving $(\tilde{\tau}_s(k)), k = 1,..,K, s = 1,..,S$ and the initial assignments λ_s^0 without taking into account the functional information. This ensures that the basic constraints of the algorithm are met, while the criterion J is clearly not optimized. The final value of J, and a fortiori the assignments λ_s clearly depend on this initial choice, but the fluctuations are usually relatively low compared to the difference between the original and the final value of J (around 10%).

Spatial regularization. Running this algorithm (both in the intra- and inter-subject case) usually provides extremely non-regular shapes for the resulting clusters - although the connectivity constraints are met. In practice, it is extremely useful to include spatial regularization in the procedure. A simple way to do that it to replace the features $\beta_s(v)$ by

$$\underline{\beta}_s(v) = \left[\frac{1}{\beta_0}\beta_s(v); \frac{\mu}{\tau_0}\tau(v)\right], \quad (3.23)$$

where β_0 and τ_0 are arbitrary quantities -typically the standard deviation of the input data- which are used to make the coordinates dimensionless, and μ is a constant that weights the two relative information. This parameter can be set e.g. by qualitatively assessing the parcellation in an intra-subject case.

3.3.3 The Parcel-based random effects (PRFX) procedure

Once the parcels, i.e. the assignment functions (λ_s) are computed, it is about trivial to compute parcel-based average of information of interest, e.g. activation values. The parcels are then simply handled as super-voxels in inference procedures. For simplicity, we concentrate on random-effects procedures only, given that a mixed-effects procedures entails several levels of variance: the voxel-level first-level variance $\sigma_s(v)$ in Eq. (3.1), the variance of the mean estimator (which can easily be taken into account by bootstrap), and the inter-subject variance. This point remains an interesting matter for future work.

We thus consider only the last variance term and compute the t-statistic associated with this model. The nice feature with parcellation is that the multiple comparison problem is alleviated by the replacement of N_v voxels with a much smaller number K of parcels. Assuming that the parcellation model is reasonable, this gain may not be lost by a reduction of spatial accuracy of the model.

Once again, the safest procedure consists in using a non-parametric assessment, i.e. recomputing the statistic for sign-flipped datasets, in order to get the distribution of the maximum value of the brain volume under the null hypothesis. Classical inference proceeds as usual. Still their remains the question of *what* is indeed active when the null hypothesis is rejected. According to the model, the *parcel* is active, but clearly, this corresponds to different voxels according to the subject under consideration. Our approach consists in clustering the common space with a Voronoi parcellation, based only on the spatial coordinates, i.e. assigning each site to the nearest template $\bar{\tau}_k$. This set is used for representation of the active region.

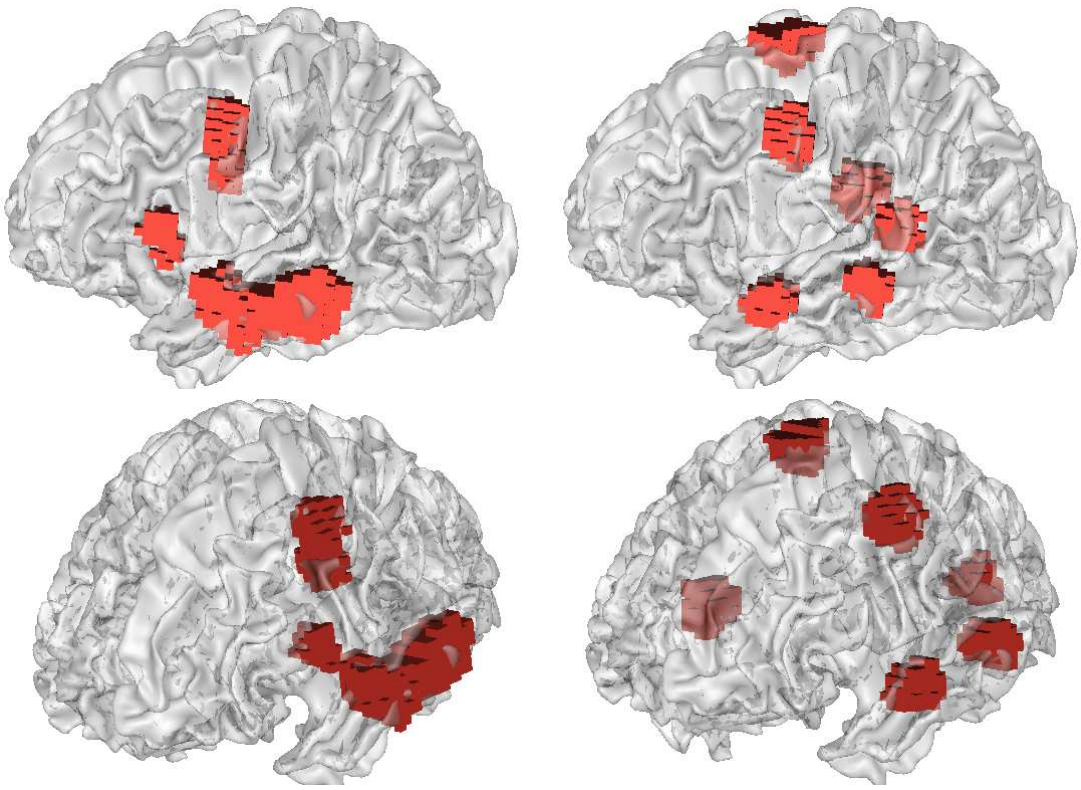
Caveat. Nothing prevents a user of this procedure to use the same data for the estimation of the parcels and the computation of random-effects. This is in fact a reasonable way to use the model. In this case, the parcellation procedure itself needs to be considered as part of the statistic, and thus included in the resampling loop when non-parametric techniques are used to assess their significance. Moreover, choosing a corrected statistic, i.e. controlling the map-level rate of false detections, is necessary to obtain a meaningful calibration of the statistic.

3.3.4 Results and discussion

We have used this method on a excerpt of the Localizer dataset that comprises 14 subjects. The contrast under study shows functional activity in a reading task (as opposed to viewing a checkerboard).

The voxel-based random effects map, thresholded at the cluster level $p < 0.5$ corrected, $z > 4.80$, yields few voxels around two foci in the left hemisphere, one in the prefrontal cortex $z = 5.18 (-48, -3, 42) mm$ and one in the posterior part of the temporal cortex $z = 4.99, (-63, -36, -3) mm$.

We compare this with the parcel-based random effects map obtained from the procedure described above (where learning and test are performed on the same data) with 400 resamplings, and with a trivial parcellation which is simply a spatial partition unrelated to any functional information. The threshold is $p < 0.05$, corrected, which corresponds to z values of $z > 4.02$ and $z > 3.72$ respectively. However, the informed parcellation is more sensitive than the other one, as can be seen in figure 3.11.



(a) (b)

Figure 3.11: Comparison of a *naive* parcel-based RFX map (a) with the proposed one (b), both thresholded at ($p < 0.05$), corrected. The supra-threshold parcels have been meshed and are represented on a typical grey-white matter interface. The solution (a) finds activity along the left temporal lobe, in left Broca's are and in the left precentral gyrus. The solution (b) outlines a much dispersed network, with three distinct foci along the temporal lobe, the precentral gyrus and the Supplementary Motor Area (SMA). It also finds one significantly active region in the right temporal lobe, but misses left Broca's area.

Essentially, both parcellations provide significantly active parcels in the loci found by the voxel-level test. But both find a richer network. Part of this gain in sensitivity in clearly explained by the reduction of the multiple comparison problem associated with the resolution decrease. Using the functional data itself to drive the parcellation seems to segregate the active set of regions into a finer grain pattern (see the temporal lobe in the left hemisphere), and to be effective in outlining isolated regions with significant activity (in the SMA and the right temporal region). On the other hand, this procedure does not detect Broca's area.

Conclusion

Parcellation is an attractive alternative to voxel-based methods for group inference, since it has the important property of reducing the multiple comparison problem, and is thus relatively sensitive with respect to traditional procedures. However, the clustering procedure developed here requires the choice of several parameters and might be difficult to calibrate. Moreover, it leaves unanswered the important question: do parcel represent intrinsic structures of the cortical surface, are are they simply a convenient way to change the image resolution ? To address this questions, we will develop a generative parcel model in Chapter 5.

Chapter 4

Building a functional brain template (a): the Functional Landmark approach

We present here our contribution of automated definition of reproducible regions across subjects, or functional landmarks. This is a summary of a series of contribution on the definition of brain functional landmarks [205, 207, 206, 210] and the comparison with anatomical landmarks [225].

4.1 Introduction

The purpose of group analysis is to yield the regions that are activated in general, or in average, in a certain population of subject. This problem is thus not different from the problem of building a functional brain template. By this we mean, that a template represents a prototype which individual instances are similar to, although the correspondence is not trivial (i.e. it is more complex than the random effects model presented in Chapter 3). This idea seems quite natural, but has received little attention in the neuroimaging community. In this chapter, we present several formalizations of this concept and some validation experiments.

Motivations As argued in section 3.3, performing group inference at the voxel level is a potentially suboptimal approach:

- Voxels are arbitrary units of the brain volume, which do not correspond perfectly after normalization (coregistration of the data to a common template); the remaining variability in many cortical regions can be estimated as 1cm in many cortical regions (see e.g. [91, 193]).

- Voxel-based analysis is plagued with a multiple comparison problem (the brain comprises about 6.10^4 voxels at 3mm resolution).

- Neuroscientists are used to thinking in terms of brain regions or brain structures, these being defined based on sulco-gyral anatomy, standard labels (e.g. Brodmann areas), functional characteristics (e.g. V1, *the fusiform face area*) and possibly anatomical/functional connectivity (e.g. *the post-cingular part of the default mode network*).

A possible solution consists in restricting the analysis to regions of interest (ROIs) defined on an individual basis [14, 179], but this often requires the acquisition of additional data, and the definition of the ROIs might be suboptimal or biased [70]. Our solution is to define such regions of interest based on multi-subject data related to a particular contrast of interest. Doing so, we might create regions tailored to each subject's functional anatomy, but in a consistent manner. As stated above, the definition of such regions is in fact a natural output of group studies, and one may compare it to more standard procedures such as those described in Sec. 3. However, the main difference here is that inference is performed on more complex objects than voxels.

Since the main criterion to validate such regions is their cross-subject consistency or reproducibility, we call them (Brain) Functional Landmarks (FLs) [205]. Ultimately, FLs can be defined for several functional contrasts, and can become the basis of anatomo-functional atlases.

On the definition of FLs In this next sections, we describe and discuss several approaches on the way to build a population-level fMRI activity model given several individual instances.

This way to analyse fMRI datasets has been pioneered in [30, 154], where the fundamental object was the scale-space activity blob. In this sequel, we revisit this kind of ideas by distinguishing and emphasizing three different sub-components of the FL definition:

i. the structural point of view, which consists in extracting objects that represent meaningful information in individual data,

ii. a probabilistic point of view, which aims at assessing the significance of these features within the group dataset, and

iii. a pattern matching point of view, which extracts the common information within a group of subjects and thus infers a population-level pattern.

4.2 Structural point of view

4.2.1 What is a structural approach for fMRI data ?

In this section, we use the term *structural* to characterize a family of image processing techniques that extract more or less complex object or primitives from a certain data domain (image or mesh). In a multiple-subject dataset, this procedure is clearly run in a parallel fashion on all individual datasets.

Let $\phi : [1, .., N_v] \rightarrow \mathbb{R}$, be such an image or texture, where the set of sites $[1, .., N_v]$ is assumed to endowed with a spatial model, e.g. the position in 3D space or on a certain brain mesh. The position

of a site v will be denoted $\tau(v)$. In this section, we will generally consider that the spatial structure is a regular 3D grid, but the generalization to other structures, e.g. meshes, is generally straightforward, assuming that a common coordinate is available in all the subjects. It is assumed here that the values of $\phi(v)$ can be interpreted as low or high evidence for the presence of an activation, which is typically the case if ϕ is in fact a statistic, e.g. a Student t-test (see Eq. 2.13).

The question is thus: Which primitives can be extracted from an fMRI image, which contain some relevant information in order to allow cross-subject comparison ? Voxel-level data is at the same time too noisy, redundant (if a voxel v is active, its neighbors are also probably active), and unnecessarily complex, thus hard to compare across subjects. Note that one can concentrate on positive values of the map, given that changing the sign will enable one to reproduce that analysis on the negative values, but with the opposite interpretation: the map-level minimum for a contrast is the maximum of the opposite contrast. Moreover, there is a consensus on the fact that values of ϕ close to 0 are of no interest. To see this (and in fact to understand all the structural approach), it useful to using a Bayesian interpretation of ϕ, which we assume to be a student test with ν degrees of freedom (see Sec. B.2):

$$p(H_1(v)|\phi(v)) = \mathcal{T}^{-1}_{\nu,\phi(v)}(\infty) - \mathcal{T}^{-1}_{\nu,\phi(v)}(0), \tag{4.1}$$

where $\mathcal{T}_{\nu,\phi(v)}$ is the non-central Student distribution with ν degrees of freedom and non-centrality parameter $\phi(v)$ (see [66]). Thus a value of $\phi(v)$ close to 0 provides little evidence for the voxel to be active.

It follows that the information of interest in map ϕ is in the maxima of this maps, or at least within its highest values [1].

4.2.2 A typology of structures of interest

In this section, it is assumed that one particular subject is under study. In a multi-subject case, the procedure is iterated across subjects. We can define at least five models, which are typically borrowed from mathematical morphology and computer vision approaches.

- Supra-threshold connected regions or clusters: this is in fact the basic output in neuroimaging inference, and is thus a basic ingredient common to all researchers. However, its has a lot of drawbacks: *i)* it depends heavily on an arbitrary threshold ϕ_c, *ii)* it blurs all the difference of activity within the supra threshold region, *iii)* it can be very inaccurate if the active region is large *iv)* it can vary a lot from subject to subject, depending on the contrast-to-noise ratio of each dataset.

- The local maxima of ϕ: extracting local maxima provides a complementary feature which has good properties in terms of sensitivity (in each active region there is indeed a local maximum), though

[1] Let us note that these considerations could be circumvented by simply observing that neuroscientists themselves, when they look a fMRI datasets, typically use a color map which emphasize the peaks, and in a blink they can identify the information conveyed by a map, by simply concentrating on the bright spots. Imitating their attitude is probably a safe choice.

not very specific (local maxima occur in the background noise too). Part of the solution consists in introducing signal and spatial/topological thresholds to select significantly high and distant local maxima. Although these thresholds may be viewed as arbitrary, their effect is less dramatic than in the case of supra-threshold clusters. As a complement, local maxima may be selected as significant or not a posteriori (see further). May be the main drawback is that the notion of the spatial extent and spatial configuration of active regions is completely lost, which may ultimately complicate the inter-subject assessment. The next methods aim at combining the advantages of supra-threshold regions and local maxima.

- Watershed segmentation: The watershed segmentation [230] consists in separating all the basins that appear when considering the function $-\phi$ as a landscape. Starting from any voxel -with the exception of frontiers, i.e. saddle points, but this particular case is handled easily-, a gradient descent will reach a certain local minimum of $-\phi$, hence this voxel belongs to the basin associated with the local minimum. In a certain sense, a watershed segmentation is simply an augmentation of the local maxima (thinking of ϕ) with the surrounding region, possibly after some thresholding. It provides some new features, such as a neighboring system between segmented structures. The main drawback is that spurious basins may be present due to noise: here we have the same lack of specificity as in the definition of local maxima. It is possible to handle this kind of case by allowing basin merges, according to some morphological criteria (see e.g. [143]); however, the implied criteria are quite difficult to manage, and to calibrate correctly on real data.

- Nested blobs: Assume that supra-threshold regions of ϕ are computed with decreasing thresholds: each local maximum corresponds to a novel cluster, and clusters merge at saddle points of ϕ. We define a blob as a maximal set of connected voxels comprising a certain number of the maps local maxima, i.e. as it appears just before crossing a saddle point. When the saddle point is crossed, the two blobs merge into a larger blob, so that the family of blobs is naturally nested. In terms of graphical model, the set ob blobs can be organized in a forest, i.e. a set of trees, in which each node represents a set of connected voxel at a certain threshold, and edges represent inclusion relationships. The leaves of the forest are associated with a unique local maxima, while the roots correspond to some supra-threshold regions. If the cluster forming threshold ϕ_c is equal to $-\infty$, the entire dataset corresponds to one single tree.

- Scale-space blobs. This alternative is based on a scale-space representation, which is a principle borrowed from computer vision [118] known as the primal sketch; essentially, the idea consists in smoothing the dataset at different scales and to extract the blobs at these different scales. The purpose of this approach is to combine the robustness of coarse scales with the precision of finer scales. The extracted blobs at each scale correspond to the leaves of the nested blob model, but the blobs are then linked to each other across scales, and/or associated through bifurcations of the scale/space structure. The final object is thus a scale-space blob [30]. This approach is clearly more elaborated than all the previous ones, but the complexity of the resulting objects might create more difficulties for calibration purpose (choice of parameters) and for subsequent steps (statistical

assessment, inter-subject comparison) ; in particular, it is no longer clear in which probabilistic space such objects are defined.

The blob model and the watershed Lets us give the formal definition of the watershed and blob models used in our work. We have a map ϕ and a threshold ϕ_c (in general a lenient threshold, corresponding e.g. to a p-value $p < 0.01$ uncorrected). The 2D and 3D topological structure is encoded in a neighborhood system $\sim$ (e.g. 18 nearest neighbours in 3D).

The watershed model labels supra threshold voxels so that, if λ is the labelling function, the following property holds:

$$\forall v \in [1..N_v] \text{ s.t. } \lambda(v) > \phi_c, \lambda(v) = \lambda(\text{argmax}_{w \sim v} \phi(w)) \tag{4.2}$$

The resulting domains $(a_j), j = 1..J$ are the level sets of λ. Each one is associated with a unique local maximum of ϕ. We then can associate each a_j with a signal level, typically $\phi(a_j) = mean_{v \in a_j}(\phi(v))$ Moreover the set $A = (a_j), j = 1..J$ is naturally endowed with a graph structure through the equivalence relationship:

$$a_j \simeq a_k \text{ iff } \exists (v, w) \in a_j \times a_k : v \sim w \tag{4.3}$$

Let us denote $\mathcal{G}$ the resulting undirected graph. $\mathcal{G}$ may contain loops. To avoid this, we can modify the relationships (which is then no longer an equivalence relationship): Let us assume that $\phi(a_j) < \phi(a_k)$

$$a_j \ll a_k \text{ iff } a_j \simeq a_k \text{ and } \phi(a_k) = max_{a_l \simeq a_j} \phi(a_l) \tag{4.4}$$

the $\ll$-relationship, after symmetrization, can be expressed as an directed acyclic graph G.

The blob model, in turn, consists in connected components of voxels that are above a saddle point of ϕ. Equivalently, a blob a_j corresponds to a maximal supra-threshold connected component of ϕ for a certain threshold ϕ_c. The important point is that the blob structure of a map ϕ are nested: two blobs separated by a certain saddle point above ϕ_c are included in a larger blob, which may itself be contained in a larger one, and so on.

The nested structure inherited from the inclusion relationship can be written as a directed acyclic graph, which in fact a forest G, i.e. a set of trees (one for each) supra-threshold component of ϕ. The interesting point is that this hierarchical structure can be used to *denoise* the graphical model. Assume e.g. that a blob a_j is smaller than a certain reference size s_m; then a_j can be merged into the parent region $a_{p(j)}$. Given that G is in general a binary tree, the parent node $a_{p(j)}$ may contain only its other child node a_k; this means that there is no reason to retain the distinction between a_k and $a_{p(j)}$: a_k may thus be merged into $a_{p(j)}$ in turn. Given an initial size threshold s_m on the blob size this procedure simply cleans the blob structure. See an illustration in Fig. 4.1.

Alternatives A few contributions have proposed some analysis schemes that are not based on mathematical morphology, but on generative models, where the landscape implied by ϕ is modeled by Gaussian structures, with sparsity constraints [162]. A recent alternative has been proposed based on Dirichlet process mixture model [108, 107], which allows to deal nicely with the unknown number of components.

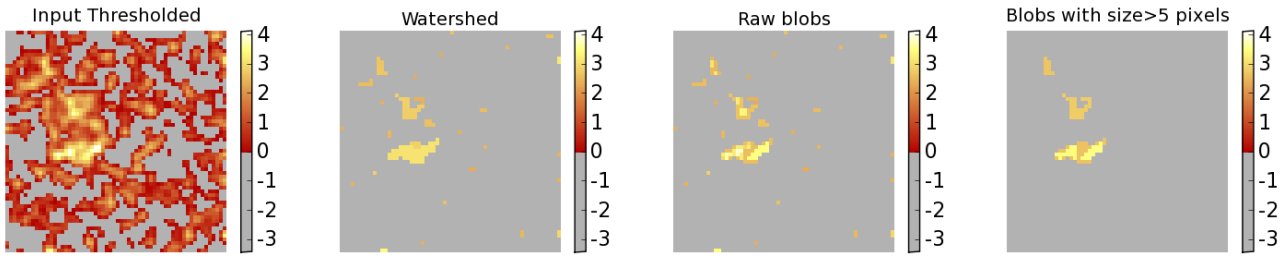


Figure 4.1: Morphological processing of an activation map. From left to right: the input map ϕ_s, can be processed subject to watershed segmentation, or to a blob extraction procedure, with a threshold on the components size $s_{min} = 0$ voxels, or $s_{min} = 5$ voxels. We prefer this last solution that discards that retains the information of the input image with a good denoising. In the three right figures, the regions (a_j) are represented by their average signal $\phi^j, j = 1..J$.

However, these models are very difficult to parametrize, and yield a computational burden, even of the case of very small datasets. Moreover, the probabilistic interpretation is usually lost through successive simplifications.

By contrast, mathematical morphology provides easy-to-use and interpretable features, and extremely efficient computations (typically of the order of 1s for real datasets). Our strategy is thus to concentrate on simple models (watershed, nested blobs) and to endow them with some probabilistic significance in a second step.

In the sequel, we denote the extracted regions $(a_j^s)_{j=1..J(s),s=1..S}$ where s is the individual dataset under study and $j \in [1, J(s)]$ is a region index. The average position of such regions is denoted t_j^s, and is typically a 3D vector. Note that in the case of the nested blob model, the position t_j^s of a non-leaf node a_j^s is taken as the center of mass of the set of blobs that are included within a_j^s (i.e. the center of mass of s_j^s,its children and so on recursively).

For each structure $a_j, j \in \{1, .., J\}$, we can derive an average signal within that structure:

$$\phi_j^s = \phi(a_j^s) = mean_{v \in a_j^s} \phi(v) \tag{4.5}$$

4.2.3 Estimating the statistical significance of extracted structures

Deciding whether an extracted structure is a true or false positive is a difficult question that requires the comparison of multi-subject data, and a spatial model. However, it may be useful to estimate at the subject level whether there is much evidence to view a structure as activated or not. In particular, this can mitigate the impact of an arbitrary threshold: two supra-threshold regions might be handled differently if the values of ϕ_j^s is high or just above threshold.

For this purpose, we propose to model the histogram of ϕ through a mixture model, that typically

includes three modes : one null mode for non activated voxels (H_0 hypothesis), one positive mode for positively activated voxels (H_1 hypothesis), and one negative mode for negatively activated voxels (H_{-1} hypothesis). Note that while the H_{-1} is of no practical interest here, it is useful to avoid any bias on the estimation of the H_0 mode.

We have considered three models:

- A Gamma-Gaussian mixture model, which uses a gamma density for H_1 and H_{-1} components (with a sign swap for the negative one), and a Gaussian model for the H_0 component. The estimation is simply carried out using an EM algorithm.

  The interesting point is that the H_1/H_{-1} components are inherently forced in the positive/negative domain, and that both densities are 0 at 0 if the shape parameter is greater than 1 ; moreover, the gamma densities are a practical choice for the putatively fat tails of the density.

  The main problem is that in many cases, the central part of the histogram is often not exactly Gaussian, so that the gamma components fit this lack of normality rather than the tails of the histogram, resulting in a completely suboptimal histogram fit. To solve this, priors must be introduced, which reduces the usefulness of this approach.

- A three-components Gaussian mixture model (GMM). However, a purely maximum likelihood estimate may result in odd behaviour due to e.g. unconstrained variances. It is thus safer to endow the GMM with conjugate prior on the components means, weights and precision. In this case, estimation is much simpler and can be carried using Gibbs sampling or Variational Bayes. We found this approach to be simple and efficient.

  The mixture model provides the mixture weights $p(H_i), i \in \{-1, 0, 1\}$, the likelihoods $p(\phi|H_i), i \in \{-1, 0, 1\}$, and the posterior $p(H_i|\phi), i \in \{-1, 0, 1\}$.

  Given the values ϕ_j in Eq. (4.5),we simply take the posterior $p(H_i|\phi_j), i \in \{-1, 0, 1\}$.

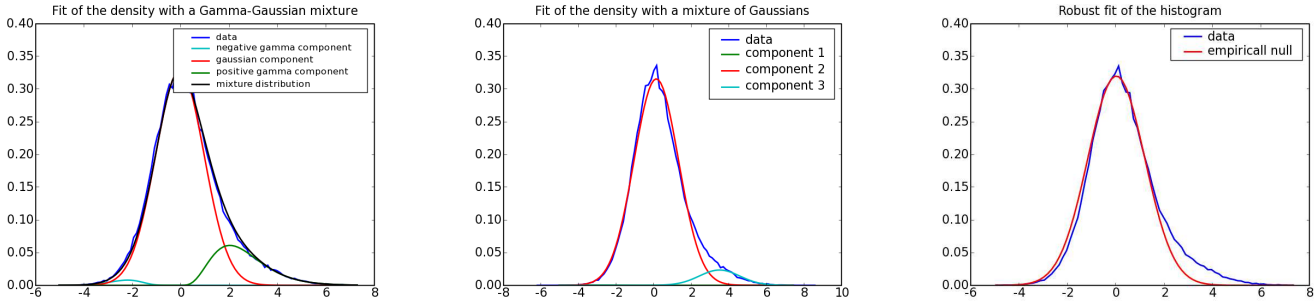
  $$p(H_i|\phi_j) \propto p(\phi_j|H_i)\pi_i \tag{4.6}$$

- Another approach consists in using the positive false discovery rate approach [194, 180]. The positive FDR is defined as

  $$pFDR(\phi) = \mathbb{E}\left(\frac{V(\phi)}{R(\phi)}|R(\phi) > 0\right) \tag{4.7}$$

  where $V(\phi)$ and $R(\phi)$ are the number of false discoveries and of supra-threshold voxels at threshold ϕ. The positive FDR can be interpreted at the posterior probability of a given voxel being inactive given the data, i.e. $p(H_0(a_j^s)|\phi_j^s)$; consequently, it also yields $p(H_1(a_j^s)) = 1 - p(H_0(a_j^s))$. The advantage of the positive FDR over the previous solutions is that it only entails the estimation of *i)* the proportion of null voxels *ii)* the distribution of null voxels. This can simply be obtained by fitting the histogram of the data with the expected null distribution [180].

- Finally the discrete local maxima approach discussed in Sec. 2.4 might be considered here.



(a) (b) (c)

Figure 4.2: Different examples of the fit of the histogram of an activity map ϕ_s: (a) a Gamma-Gaussian mixture model, (b) Gaussian mixture model (c) estimation of the null mode of the histogram with the method of [180]. In any case, it is possible to derive a probability that a given value of ϕ corresponds to an activation, but the results are model-dependent.

4.3 Probabilistic assessment of the structures across subjects

Principle The next step consists in validating the ROIs at the group level, i.e. to retain only regions that are at similar locations across subjects. We do so by computing some posterior significance values $p(H_i(a_j^s)), i \in \{0,1\}$ that represent the evidence that a_j^s is a landmark region based on spatial information.

The essential hypothesis is that true maxima tend to be clustered in the common space, while false positive occur at any location in the standard space. Hereafter, we present two points of view: one is based on non-parametric assessment on the spatial density [207, 206], and the other one uses a parametric model that relies on Dirichlet Process Mixture Model [210]. This second model can incorporate signal-based priors such as those defined in Eq. (4.6).

Non-parametric approach Given that the threshold ϕ_c used for the initial thresholding should preferably be kept low to avoid false negatives, the first step necessarily results in several false positives. A statistical test on the spatial distribution of the maxima is thus performed to control the false positive rate. Only regions with across-subject reproducible activity are of interest. Thus we build a spatial statistic to remove the local maxima in each subject that are far from local maxima in other subjects. This spatial statistic is the density of presence of supra-threshold local maxima in the other subjects. Let τ be an arbitrary position in the common space. For each subject $s \in \{1, .., S\}$, we define

$$\mathcal{D}_s(\tau) = \sum_{\sigma \in \{1,..,S\}-\{s\}} \sum_{i=1}^{I(\sigma)} \exp\left(-\frac{\|\tau - t_j^\sigma\|^2}{2\delta_\tau^2}\right) \tag{4.8}$$

The parameter δ_τ represents an inter-subject spatial variability and is typically set to 5mm.

The distribution of the quantity $\mathcal{D}_s(\tau)$ can then be compared, in every location, to its distribution under the null hypothesis H_0. The null hypothesis that we consider here is global, i.e. it means that there is no specifically task-related region in the brain. Under this assumption, the spatial density of local maxima is uniform in the brain volume. We estimate the distribution of $\mathcal{D}_s(\tau)$ under H_0 by random resampling of the position of the activation maxima $(a_i^\sigma)_{i=1..J(\sigma),\sigma\neq s}$ within the brain volume. Let $\tilde{\mathcal{D}}_s$ be the surrogate distribution obtained after k resamplings ($k = 10$ typically).

Then, let α be a significance level, and let u_α be a threshold on the values of $\mathcal{D}_s$ such that $P(\mathcal{D}_s(\tau) > u_\alpha|H_0) < \alpha$; an estimator of u_α is given by the α-quantile of the density $\tilde{\mathcal{D}}_s$

$$u_\alpha = \text{arginf}_u \left[\frac{1}{\Omega}\int \mathbb{I}_{\tilde{\mathcal{D}}_s > u}(\tau)d\tau < \alpha\right] \tag{4.9}$$

where $\Omega = \int d\tau$ is the brain volume.

Importantly, the test is performed for a small number of spatial locations $(t_j^s)_{i=1..J(s)}$. Hence its significance can be corrected using a Bonferroni procedure, i.e., by replacing α by $\frac{\alpha}{J(s)}$ in Eq. (4.9). An example is provided in Fig. 4.3.

One might be concerned with the behaviour of the method, assuming that the null hypothesis has been rejected in some regions of the brain: Does the test remain valid in the other regions, given that the global null hypothesis of a uniform density of maxima has been rejected ? In fact, in such case, the resampled distribution $\tilde{\mathcal{D}}_s$ is an *overestimation* of the true null distribution under the null hypothesis $P(\mathcal{D}_s|H_0)$, which means that the ensuing test is conservative, hence valid. This fact is evident in Fig. 4.3, where the (null) mode of the resampled distribution is shifted to the right, with respect to the mode of the non-resampled distribution.

The test is iterated in all the subjects, then non-significant maxima at the desired significance level are rejected. The process can be iterated in order to refine the spatial model. Let $\mathcal{J}(s) \leq J(s), s \in \{1, .., S\}$ be the number of remaining regions in each subject. Since we control the (corrected) probability that one false positive region might show up in a given subject at level α, given $\nu \in \{1..S\}$, the probability of one false positive region in ν subjects over S is controlled by the binomial law $\mathcal{B}(\nu, S, \alpha)$.

Parametric approach The second approach that we propose is formulated in a Bayesian setting:

$$p(H_i(a_j^s)|t_j^s, \phi^s) = \frac{p(t_j^s|H_i(a_j^s), \phi^s)p(H_i(a_j^s)|\phi^s)}{p(t_j^s|\phi^s)} \tag{4.10}$$

where $p(t_j^s|\phi^s) = \sum_i p(t_j^s|H_i(a_j^s), \phi^s)p(H_i(a_j^s)|\phi^s)$. Moreover, we further assume the conditional independence of the positions of ROIs and of their significance given a certain hypothesis on these ROIs $p(t_j^s|H_i(a_j^s), \phi^s) = p(t_j^s|H_i(a_j^s))$. In order to estimate the spatial densities $p(t_j^s|H_i(a_j^s))$ we must specify a spatial model under each alternative.

- under H_0, the selected ROIs are distributed uniformly across the brain, so that $p(t_j^s|H_0(a_j^s)) = 1/|\Omega|$, where Ω is the brain volume

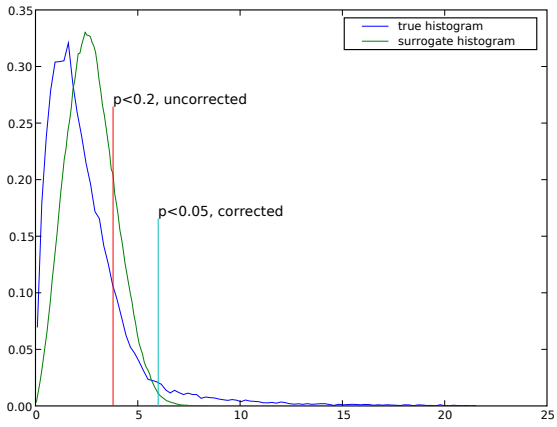


Figure 4.3: Modeling the density of $\mathcal{D}_s$ in the volume. Randomly reshuffling the position of the maxima of activity in subjects $\sigma \in \{1, .., S\} - \{s\}$ yields an empirical histogram of $\mathcal{D}_s$ under the null hypothesis (green), which can be used to define critical values u_α, which can be corrected for multiple comparisons or not, to threshold the density $\mathcal{D}$.

- under H_1, the distribution $p(t_j^s|H_1(a_j^s))$ is unknown, but is expected to be clustered in some regions of the brain. To model this density, we use a Gaussian Mixture Model with an unspecified number of components implemented through a Dirichlet Process Mixture Model (DPMM).

A Dirichlet Process $DP(\theta, G)$ with a concentration parameter $\theta > 0$ and a base measure G can be used as a nonparametric prior distribution on mixing proportion parameters in a mixture model when the number of components is unknown a priori [145, 171]. The generative process for a mixture of Gaussian distributions with component means μ_k and covariance Λ_k and DP prior $DP(\theta, G)$ can be written

$$\begin{aligned} g &\sim DP(\theta, G), \\ (\mu_j, \Lambda_j) &\sim g, \\ t_j &\sim \mathcal{N}(t_j; \mu_j, \Lambda_j), \end{aligned} \tag{4.11}$$

where g is a random prior distribution on the parameters sampled from $DP(\theta, G)$. Essentially, the Dirichlet process $DP(\theta, G)$ converts the continuous distribution G into a discrete one g. The fundamental property of Dirichlet process is the following: if $\{x_i = (\mu_i, \Lambda_i)\}$ are samples from g, then the distribution of g given $(\{x_i\})$ is

$$g|\{x_i\} \sim DP\left(\theta G + \sum_i \delta_{x_i}\right) \tag{4.12}$$

Thus the concentration parameter *theta* allow the creation of new classes given the already available ones. In practice, g can be integrated out, so that

$$x_i|\{x_{-i}\},\theta,G \sim \theta G + \sum_i \delta_{x_{-i}} \tag{4.13}$$

The base measure G is uniform over the brain volume for the mean parameters, and is an Inverse Wishart distribution for the covariance

$$G = p(\mu,\Lambda_k) = \frac{1}{|\Omega|} IW(\nu\sigma^2\mathbb{I},\nu), \tag{4.14}$$

where σ and ν are the hyper-parameters of the model. The estimation of such models can be performed fairly easily using MCMC techniques, like Gibbs sampling. By keeping the hyper-parameters fixed, we can moreover integrate out all the random variables except z_j^s (*collapsed* Gibbs sampling procedure, which is known to speed up convergence [121]). The algorithm iteratively samples z_j^s from

$$p(z_j^s = k|t_j^s, z^{-s}, \theta) \propto p(t_j^s|t^{-s}, z^{-s}, z_j^s = k)p(z_j^s = k|z^{-s},\theta), \tag{4.15}$$

where z^{-s} represents the values of z for the ROIs in subjects other than s and t^{-s} represents all the ROI positions in subjects other than s. The first term of Eq. (4.15) simply corresponds to a 3D Gaussian cluster model

$$p(t_j^s|t^{-s}, z^{-s}, z_j^s = k) = \mathcal{N}(t_j^s;\mu_k,\Lambda_k) \tag{4.16}$$

or a constant $\frac{1}{|\Omega|}$ if the component is empty. The second term comes from the partition structure of the Dirichlet process:

$$p(z_j^s = k|z^{-s},\theta) = \begin{cases} \frac{n_k^{-s}}{\theta+N^{-s}} \text{ if } n_k^{-s} > 0, \\ \frac{\theta}{\theta+N^{-s}} \text{ otherwise,} \end{cases} \tag{4.17}$$

where n_k^{-s} is the number of instance of $z = k$ for all regions in subjects other than s and $N^{-s} = \sum_k n_k^{-s}$; $n_k^{-s} > 0$ amounts to considering that k is a previously seen component; k is unvisited otherwise. In our setting, we include an alternative case, in which the region a_j^s is a false positive. The sampling scheme is thus

$$p(H_0(a_j^s)|t_j^s,\phi^s) = \frac{1}{c}\frac{1}{|\Omega|}p(H_0(a_j^s)|\phi^s), \tag{4.18}$$

$$p(H_1(a_j^s), z_j^s = k|t_j^s, z^{-s},\theta,\phi^s) = \frac{1}{c}\begin{cases} \frac{\theta}{\theta+N^{-s}}\frac{1}{|\Omega|}p(H_1(a_j^s)|\phi^s) \text{ if } n_k^{-s} = 0, \\ \frac{n_k^{-s}}{\theta+N^{-s}}\mathcal{N}(t_j^s;\mu_k,\Lambda_k)p(H_1(a_j^s)|\phi^s) \text{ otherwise,} \end{cases} \tag{4.19}$$

where c is the normalizing constant (simply obtained by summation). The parameters $(\mu_k)_{k=1..K}$ and $(\Lambda_k)_{k=1..K}$ are updated naturally according to the current value of (z_j^s). Once the Markov chain has run for long enough, the samples of the chain will be sampled from the true posterior $p(H_i(a_j^s), z_j^s =$

$k|t_j^s, z^{-s}, \theta, \phi^s)$, and the approximate distribution of $p(H_i(a_j^s)|t_j^s, z^{-s}, \theta, \phi^s)$ is given by averaging over the samples. Inference proceeds by selecting the regions a_j^s for which $p(H_1(a_j^s)|t_j^s, z^{-s}, \theta, \phi^s)$ is above some threshold q. This reduces the number of regions to $\mathcal{J}(s) < J(s)$. We found that 1000 iterations over all the ROIs of all subjects are sufficient to achieve a reliable estimate of the posterior.

In summary, the spatial model used here yields the posterior probabilities of the target regions being truly activated regions, given that truly activated regions should be clustered spatially across subjects. This procedure thus infers what are the true regions of activity among all the candidate regions, in each subject.

$$p(H_1(a_j^s)|t_j^s, \phi^s) = \sum_k p(H_1(a_j^s), z_j^s = k|t_j^s, \phi^s) = 1 - p(H_0(a_j^s)|t_j^s, \phi^s) \tag{4.20}$$

this quantity can be readily used to decide which regions are significantly active, given the available data.

In addition, the algorithm also yields the spatial likelihood of the position of activated regions $p(t|H_1)$, which is a very useful feature of the method.

Choosing the priors and parameters The DPMM approach nicely bypasses the troublesome choice of the number of components in the spatial mixture model; It simply requires the definition of the cluster creation parameters θ, and of the priors (σ, ν) for the covariance. We choose $\theta = 0.5$ as usually done in the literature. Once again, given that the method does intend to infer the number of clusters, this choice is not problematic. For 3D problems, we take $\sigma = 5$mm, given that this corresponds to a Full Width at Half Maximum of $\sigma\sqrt{\log(256)} \sim 11.8$mm. Finally, the prior on the degrees of freedom is chosen as $\nu = 10$ to avoid singular solutions even if the clusters are weakly populated ($\nu = \infty$ is also possible).

In terms of results, we found that σ had the largest impact on the results, but even then it is not dramatic: the flexibility of the DPMM easily compensates for suboptimal parameter choices.

4.4 Pattern matching point of view

The previous step provides us with a set of statistically significant and reasonably clustered regions across subjects. However, it is not clear at this point which correspondence exists between these regions. Note that a cluster-based model such as in Eq. (4.19) clearly implies such a correspondence across subjects, but we prefer not to use it: *i)* in the Gibbs sampling procedure, clusters appear and disappear dynamically, so that it is very difficult to assess a clustering model across samples of (z_k); *ii)* so far, these correspondences are based on the absolute position of regions in a common space, which is subject to anatomical variability. We thus incorporate some information on the relative position of regions, in order to obtain more reliable correspondences.

A graph matching approach for the non-hierarchical model Here it is assumed that the available regions are simply contiguous, without any notion of hierarchy. The contiguity model is expressed in a graph that can be acyclic (G^s) or not ($\mathcal{G}^s$), as defined in section 4.2. Note however that non-significant nodes have been removed. We base our search of inter-subject correspondences on the *relative positions*

of activated areas. While this does not rely on a physical model, it builds on the heuristic that across subjects, positions of activated regions should be locally similar, though not identical in the common space. For instance, it is logical to favor configurations in which couples of neighboring regions with similar relative positions in two subjects will be grouped in two different cliques according to their relative position.

Probabilistic associations are then searched for each pair of subjects, using a belief propagation (BP) algorithm [157]. Given a reference subject s_1 and a target subject s_2, the associations are initialized as

$$p(a_i^{s_2} \leftarrow a_j^{s_1}) \propto \exp\left(-\frac{\|t_i^{s_2} - t_j^{s_1}\|^2}{2\delta_\tau^2}\right) \tag{4.21}$$

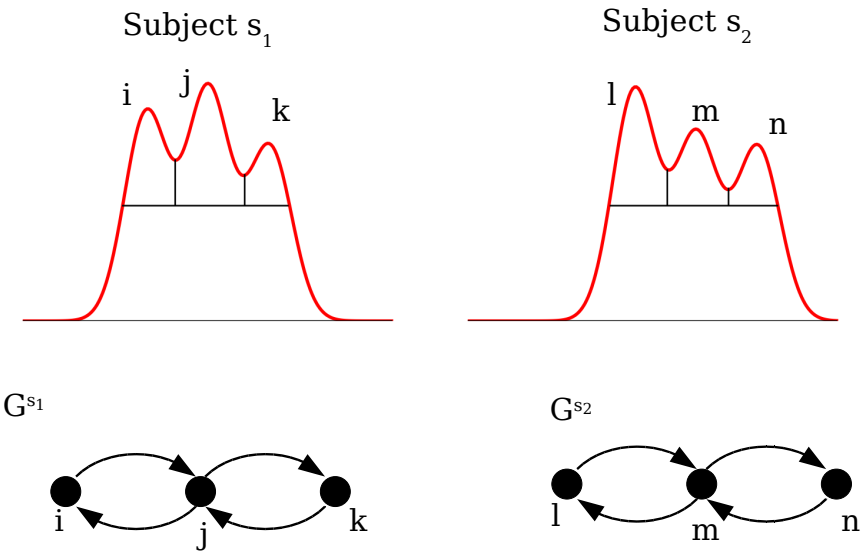
with appropriate normalization $\left(\sum_{j=1}^{\mathcal{J}(s_1)} p(a_i^{s_2} \leftarrow a_j^{s_1}) = 1\right)$, where $P(a_i^{s_2} \leftarrow a_j^{s_1})$ stands for the probability that the region $a_i^{s_2}$ in subject s_2 is the homologue of region $a_j^{s_1}$ in subject s_1. These probabilities are refined by belief propagation; for each edge (jk) of the graph G^{s_1} or $\mathcal{G}^{s_1}$, messages are sent from $a_j^{s_1}$ to $a_k^{s_1}$ to quantify the probability of association between $a_k^{s_1}$ and $(a_i^{s_2})_{i=1..\mathcal{J}(s_2)}$:

$$m_{jk}(i) \propto \sum_{l=1}^{\mathcal{J}(s_2)} p(a_l^{s_2} \leftarrow a_j^{s_1}) \exp\left(-\frac{\|(t_i^{s_2} - t_l^{s_2}) - (t_k^{s_1} - t_j^{s_1})\|^2}{2\delta_\tau^2}\right) \tag{4.22}$$

with appropriate normalization $(\sum_{i=1}^{\mathcal{J}(s_2)} m_{jk}(i) = 1)$. Eq. (4.22) simply means that whenever the positions of local maxima $t_l^{s_2}$ and $t_i^{s_2}$ in subject s_2 and $t_j^{s_1}$ and $t_k^{s_1}$ in subject s_1 form a parallelogram, the configurations are favored in which $a_l^{s_2}$ and $a_j^{s_1}$ on the one hand, $a_i^{s_2}$ and $a_k^{s_1}$ on the other hand, are associated. As shown in Fig. 4.4, taking into account the *relative* positions of the maxima improves cross-subjects correspondences by compensating global translation effects. The beliefs $p(a_i^{s_2} \leftarrow a_j^{s_1})$ and messages are then updated and normalized according to the formal laws of BP [157, 245]. Note that the graphs G^s have no loops, so that convergence is straightforward. We have also used loopy belief propagation, based on the graphs $\mathcal{G}^s$, which did not raise any issue concerning the convergence of the correspondence probabilities in practical cases. Furthermore, the choice of G^s or $\mathcal{G}^s$ was not found to be crucial in the method. In our experiments, we use G^s by default.

The estimation of the probabilities is performed on each pair of subjects in the group. As an important note, all the quantities used here are asymmetric. In particular, the graphs G^{s_1} and G^{s_2} have a priori different structures, so that the probabilities $P(a_i^{s_2} \leftarrow a_j^{s_1})$ and $P(a_j^{s_1} \leftarrow a_i^{s_2})$ might be quite different after convergence. This is particularly obvious in the case of many-to-one correspondences, given that the number of local maxima in a given region may vary a lot across subjects. The next step essentially chooses which of these correspondence are meaningful at the population level.

A Bayesian networks approach for the nested blob (hierarchical) model. Here, it is assumed that the regions are arranged in a hierarchical model, namely the nested blob model, which amounts to defining a tree structure in each subject. Although the DPMM yields a clustering of the individual



Initialization

	$P(a_l^{s_2} \leftarrow .)$	$P(a_m^{s_2} \leftarrow .)$	$P(a_n^{s_2} \leftarrow .)$
$P(. \leftarrow a_i^{s_1})$	**0.59**	0.31	0.10
$P(. \leftarrow a_j^{s_1})$	**0.42**	0.38	0.20
$P(. \leftarrow a_k^{s_1})$	0.25	**0.39**	0.35

After convergence

	$P(a_l^{s_2} \leftarrow .)$	$P(a_m^{s_2} \leftarrow .)$	$P(a_n^{s_2} \leftarrow .)$
$P(. \leftarrow a_i^{s_1})$	**0.74**	0.23	0.03
$P(. \leftarrow a_j^{s_1})$	0.41	**0.46**	0.13
$P(. \leftarrow a_k^{s_1})$	0.14	0.40	**0.46**

Figure 4.4: Illustration of the use of the Belief Propagation algorithm to find correspondences between maxima within a pair of subjects. This is a toy dataset, in a one-dimensional space. The activity maps of subjects s_1 and s_2 are shown on the top of the figure, together with a watershed segmentation. In that case, $t_i^{s_1} = 0, t_j^{s_1} = 1$ and $t_k^{s_1} = 2$, while $t_l^{s_2} = 0.7, t_m^{s_2} = 1.7$ and $t_n^{s_2} = 2.7$; $\delta_\tau = 1.4$. The related graphs G^{s_1} and G^{s_2} are shown below; in this case they are isomorphic. The associations are initialized using Eq. (4.21), and then refined using Eq. (4.22): Clearly the message passing algorithm enhances the probabilities $P(a_l^{s_2} \leftarrow a_i^{s_1}), P(a_m^{s_2} \leftarrow a_j^{s_1})$, and $P(a_n^{s_2} \leftarrow a_l^{s_1})$, thus compensates the effect of the global translation between the two datasets. This effect is also present with the converse probabilities $p(a_.^{s_1} \leftarrow a_.^{s_2})$.

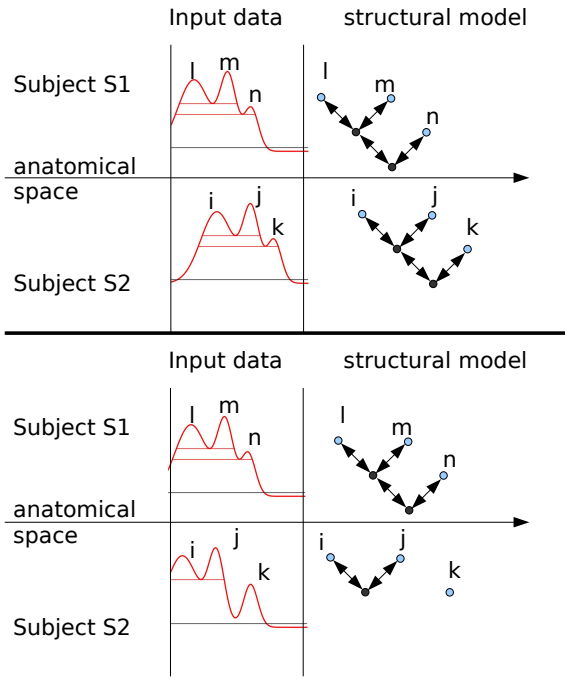
regions through the z_j^s variables in Eq. (4.19), the inter-subject correspondence implied by this model may not be very reliable, due to the relative simplicity of the Gaussian generative model. In order to improve it, it is useful to take into account the topology of the individual patterns, which is coded in the graphical model of the nested level sets model defined in Sec. 4.2. Recall that these nested ROIs in each subjects are embedded in a *forest* (i.e. multiple trees) graphical model. Only the *leaves* of the forest will be included in the association model, since they represent maxima, i.e. putative foci of activity. Other (*saddle*, see section 4.2) regions are only used in order to code the large-scale structure of the functional activity pattern in the data. This is an important because, this indicate that two separate peaks of activity may in fact be included within a crest line, and should thus be dealt with as such when building the functional template. Probabilistic associations are searched in each pair of subjects. Given a reference subject s_1 and a target subject s_2, let $p(a_i^{s_2} \leftarrow a_j^{s_1})$ be the probability that region a_j in subject s_1 is associated with region a_i in subject s_2; let $a_{pa(j)}^{s_1}$ be the *parent* of $a_j^{s_1}$ in the graphical sense. The global association model is given by a Bayesian Network (BN) defined by the conditional probabilities:

$$p(a_k^{s_2} \leftarrow a_j^{s_1} | (a_l^{s_2} \leftarrow a_{pa(j)}^{s_1})) \propto \delta_{a_{pa(k)}^{s_2} = a_l^{s_2}} \mathcal{K}_\sigma(t_j^{s_1} - t_k^{s_2} - (t_{pa(j)}^{s_1} - t_l^{s_2})) \tag{4.23}$$

Here $\mathcal{K}_\sigma$ is a standard Radial Basis Function (RBF) kernel with parameter σ. The second term in Eq. (4.23) is crucial, since it implies that the children of two nodes in subjects s_1 and s_2 with a similar relative position will have a stronger probability of association, which is our core intuition. The associations are initialized at the leaf level using Eq. (4.21). The association probabilities are updated using a belief propagation (BP) algorithm [157]. This procedure is known to converge given that all the graphs involved in the model are trees. An interpretation and examples of this procedure are shown in Fig. 4.5.

From pairwise correspondences to a group model All the probabilities of all pairwise associations between subjects are then arranged in a common belief matrix B. A row of B contains all the probabilities that a region a_i^s in a subject is associated with all regions $(a_j^\sigma)_{j=1..\mathcal{J}(\sigma),\sigma\in\{1,..,S\}}$ in the other subjects; note that the associations with the other maxima $(a_j^s)j = 1..\mathcal{J}(s)$ within the subject s itself are null. We can also interpret B as the adjacency matrix of a probabilistic association graph between regions that are pooled across subjects. Some of these associations may not be realistic, but in that case the two maxima should not have strong common associations with other maxima. To deal with such cases, we proceed with the extraction of the maximal cliques of the belief matrix, i.e. groups of maxima that have mutually strong associations. In our case, the association probabilities are asymmetric, thus the maximal clique approach requires that association probabilities are *relatively* high bidirectionally, in a sense detailed thereafter.

Many clustering procedures are possible, e.g. hierarchical clustering techniques, using average or maximal linkage heuristics or replicator dynamics (RD). Since the latter procedure is more data-driven (it does not require a prior definition of the number q of clusters to be found), we describe it in more details, but we also suggest to use average-link agglomerative clustering, where the number of desired clusters is $q = \text{mean}_s \mathcal{J}(s)$ the average number of regions per subject. Note that in that case, the association probabilities are symmetrized.



Initial probabilities

	$p(i \leftarrow .)$	$p(j \leftarrow .)$	$p(k \leftarrow .)$
$p(. \leftarrow l)$	**0.93**	0.07	0.00
$p(. \leftarrow m)$	**0.55**	0.37	0.08
$p(. \leftarrow n)$	0.17	**0.52**	0.31

Final probabilities

	$p(i \leftarrow .)$	$p(j \leftarrow .)$	$p(k \leftarrow .)$
$p(. \leftarrow l)$	**0.98**	0.02	0.00
$p(. \leftarrow m)$	0.13	**0.84**	0.03
$p(. \leftarrow n)$	0.10	0.18	**0.72**

	$p(l \leftarrow .)$	$p(m \leftarrow .)$	$p(n \leftarrow .)$
$p(. \leftarrow i)$	0.33	**0.49**	0.18
$p(. \leftarrow j)$	0.03	0.37	**0.60**
$p(. \leftarrow k)$	0.00	0.18	**0.82**

	$p(l \leftarrow .)$	$p(m \leftarrow .)$	$p(n \leftarrow .)$
$p(. \leftarrow i)$	**0.66**	0.32	0.11
$p(. \leftarrow j)$	0.00	**0.53**	0.47
$p(. \leftarrow k)$	0.00	0.03	**0.97**

Initial probabilities

	$p(i \leftarrow .)$	$p(j \leftarrow .)$	$p(k \leftarrow .)$
$p(. \leftarrow l)$	0.40	**0.57**	0.03
$p(. \leftarrow m)$	0.01	0.27	**0.72**
$p(. \leftarrow n)$	0.00	0.12	**0.88**

Final probabilities

	$p(i \leftarrow .)$	$p(j \leftarrow .)$	$p(k \leftarrow .)$
$p(. \leftarrow l)$	**0.97**	0.00	0.03
$p(. \leftarrow m)$	0.00	**0.28**	0.72
$p(. \leftarrow n)$	0.02	0.12	**0.86**

	$p(l \leftarrow .)$	$p(m \leftarrow .)$	$p(n \leftarrow .)$
$p(. \leftarrow i)$	**0.98**	0.02	0.00
$p(. \leftarrow j)$	**0.66**	0.24	0.10
$p(. \leftarrow k)$	0.02	0.46	**0.52**

	$p(l \leftarrow .)$	$p(m \leftarrow .)$	$p(n \leftarrow .)$
$p(. \leftarrow i)$	**1.00**	0.00	0.00
$p(. \leftarrow j)$	0.01	**0.99**	0.00
$p(. \leftarrow k)$	0.02	0.46	**0.52**

Figure 4.5: Illustration of the probabilistic correspondence computation algorithm on a toy (1D) example: Recall that the input data corresponds to a synthetic model of the level sets of two activation maps of two subjects. The maps maxima are called (l, m, n) and (i, j, k) in the two subjects. (Top) the underlying structure is the same in the two subjects, but there is a translation, so that the trivial association scheme (see Eq. (4.21)) provides unreliable initial probabilities. After convergence of the probabilities through the BN model (see Eq. (4.23)), the probabilities are closer to the ground truth (diagonal terms get higher values). (Bottom) the underlying structure is different in the two subjects. Nevertheless, the algorithm considerably improves the correspondences.

A formal definition of graph-theoretical cliques is given in [156], in the case where the matrix B is symmetric. They are termed the *dominant sets* of the graph, and their definition relies on two conditions: *i)* that the similarity value of each element of the clique should be high enough with respect to the average similarity of the other elements *ii)* that any element outside the clique should have a weaker similarity with the clique than the elements of the clique, where the similarity values are computed from the affinity values between graph neighbors. Finally, it is shown in [156] that it amounts to define a membership vector x on the graph vertices, and then to solve the program

$$\text{maximize } x^T B x \text{ subject to } x \geq 0 \text{ and } x^T e = 1 \tag{4.24}$$

where e is the vector of ones with the same size of x. Finally, still in the case where B is symmetric, this problem can be solved using replicator dynamics equations to B (see e.g. [123, 147]). Replicator dynamics consist in initializing, randomly or not, a positive vector $x^{(0)}$ whose length is equal to the total number of vertices, and then in iterating the update rule

$$x^{(i+1)} = \frac{(Bx^{(i)}).x^{(i)}}{{x^{(i)}}^T B x^{(i)}} \tag{4.25}$$

where $*.*$ stands for the element-wise product. After a few iterations, almost all the components of x vanish, and the other ones correspond to a maximal clique of the belief graph. The clique is removed, and the process is repeated until no non-trivial clique is found. Other rules than the replicator dynamics can be used instead [158], but we experienced that Eq. (4.25) works efficiently. We noticed that this procedure tends to over-segment the graph B, which is natural due to the restrictive definition of the maximal cliques (or *dominant sets*, see above), but this is not a problematic issue, since further merging of cliques remains possible. For instance, the procedure can be iterated based on cliques instead of regions, thus yielding larger cliques.

Finally, all the cliques that contain maxima from at least ν (e.g. $\frac{S}{2}$) subjects over S are retained. Note that this procedure allow an explicit control of the rate of false detections when the classical procedure described in Sec. 4.2.3 is used.

Derivation of a group template This procedure provides us with clusters of activated regions defined across subjects. It does not require that all subjects are represented for a given activated region, and therefore is able to account for some inter-individual differences. A large cluster means that subjects *typically* have an activated region that corresponds to this cluster.

In order to make group maps, we assume that the positions (t_s^i) of the maxima within each clique are normally distributed, and thus represent the cliques through their 95% confidence regions (CR) in the common (MNI) space. Note that the normality hypothesis used here is about inter-subject ROI positions in the reference space, and not about inter-subject signals, which is a key difference with standard techniques; this assumption is used only to define the inter-subject activated regions in the common space.

As a matter of interpretation, these CR regions are quite close to reproducibility maps [119], i.e. maps that count the number of times a voxel is declared active across subjects in some group, because each CR is associated with zero or one particular region in each subject. The interpretation is thus that *a local peak of activity for the proposed task is expected to be observed within the area defined by the CR in a proportion $\frac{\nu}{S}$ of the population.* This is quite different from fixed-effects analyses, which disregard inter-subject variability, and random or mixed-effects analyses (see Sec. 3.1), which yield the probability that the mean effect observed in the population will be positive.

4.5 Use in group studies

We take an excerpt of $S = 130$ subjects of the localizer database described in Sec. B.4.1; we concentrate on the contrast that yields the amount of activity for a subtraction task.

We limit ourself to a brief comparison of our method with a Mixed Effects analysis thresholded at the cluster level (CMFX, $p < 0.05$, corrected) on a subset of 10 randomly chosen subjects. CMFX can be shown to be a particularly sensitive and reliable marker of brain activations. The results are given in Fig. 4.6(f,g). The proposed method extracts 17 regions of activity, while the CMFX extracts 5 supra-threshold clusters (Fig. 4.6(h)); the reason is that *i)* the CMFX merges several regions that are distinguished with our technique *ii)* some small activated clusters detected with our method are not significant for the CMFX test. Additionally, we present 5 individual maps in Fig. 4.6(a-e), in order to suggest that the results are not limited to group maps, but include individual patterns. Corresponding regions have the same color.

Taking advantage of the large cohort, we controlled the reliability of the results obtained in small groups of subjects by assessing the reproducibility of the resulting regions across $R = 10$ disjoint groups of 13 subjects. We consider group-level binary maps obtained with different inference procedures, extract the clusters of size greater than $\eta = 10$ voxels and compute the following penalty, which measures how far supra-threshold clusters are from each other across groups [see also Eq. (3.17)]:

$$\Psi = \frac{1}{R(R-1)} \sum_{r=1}^{R} \sum_{s \in \{1,..,R\}-\{r\}} \frac{1}{I(r)} \sum_{i=1}^{I(r)} min_{j \in \{1,..,I(s)\}} \psi(\|t_i^r - t_j^s\|), \tag{4.26}$$

where (t_i^r) are the positions of the cluster centers, and $\psi(x) = 1 - \exp\left(-\frac{x^2}{2\sigma^2}\right)$. We do this for five statistics: voxel-based Random Effects, thresholded at $p < 0.001$, uncorrected (RFX), cluster-level RFX, where the map is thresholded at the $p < 0.01$,uncorrected level, then at the $p < 0.05$, corrected, cluster-level (CRFX), the same procedures and thresholds with Mixed-Effects models (MFX/CMFX), the group-level confidence regions obtained with the two following procedure:

- Blob model, followed by the Bayesian assessment of the individual activation, thresholding of the posterior probability at 0.9, search for pairwise correspondences using the nested blob model; the resulting regions were selected when they reached the $\nu = 0.4 \times S$ reproducibility threshold. This procedure yields the CR score.

- The Watershed-base model, followed by a classical test on the high density regions, and the non-hierarchical between-subject correspondence model. The resulting regions were selected when they reached the $\nu = 0.4S$ reproducibility threshold. This procedure yields the CR* score.

All the significance levels are obtained using permutation techniques, so that the tests can be considered as exact [141]. The results are given in Fig. 4.6(i). It turns out that the proposed method gets the lowest penalty in average, similar to CR* and CMFX, and moreover with less variability than the other methods. This proves its reliability for group studies. Note that the good results of CMFX are mitigated by the fact that this method only detects large clusters (see Fig. 4.6(h)).

We have also studied the impact of the choice of the graphical model in the non-hierarchical graphical model on the reproducibility indexes obtained by jackknife subsampling, where the κ index is used (see Sec. 3.2.2 the larger the better): we found that the CR regions were more reproducible when obtained using either the acyclic graph G^s ($\kappa = 0.589 \pm 0.01$) or the spatial connectivity graph $\mathcal{G}^s$ ($\kappa = 0.597 \pm 0.01$) than a trivial graph without edges (i.e. without the belief propagation algorithm, see Eq. (4.22), $\kappa = 0.539 \pm 0.01$).

4.6 Comparison of anatomical and functional features on the cortical surface

Once Brain Functional Landmarks have been extracted, it is possible to compare their position with the position of anatomical landmarks, such as the main cortical sulci.

Although it has been suggested that surface-based cortical mapping may be more precise than volume-based brain mapping [55], a quantitative analysis of this approach is still lacking. In this work, we propose to define the position of regions found in functional neuroimaging protocols based on anatomical features extracted individually. On the cortical surface, the sulci represent the main landmarks that might be used to define the position of the regions [173]. More specifically, we propose to use the geodesic distance between target points on the cortical surface and their neighboring sulci to define their position. For instance, in general any point on the cortical surface is uniquely defined by its distance to the three nearest sulci, and we use explicitly this *triangulation* principle to define the position of a region across subjects.

In order to test the validity of this approach, we *i)* use some functional landmark (FL) regions, which are reproducible foci of activity in a given group of subjects, *ii)* measure the variability of the position of these FLs when the data is analysed in the standard space and then projected onto the individual cortical surface and *iii)* compare this with the spatial variability obtained with our triangulation framework. Using cross-validation techniques, we show that the triangulation framework allows in some cases a more precise localization of some functional regions. We discuss the implications and the limits of this approach.

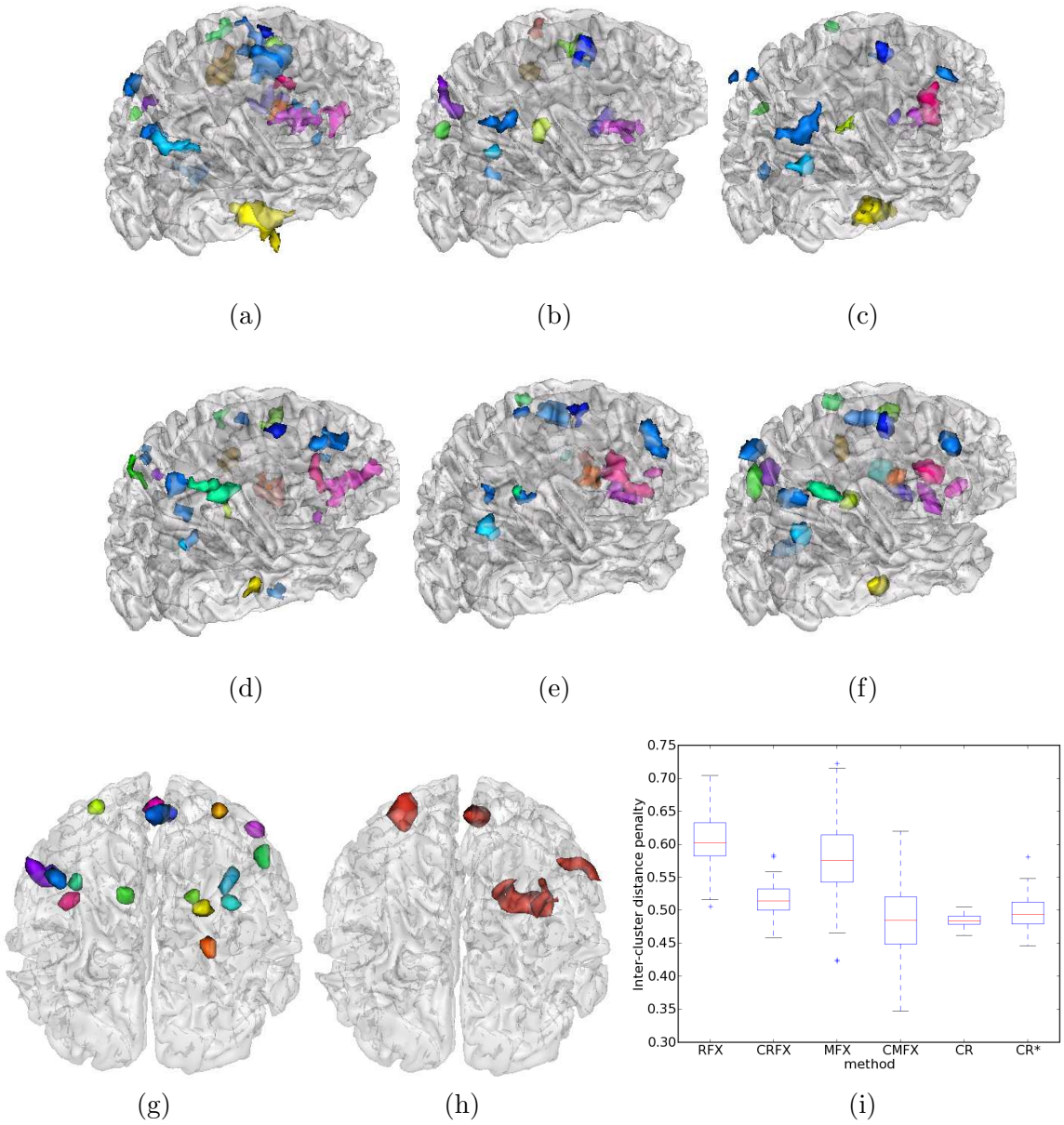


Figure 4.6: (a-e) Analysis of a particular group of 10 subjects: confidence regions found with our method (17 regions) in five subjects and group template (f,g); (h) regions found with a Mixed-Effects test, thresholded at the cluster level ($p < 0.05$ corrected, 5 regions). (i) Inter-cluster distance penalty computed for 13 disjoint random samples of 10 subjects, and 6 different group analyzes techniques: Random Effects (RFX), cluster-level RFX (CRFX), Mixed Effects (MFX), cluster-level Mixed Effects (CMFX), our estimation of regions where activity occurs at the group level (CR), and the regions obtained with the method in [207] (CR*). CR, CR* and CMFX get the lowest penalty, which corresponds to an average inter-cluster distance of 6mm.

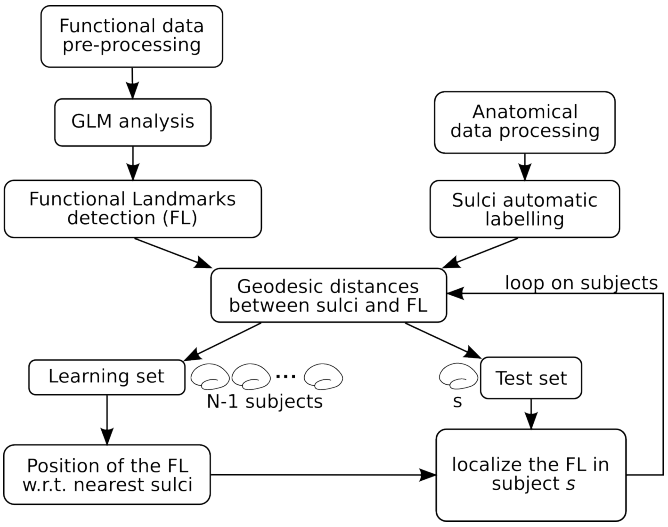


Figure 4.7: Flowchart of the method to compare the reliability of the position of brain functional landmarks on the cortical surface, by using sulci, and in the brain volume, using standard coordinate systems. The input consists of activation images and labelled sulci. The core part of this algorithm makes cross-validated predictions of the positions of the candidate regions in individual space.

4.6.1 Localizing functional regions on the cortical surface

Let $\mathcal{P}^s(.)$ be the function that projects any point of the MNI space onto the cortical surface of the subject s using a nearest-node approach.

Note that the projection is a crucial step, which is problematic in general, given that the EPI images used to obtain functional landmarks are distorted with respect to the anatomy. In this study, the EPI images were corrected to reduce the distortions, but the correction may not be perfect, and a slight bias in the 3D space can have a dramatic effect when projecting the data onto the cortical surface, e.g. projecting a FL onto a wrong gyrus. Let $\gamma_i^s = \mathcal{P}^s(t_i^s)$ be the projection of (t_i^s) onto the grey-white matter interface. The three nearest sulcal lines on the surface around γ_i^s are identified. Let $\Gamma_i(1), .., \Gamma_i(3)$ be these sulci. Note that they are chosen at the group level, and are thus the same across subjects, even if another sulcus is closer in one of the subjects.

We propose to characterize γ_i^s by its geodesic distance to the nearest sulci, $d_{ij}^s = d(\gamma_i^s, \Gamma_i(j))$. We check the worthiness of this characterization which we call *triangulation*, using cross-validation: let $(1..S(i))$ be the set of subjects in which the FL could be reliably defined; then let

$$\begin{aligned} \overline{d_{ij}^s} &= mean_{\sigma \in \mathcal{S}(i)-\{s\}} d_{ij}^\sigma \\ \widetilde{d_{ij}^s} &= \sqrt{var_{\sigma \in \mathcal{S}(i)-\{s\}} d_{ij}^\sigma} \end{aligned}$$

be the mean value and the standard deviation of d_{ij}^s in all subjects but s. Finally, let $\widehat{\gamma_i^s}$ be the point of

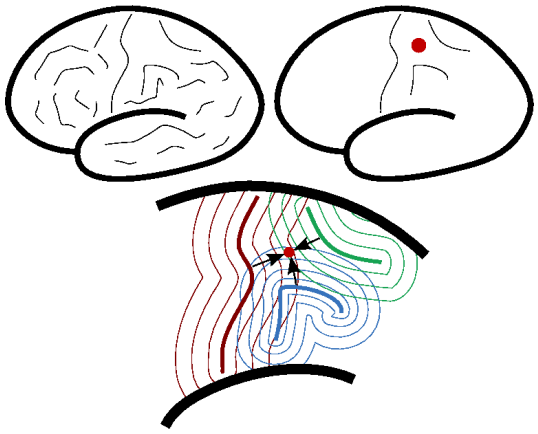


Figure 4.8: The position of the functional landmark (red circle) is defined by triangulation, as the minimum of criterion defined in Eq. (4.27).

the cortical surface of subject s that minimizes

$$\mathcal{J}^s(\gamma) = \sum_{j=1}^{3} \frac{(d(\gamma, \Gamma_i(j)) - \overline{d_{ij}^s})^2}{\widetilde{d_{ij}^s}^2}. \tag{4.27}$$

The steps performed to obtain $\widehat{\gamma_i^s}$ are summarized in Figs. 4.7 and 4.8. The minimization of the criterion (4.27) is performed by simple gradient descent. If the local referential is adequate, $\widehat{\gamma_i^s}$ should be very close to the actual functional landmark γ_i^s. We measure the discrepancy by the geodesic distance between the two points, $\mathcal{D}_i^s = d(\gamma_i^s, \widehat{\gamma_i^s})$. $\mathcal{D}_i^s$ represents the functional variability within the local referential based on the sulci.

To assess the quality of this method, we compare it with a volume-based approach: let $\overline{t_i^s} = \frac{\sum_{\sigma \in \mathcal{S}(i)-\{s\}} t_i^\sigma}{|\mathcal{S}(i)|-1}$ be the average position of the functional landmark i in MNI space, excluding subject s. We consider its projection on the cortical surface of the subject s: $\overline{\gamma_i^s} = \mathcal{P}^s(\overline{t_i^s})$. The quality of the standard normalization procedure is then defined as $\Delta_i^s = d(\gamma_i^s, \overline{\gamma_i^s})$. In the next section, we compare the distributions of $(\mathcal{D}_i^s)$ and (Δ_i^s) for several functional landmarks.

4.6.2 Results and Discussion

Three functional landmarks were investigated (sulci names are in accordance with Ono's atlas [153]):

i. The first landmark (represented by index (1) in Fig. 4.10, left) which corresponds to a reading activation, is located near the descending terminal ramus of the posterior ramus of the sylvian fissure, and referenced by the primary intermediate sulcus and the Superior Temporal Sulcus. This landmark could be identified in 16 subjects of the database.

ii. The second landmark (represented by index (2) in Fig. 4.10, left) corresponds to a computation task and is located in the IPS (parietal lobe) and referenced by the ascending terminal ramus of

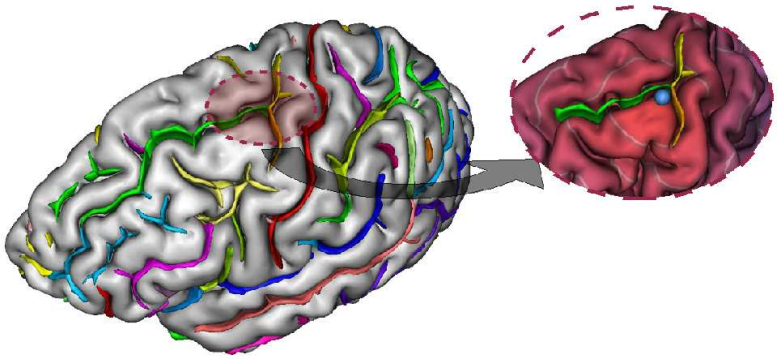


Figure 4.9: Within labelled sulci framework, the activation landmark of computation (blue ball on the picture) in the prefrontal lobe is localized near the intersection of the three sulci: superior precentral, marginal precentral and superior frontal.

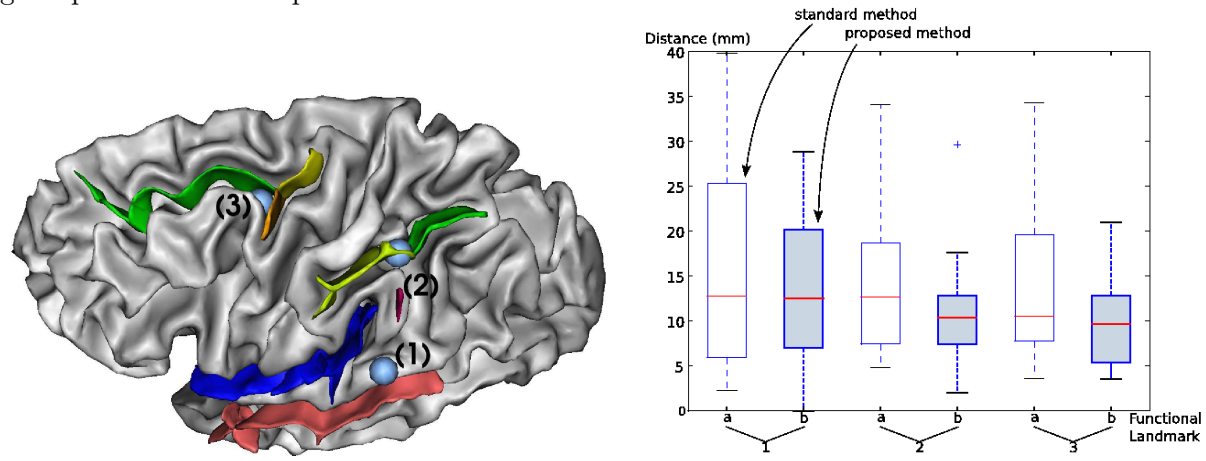


Figure 4.10: (left) The three balls represent the three functional landmarks. Each ball (landmark) is localized by the three nearest sulci. (right) Box-plot of the distances (Δ_i^s) (a) and ($\mathcal{D}_i^s$) (b) across subjects. These results correspond to three Functional Landmarks: (1) one related to a reading contrast, (2) one related to a subtraction contrast in parietal lobe and (3) the same contrast in the Frontal Eye Field.

the posterior ramus of the sylvian fissure, the inferior precentral sulcus and the main intraparietal sulcus. This landmark could be identified in 20 subjects of the database.

iii. The third landmark (represented by index (3) in Fig. 4.10, left) also corresponds to computation and is located in the Frontal Eye Field (FEF, PreCentral lobe). It is located at the intersection of three sulci: superior frontal sulcus, marginal precentral sulcus, superior precentral sulcus. This landmark was found in 22 subjects of the database.

The box-plots of the distribution of (Δ_i^s) and $(\mathcal{D}_i^s)$ are given in the first and second column of Fig. 4.10, right respectively. In the three cases the values of $(\mathcal{D}_i^s)$ are lower than those of (Δ_i^s) on average. Student t-test shows that the effect is almost significant in the first case ($p < 0.052$), and significant in the other cases ($p < 0.048$ and $p < 0.009$ respectively).

Our results show that a local referential can improve the accuracy of the position of some functional regions on the cortical surface. It is important to note that the detection of functional landmarks is done in the normalized space; thus our procedure to compare both referentials might be slightly biased in favor of the volume-based position definition technique. If this is the case, our results are conservative.

Our approach, as any surface-based brain mapping technique, is rather sensitive to anatomo-functional distortions that may remain after distortion corrections and anatomo-functional coregistration. This may explain why the distances (Δ_i^s) and $(\mathcal{D}_i^s)$ remain quite high in some -rare- cases (up to 30-40 mm on the cortical surface, see Fig. 4.10).

It is also crucial that a coherent labelling system is applied to all subjects. Errors in anatomical segmentation of the cortical surface can result in a wrong identification of some sulci and therefore a wrong definition of brain functional architecture. This bias may be significant for the whole database if the number of subjects is small.

It is important to note that the characterization of functional regions by their distances to several neighboring sulci is not the only approach for anatomical landmark-based brain mapping. In particular, for some brain regions just one or two sulci may be found near functional landmarks, in which case the present method cannot be applied. In this kind of case, e.g. close to the central sulcus or the superior temporal sulcus, it might be more adapted to use a local coordinate system along the sulcus.

In some brain regions, e.g. the frontal cortex, the identification of nearest sulci may be difficult. If the form of a sulcus is strongly variable between subjects, it can not be used as a reference. Moreover, several sulci are discontinuous, have several branches or a complex geometric form. Deciding which sulci are more reliable for functional brain mapping is an important topic for future research.

Importantly, sulcal variability is not necessarily a confound; it may also convey important information to describe different sub-populations in a group of subjects. Understanding how functional and anatomical regions jointly characterize such sub-populations is thus an important matter for the future.

Conclusion When local coordinate systems cannot be used, there remains the possibility of using more global coordinate systems [55, 26]. The comparison of local and global referential for the cortical

surface will be an important topic for future research. Using sulci-based referential remain in general more complex than traditional normalization techniques. On the other hand, it may be more helpful to describe and define precisely some brain regions. Finally, it might be possible that some functional regions cannot be reliably associated with anatomical landmarks. This is an interesting question, which deserves a more systematic assessment.

Chapter 5

Building a functional brain template (b): Anatomo-functional brain parcellations

In this chapter, we address the same question as in the previous chapter -how to build a template of the brain functional structures from an fMRI dataset- but using a *dense* approach instead of the sparse functional landmark point of view. This is the opportunity to revisit the concept of anatomo-functional multi-subject parcellation, with a more probabilistic point of view. Some contributions have been published in [217, 224].

5.1 Introduction

One of the main goals in neuroimaging is to match functional and anatomical regions across subjects, in order to better understand and characterize brain regions. Parcelling the cortical surface is a particularly intuitive approach, because it explicitly segments the cortex into modules with an anatomical and functional definition that is consistent at the population level.

Anatomical parcellation approaches typically define a coordinate system on the cortical surface that represents the position of the main anatomical features (identified sulci, curvature), and maps the cortex to a sphere [54, 27]. Then the cortical maps are further segmented into gyri, which are defined with respect to the main sulci of the brain [57], or directly in the sulcus-based coordinate system [26]. Although the nomenclature of the gyri may vary across publications and softwares, the advantage of these approaches is that they provide a relatively standard division of the cortex into regions; still this description is quite coarse (30 to 60 regions), which limits its usefulness for cortical surface mapping.

In order to more finely delineate cortical regions, functional information, obtained using fMRI data, provides further insights, and can easily be compared across subjects. A few approaches have been proposed to identify reproducible functional activity areas among a group of subjects. A method based on replicator dynamics was introduced [148] to subdivide a pre-defined cortical region; coherent subdivisions were found across subjects. In [203] a method based on spectral clustering performed the delineation of

homogeneous and connected regions which had similar position and functional activity across subjects. Finally, subdivisions of the cortex based on functional time courses only was proposed in [79], and essentially performed a clustering of functional signal. This provides a relatively coarse but prior free and consistent parcellation of the cortex based on fMRI data only. These approaches usually do not consider the anatomical information related to the data or simply reduce it to the three-dimensional coordinate system. By contrast, it is important for interpretation purposes to relate functional information to anatomical structures [14, 39]. Moreover, these approaches have not addressed the question of model selection. Specifically they do not perform a probabilistic comparison of different parcellation techniques. This point was partially addressed in [217], but the spatial model used there was only an approximation of the correct model.

In this section, we describe a novel approach that combines anatomical parcellation of the data into gyri and functional information that further refine the parcellation. The final parcellation procedure is cast into a probabilistic framework and the parameters of the model are identified using a modified EM algorithm, partly reminiscent of [238]. Finally, cross-validation procedures are used to optimize the number of components. Starting from 47 initial anatomical parcels (gyri), 254 and 229 anatomo-functional parcels are found in the left and right hemisphere respectively.

5.2 Model and methods

5.2.1 Data description and pre-processing

The database used here is an excerpt of $S = 25$ subjects of the database described in Sec. B.4.1. The Brainvisa package was used to segment different anatomical compartments (hemispheres, white matter, gray matter, cerebellum) from the T1 image of the brain of each subject, providing white and gray matter mesh, and segmenting the sulci [173] (see also Sec. 2.6). To label the sulci, automatic classification methods were used to extrapolate labels from the manually constructed database [164]. This sequence of processing steps was applied systematically to all available brains and the quality of resulting segmentation was checked. In particular, for all selected subjects the labeling was checked and in some cases corrected to solve inter-subject inconsistencies. Next, we used the method described in [27] to obtain a surface-based coordinate system that takes into account the main sulci, and the methods described in [26] to subdivide the cortical surface into gyri. The coordinate system maps the cortex onto a sphere, and will thus be denoted as (θ, ϕ).

For all subjects a standard preprocessing of fMRI data (distortion correction, correction of differences in slice timing, motion correction and anatomo-functional co-registration) was performed using the SPM5 software. Functional images were then projected onto the gray/white interface using the method described in [154]. Subsequently, on each functional dataset a GLM analysis was carried out to obtain task-related activity maps for different contrasts of experimental conditions. ($n_f = 4$) such contrasts are used in this paper: *i)* left versus right button presses, *ii)* sentence listening versus sentence reading, *iii)* computation versus reading *iv)* reading versus passive checkerboard viewing. A schematic flowchart of this processing sequence is shown in the upper part of Fig. 5.1.

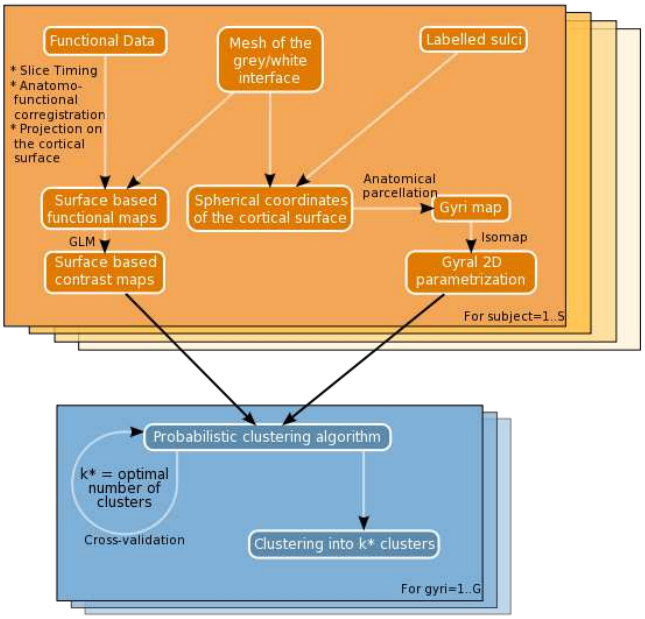


Figure 5.1: Flowchart of the method. The upper part shows the pre-processing steps of the method up to GLM and global parametrization of the cortical surface. (Bottom) sub-division into parcels is obtained by probabilistic clustering algorithm.

5.2.2 Local coordinates and distances

In this section we derive a coordinate system adapted to the representation of a gyrus which closely approximates the spherical coordinate system that describes cortical topography. Let (θ, ϕ) be this system of spherical coordinates. The distance between two nodes $p_1(\theta_1, \phi_1)$ and $p_2(\theta_2, \phi_2)$ is given by:

$$d_S(p_1, p_2) = \arccos\{\cos(\theta_1 - \theta_2) + (\cos(\phi_1 - \phi_2) - 1)\sin(\theta_1)\sin(\theta_2)\} \tag{5.1}$$

Since it is convenient to work neither with this non-Euclidean metric nor with three-dimensional coordinates, we create a new local coordinate system that respects the relation (5.1), by using a multi-dimensional scaling algorithm: Given a matrix $\mathcal{D}_S$ of squared distances between any pairs of points on the gyrus, we find the two-dimensional embedding $t = (t_1, t_2)$ that minimally distorts $\mathcal{D}_S$ in the least-square sense: $t = argmin_{\sqcup \in L^2(G)} \|\bar{\mathcal{D}}_S - \sqcup^T \sqcup\|^2$, where $\bar{\mathcal{D}}_S$ is the matrix $\mathcal{D}_S$ after centering, and $L^2(G)$ is the set of two-dimensional square-integrable functions on the gyrus. The solution is simply provided by a singular value decomposition of $\mathcal{D}_S$, because the gyrus has a trivial topology. Note that this is similar to the isomap algorithm. Finally the case of large patches is easily dealt with using the Nyström approach [237]. Importantly, this computation can be done across subjects, so that all the subjects are finally in the same local coordinate system, which is quasi isometric to the spherical coordinate system.

A probabilistic clustering algorithm is used to parcel the data into functionally homogeneous and spatially coherent regions.

5.2.3 Notations and model

In this section, the data are pooled from the set of subjects; the clustering is thus directly performed on the multi-subject data. Let $\{t_i\}_{i=1..I}$ be a set of coordinates that represent the position of cortical sites, and let $Y = \{y_i\}_{i=1..I}$ be n_f-dimensional vectors that represent the functional activity related to these sites. Let $K > 0$ be the number of components of the probabilistic model and let $(w_{ik})_{i=1..I,k=1..K}$ denote the probability that the site i belongs to component $k \in [1..K]$. Let $\Theta = (\theta_k)_{k=1..K}$ be the set of p parameters of the normal densities that represent the functional information related to the classes. The likelihood of the observation is thus

$$p(Y|\Theta) = \prod_{i=1}^{I} \left(\sum_{k=1}^{K} w_{ik} \mathcal{N}(y_i; \theta_k) \right), \tag{5.2}$$

which assumes conditional independence of the functional information given the class parameters, and would be a standard Gaussian Mixture Model is the loading w_{ik} were functions of k only. Here w_{ik} are a function of t_i and a set of two-dimensional coordinates $\mathcal{T} = (\tau_k)_{k=1..K}$ that describe the position of the clusters on the cortex, and a constant scalar γ:

$$w_{ik}(t_i, \mathcal{T}) = \frac{\exp(-\frac{\|t_i - \tau_k\|^2}{2\gamma^2})}{\sum_{l=1}^{K} \exp(-\frac{\|t_i - \tau_k\|^2}{2\gamma^2})} \tag{5.3}$$

5.2.4 Estimation of the model

We estimate the models parameters $(\Theta, \mathcal{T})$ in order to maximize the log-likelihood of the data.

$$\mathcal{L}(Y|\Theta, \mathcal{T}) = \sum_{i=1}^{I} \log \left(\sum_{k=1}^{K} w_{ik}(t_i, \mathcal{T}) \mathcal{N}(y_i; \theta_k) \right) \tag{5.4}$$

We use an alternate optimization scheme, in which Θ and $\mathcal{T}$ are optimized in turn. $max_{\Theta} p(Y|\Theta, \mathcal{T})$ is obtained through an EM algorithm while $max_{\mathcal{T}} p(Y|\Theta, \mathcal{T})$ is obtained through gradient descent. In fact, we update the E,M and position estimation (C) steps.

- E-step: let Z=$(z_i)_{i=1..I}$ be the allocation variables.

$$p(z_i = k) = \frac{w_{ik} \mathcal{N}(y_i; \theta_k)}{\sum_{l=1}^{K} w_{il} \mathcal{N}(y_i; \theta_l)} \tag{5.5}$$

- M-step: $\Theta = argmax_{\theta} \mathbb{E}_z \log p(Y, Z|\theta, \mathcal{T})$, which gives, for the mean and covariance parameters:

$$\mu_k = \frac{\sum_{i=1}^{I} p(z_i = k) y_i}{\sum_i p(z_i = k)} \tag{5.6}$$

$$\Sigma_k = \frac{\sum_{i=1}^{I} p(z_i = k)(y_i - \mu_k)^T (y_i - \mu_k)}{\sum_{i=1}^{I} p(z_i = k)} \tag{5.7}$$

- C-step:

$$\frac{\nabla \mathcal{L}}{\nabla \tau_k} = \frac{1}{\gamma^2} \sum_{i=1}^{I} (t_i - \tau_k) w_{ik} \left(\frac{\mathcal{N}(y_i; \theta_k)}{\sum_{l=1}^{K} w_{il} \mathcal{N}(y_i; \theta_l)} - 1 \right) \tag{5.8}$$

  Thus, we obtain simply

$$\tau_k = \frac{\sum_{i=1}^{I} t_i \omega_{ik}}{\sum_{i=1}^{I} \omega_{ik}} \tag{5.9}$$

  where $\omega_{ik} = w_{ik} \left(\frac{\mathcal{N}(y_i;\theta_k)}{\sum_{l=1}^{K} w_{il} \mathcal{N}(y_i;\theta_l)} - 1 \right)$ Note that a possibly more stable solution consists in applying the mean shift-like model [28]: $\tau_k \rightarrow \tau_k + \delta\tau_k$, where

$$\delta\tau_k = \frac{1}{I} \sum_{i=1}^{I} (t_i - \tau_k) \omega_{ik} \tag{5.10}$$

Alternating these three steps (5.5,5.6-5.7,5.10) is very effective in practice. In [224], a variant of this algorithm based on a Variational Bayes approach was used, but we found it to be more difficult to handle than the maximum likelihood approach described here.

5.2.5 Introducing Random Effects

Given that the data (Y, T) is sampled from multiple subjects, it seems adapted to introduce a random effect model in the signal modelling part. Let $(Y^s, T^s)_{s=1..S}$ be the data in each subject. In this work, we still assume that the parcel positions $\mathcal{T}$ are identical across subjects. We further develop the signal parameters $\Theta = (\theta_k)_{k=1..K} = (\mu_k, \Sigma_k)_{k=1..K}$ using a group and a subject level. The signal model is then: $\forall k \in \{1..K\}$, let (μ_k, Σ_k) be the population parameter, and $((\mu_k^s, \Sigma_k^s)_{s=1..S})$ be the individual parameters:

$$p(\mu_k^s | \mu_k, \Sigma_k) = \mathcal{N}(\mu_k^s; \mu_k, \Sigma_k) \tag{5.11}$$

$$p(y_i^s | z_i = k, \mu_k^s, \Sigma_k^s) = \mathcal{N}(y_i^s; \mu_k^s, \Sigma_k^s) \tag{5.12}$$

The parameters of the model become then $\Theta = (\mu_k, \Sigma_k, (\Sigma_k^s)_{s=1..S})_{k=1..K}$, while the log-likelihood of the data can be rewritten:

$$\mathcal{L}(Y|\Theta, \mathcal{T}) = \sum_{s=1}^{S} \sum_{i=1}^{I_s} \log \left(\sum_{k=1}^{K} w_{ik}(t_i^s, \mathcal{T}) \mathcal{N}(y_i^s; \mu_k, \Sigma_k + \Sigma_k^s) \right) \tag{5.13}$$

An illustration of the generative model is given in Fig. 5.2.

In the estimation procedure, the E and C steps are simply modified by substituting the terms $\mathcal{N}(y_i^s; \mu_k, \Sigma_k + \Sigma_k^s)$ for $\mathcal{N}(y_i^s; \theta_k)$. For the M step, we use an internal EM algorithm, where (μ_k^s) are the hidden variables:

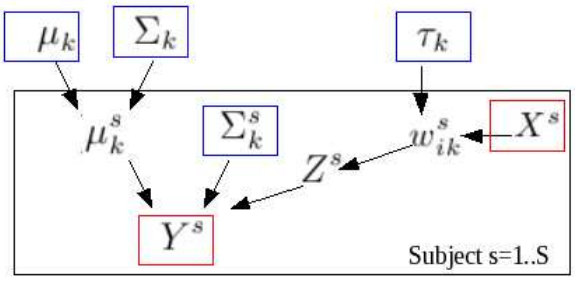


Figure 5.2: Generative model for the random-effects parcellation: in each subject $s \in 1..S$, the observed data Y_s, results from a spatial model, shared across subjects, that provides the probability w that each surface point belongs to a parcel, and a hierarchical effects model of the functional parameters, with both subject-specific (μ_k^s, Σ_k^s) and group-level (μ_k, Σ_k) mean and covariance parameters. The input data is outlined in red.

- E-step: estimation of $p(\mu_k^s)$

$$p(\mu_k^s|\mu_k, \Sigma_k, \Sigma_k^s, Y^s, Z^s) = \mathcal{N}(\mu_k^s; \Lambda_k^s \left[(\Sigma_k)^{-1}\mu_k + m_k^s(\Sigma_k^s)^{-1}n_k^s\right], \Lambda_k^s) \tag{5.14}$$

  where $m_k^s = \frac{\sum_{i=1}^{I_s} p(z_i=k)y_i}{\sum_{i=1}^{I_s} p(z_i=k)}$, $n_k^s = \sum_{i=1}^{I_s} p(z_i = k)$ and $\Lambda_k^s = \left[(\Sigma_k)^{-1} + n_k^s(\Sigma_k^s)^{-1}\right]^{-1}$

- M-step: $(\mu_k, \Sigma_k, \Sigma_k^s) = \text{argmax}\left(\mathbb{E}_{\mu_k^s} p(Y^s, \mu_k^s|\mu_k, \Sigma_k, \Sigma_k^s, Z^s)\right)$, which yields

$$\mu_k = \frac{1}{S}\sum_{s=1}^{S}\mu_k^s \tag{5.15}$$

$$\Sigma_k = \frac{1}{S}\sum_{s=1}^{S}(\mu_k^s - \mu_k)^T(\mu_k^s - \mu_k) \tag{5.16}$$

$$\Sigma_k^s = \frac{\sum_{i=1}^{I_s} p(z_i = k)(y_i - \mu_k^s)^T(y_i - \mu_k^s)}{\sum_{i=1}^{I_s} p(z_i = k)} \tag{5.17}$$

Furthermore, we use in this case a regularization procedure [65] in order to ensure that the different terms do not converge toward a degenerated case (e.g. null variance).

The important aspect with the random effects model is that it allows second level inference: let c be a certain contrast of experimental conditions; if we define second level statistics as

$$t_{RFX}(k) = \frac{c^T\mu_k}{\sqrt{c^T\Sigma_k c}}\sqrt{S-1} \tag{5.18}$$

Note that the distribution of this quantity cannot be assumed as known under the null hypothesis, but a corrected threshold can be derived through statistical resampling procedures (see Sec. 5.2.7).

Interestingly, this model and the ensuing test can be viewed as a generative alternative to the model proposed in Sec. 3.3. This is also a more elaborate point of view, and is clearly preferable, unless one is more interested in having crisp parcels in each individual.

5.2.6 Optimizing the model

Two main parameters of the model need to be optimized, namely K, which represents the number of components needed in the mixture model, and γ, which controls the strength of the spatial information [see Eq. (5.3)]. In order to have an interpretable procedure, and to avoid any confounding effect of the choice of the prior parameters, we proceed by cross-validation on the subjects data: For each subject $s \in \{1, .., S\}$, we estimate the parameters $(\mathcal{T}, \Theta)$ by pulling the data from all the subjects of the cohort but s, and then assess the goodness of fit of the resulting model on the data of s using Eq. (5.2). We do so for various choices of K and γ and select the best one according the maximal average cross-validated likelihood criterion. Let $K^\star$ be the best value for K.

Some shortcuts An important shortcut that we have used in our work was to assume that the covariance matrices Σ were diagonal, which makes the estimation as stable and straightforward as in the scalar case. This amounts to assuming the between-contrasts correlation of activity given the voxel class does not need to be explicitly modelled.

5.2.7 Random-effects (RFX) inference procedure

In order to control the specificity of the parcel-based statistical procedure, we need to know the distribution of the statistic (5.18) under the null hypothesis, i.e. when no activation is present. This cannot be done analytically, because the value of the statistic depends on the whole parcellation procedure. We tabulate the distribution of the null hypothesis by randomly swapping the sign of the data related to the tested contrast across subjects, and then recomputing the parcels and the associated statistic $\tilde{t}$.

Next, this procedure has to be carried out on the whole volume. After $R = 10^3$ randomizations, the maximal parcel-level statistic across gyri is tabulated:

$$\bar{t}_r = \max_{g \in 1..G} \max_{k \in \{1,..,K^\star(g)\}} \tilde{t}_k, \forall r \in \{1, .., R\} \tag{5.19}$$

and the surface-corrected threshold t_α for a specificity α ($\alpha = 0.05$ typically) is chosen as the $(1 - \alpha)$ quantile of the distribution of $(\bar{t}_r)$. Importantly, the interpretation of a parcel-based t-value being greater than t_α is that the probability of this value appearing in any parcel of any gyrus by chance is lower than α.

5.3 Experiments and Results

5.3.1 An Experiment on anatomo-function parcellations

To test the algorithm detailed in Sec. 5.2.4, we simulated different sets of five constant patches added with noise, within a square region, with a learning and a test set. γ and $K^\star$ were estimated using our method and the algorithm used in [217]. This was repeated 100 times. The bottom part of Fig. 5.3 shows *i)* a typical input of the simulation experiment, *ii)* a typical and approximately correct partition into five regions, *iii)* the values of $K^\star$ across 100 simulations, the corresponding values of $K^\star$ with the algorithm [217]. Our algorithm finds much more frequently the correct solution than the other, which typically overestimates K.

Using a maximum a posteriori partition of the data, the individual meshes can be labelled according to the spatial model. The resulting parcels are clearly interpretable through both the gyrus that contains them and the functional activity that characterizes the parcel.

The parcellation is also geometrically consistent across subjects: On the frontal-ventral gyrus (**B** in fig. 5.3) we identify the border between two large parcels clearly on the crest of the gyrus and the small third parcel is always on the same end of the gyrus. On the opercular section of the Broca's area (**C** fig. 5.3), four distinct parcels are positioned in the same way, and are characterized functionally in Fig. 5.3. For instance, sub-region 4 has a more asymmetric activity in the motor task, and more involved in the auditory task, while sub-region 2 is more involved in the computation task, and minimally involved in reading and listening sentences; sub-regions 1 and 3 have intermediate behaviour between these two cases. Altogether, the number of regions found in the model is 254 in the left cortex and 229 in the right cortex.

5.3.2 Discussion

We have shown that by combining anatomical and functional information on the cortical surface it is possible to identify the structure of the brain anatomo-functional partitioning on three hierarchical levels: Lobe $\Rightarrow$ Gyrus $\Rightarrow$ Functional parcel/region. With respect to previous anatomo-functional parcellation [148, 203, 217, 79], this clarifies the definition of the entities found by this approach. The last level division is provided by functional data, which means that the parcellation ultimately describes some local gradient of functional information within a given anatomical gyrus. This approach is intended to provide useful insights in the organisation of brain functional topography. A related observation is that in most of the cases the limits between parcels occur at the crest line of the gyri (typically on the top of the gyri). Deciding whether this is an intrinsic feature of functional anatomy or an effect of functional data projection deserves further investigation.

Ultimately, the optimal number of subdivisions per gyri is related to the across-subjects consistency of the functional features used to perform the subdivision, hence functional variability at the population-level. A compromise is thus found between a large number of small parcels that do not generalize well across subjects, and a coarse parcellation into a few sub-regions that is not sufficient to describe the available functional information. Based on 25 subjects, our algorithm provided a subdivision into

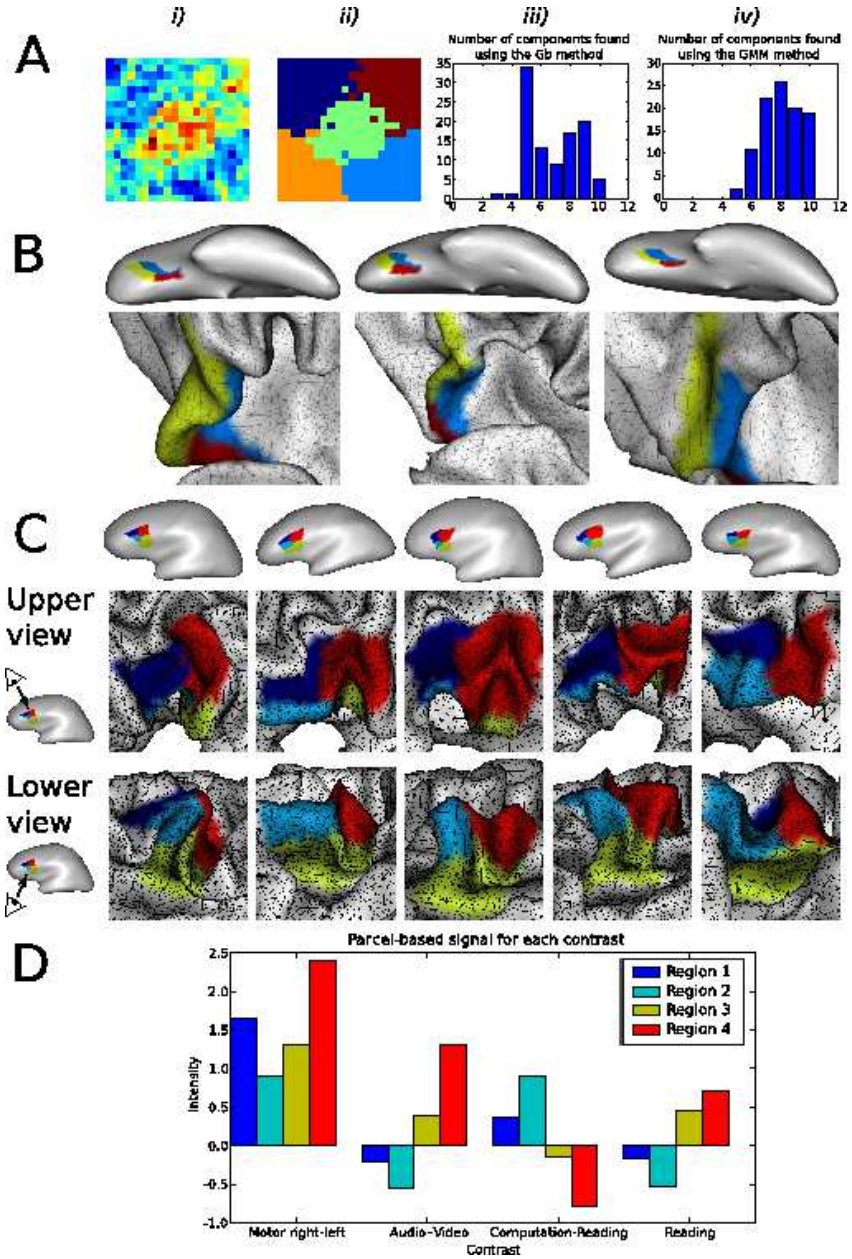


Figure 5.3: **A** Validation of our approach on simulated data: *i)* example of a simulated dataset *ii)* resulting parcellation *iii)* histogram of $K^\star$ for 100 random draws *iv)* idem with the method of [217]. **B** Sub-division on the frontal-ventral gyrus. Two larges parcels have their boundary on the crest of the gyrus. **C** Pattern of parcels on the opercular section of the Broca's area. In these five subjects, the relative position of the regions is remarkably stable. **D** Functional information related to the different parcels of **C**.

254 regions on the left hemisphere, and 229 on the right hemisphere. It is comparable, and somewhat finer than what was found using 3D parcellation [217] (about 500 regions, but for the entire brain). However, the present approach is based on a more appropriate probabilistic model, and the former had no anatomical model. Moreover, the number of regions clearly depends on the available functional information, and will probably be further refined in the future. This kind of result also depends strongly on the right correction of the EPI distortion and precise anatomo-functional co-registration. Misfits shift functional data in individual subjects and obviously might introduce significant bias. Interestingly, the left/right asymmetry in this number indicates that more information is available to segment brain regions in the left hemisphere, which coincides with the fact that reading is known to yield a much wider activation network on the left cortex in average. Future improvements of the method include a hierarchical model for the functional information related to the components in Eq. (5.2), which will better account for the inter-individual variability. Moreover, some simplifications in our estimation procedure may be improved in the future.

Accumulating knowledge on the spatial localization of functional activity in various experimental contexts is an important challenge for neuroimaging. In this work we have merged different functional and anatomical information into a coherent probabilistic framework and inferred an optimized parcellation of the cortex, which could thus be a basis for future anatomo-functional atlases.

5.3.3 Revisiting parcel-based random effects

Next, we revisit the parcel-based random effects framework described in 3.3, but within the probabilistic parcellation framework described in Sec. 5.2.5

Assessment of the PRFX statistic. Average left and right brain hemisphere meshes have been derived and are used for display. These average meshes are then parcelled using the maximum a posteriori label of each node given its position and the $\mathcal{T}$ parameters learnt by the algorithm. Our parcel-based inference (PRFX) is compared to more classical statistical procedures used on the cortical surface: *i)* the signal-level control procedure yields the threshold t_α so that the probability of the random effects statistic at a single node being greater than t_α is less than α (0.05 typically); it is obtained by tabulating the maximal t-value of any node under the null hypothesis by a permutation procedure (VRFX). *ii)* The cluster-level procedure that considers all the clusters (connected components) of nodes with a t values higher than a certain threshold (we take the threshold corresponding to $p < 0.01$, uncorrected) and tabulates the distribution of the largest cluster under the null hypothesis using the same permutation approach (CRFX), so that the risk of detecting one cluster larger than a given threshold size is less than α.

We concentrate on the contrast *sentence reading minus checkerboard viewing* that yields regions specifically involved in the reading task. Altogether, the parcellation outlines about 500 regions in each hemisphere. The images are given in Fig. 5.4.

In the left hemisphere, activation specific to the reading task is found along the Superior Temporal Sulcus (superior and middle temporal gyrus), in Broca's area, in the middle part of the pre-central gyrus,

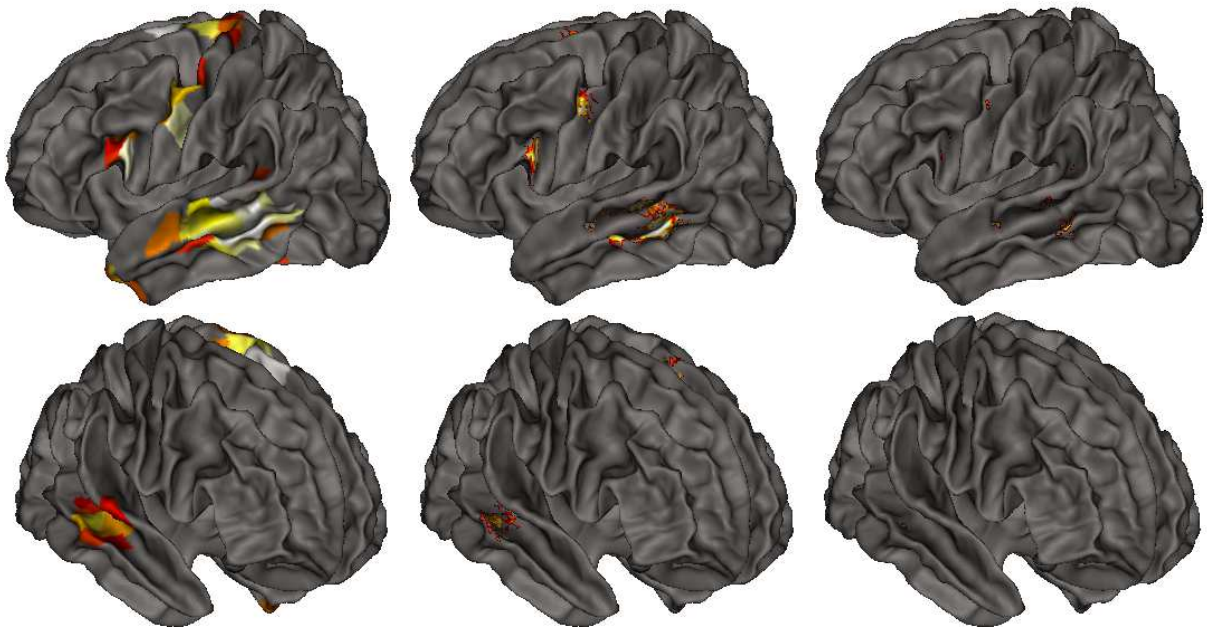


Figure 5.4: Outcome of the parcel-based (left), cluster-based (middle) and node-based (right) random effects analyses in the left(top) and right (bottom) hemisphere. All the maps are corrected at the $p < 0.05$ parcel-, cluster- and voxel-level, respectively.

and in the Superior Frontal Gyrus (Supplementary Motor Areas) by both PRFX and CRFX. Moreover, PRFX also detects significant activity in the temporal pole and the inferior temporal gyrus. VRFX detects tiny spots, barely visible in Fig. 5.4 in all these regions, except the superior Frontal region.

In the right hemisphere, all three methods detect some activity in the Superior Temporal Sulcus (middle temporal gyrus), but only PRFX detects activity in the right SMA.

Overall, the PRFX procedure is more sensitive than the other techniques. We also tested other contrasts and found that the PRFX procedure is at least as sensitive as the others. The VRFX procedure detects very few active nodes, but with a stronger control, in the sense that the null hypothesis is indeed rejected in each detected node. Nevertheless, it detects at least one active node in almost all the regions found with the other approaches.

The CRFX procedure detects more extended regions than VRFX, but rejects only the global null hypothesis in these clusters, i.e. it allows no statement on any detected node. The same applies for the Parcel-based random effect procedure: it allows the rejection of the null hypothesis in a certain portion of a pre-defined gyrus, not on all the nodes of the finally detected region.

5.3.4 Discussion

We have introduced the distinction between two sources of variance: *i)* the variance related to the spatial spread of the parcels, and thus simply represents the resolution which is chosen to analyse the data and *ii)* the between-subject variance, that represents the intrinsic functional variability between individuals, as well as potential spatial misfits. Besides, the introduction of the different variance components allows group-level inference, i.e. the computation of statistics that represent magnitude of the average effect in the population, when compared to between-subjects fMRI signal variability.

As can clearly be seen in the Section 5.3.3, the method compares very favorably in terms of sensitivity with random effects analyses performed on the coregistered and resampled meshes, both at the cluster- and vertex- level. Importantly, the method outlines extended regions or parcels, which potentially provides a less biased representation than a few mesh vertices: indeed, parcels naturally represent the uncertainty of the region position in the standard space. Another advantage of this approach is that the parcels outlined by the algorithm are confined within a certain brain gyrus, so that results are gyrus-specific.

The proposed model still requires the calibration of two parameters γ and K, which can be made automatically using model selection procedures (BIC, cross-validation), even though there is a price to pay in term of computation when cross-validation is used. When these parameters are fixed, the proposed model is not expensive computationally, so that permutation-based tests remain affordable. Moreover, the computation can be performed in parallel for the different gyri. Using a python implementation, we could run the whole framework in less than 24 hours.

A relatively straightforward extension of the present framework includes the adaptation to more complex populations, where behavioural or clinical score are available to characterize the between subject variability of subgroup structure in the observed population. This might be particularly useful to derive interpretable, i.e. few discriminative features to separate the populations. A related idea consists in using the present model to quantify the differences between individuals, generalizing procedures such as [103].

Chapter 6

Modelling and understanding inter-subject variability

Subject to subject variability is the major confound in one or two-sample random effect analyses, but more interestingly it may also be related to different genetic or behavioural features of the individuals and thus, become the object of further study [167].

In this section, we investigate different approaches to analyse the relationship between neuroimaging and other variables, such as genetic and behavioural information. Although such approaches are probably not sufficient to fully assess the *causality* of various factors on brain structure or activity and on behaviour, they are the only way to make sense of the observed variability.

We essentially describe two possibilities. The first one is based on a classical regression analysis, i.e. an extension of the group analysis methods described in chapter 3 to more general cases, where general factors are introduced to characterize the individuals. The second approach performs the inverse inference, i.e. aims at predicting some individual characteristics given the individual data. This yields a diagnostic approach. This is based on classification and regression techniques based on very high-dimensional data.

6.1 Regression analysis

This approach is meant to localize the regions where there is an impact of inter-subject characteristics on brain activity as measured by the BOLD signal. Regression analysis is based on the following data model (using the notations of Section 3)

$$Y_s = X_s\beta_s + \varepsilon_s, \varepsilon_s \propto \mathcal{N}(0, \sigma_s^2) \quad (6.1)$$

$$\beta_s = X_g\bar{\beta} + \eta_s, \quad (6.2)$$

where $(X_s)_{s=1..S}$ and X_g are the individual design matrices and the group level design matrix that encodes the inter-subject characteristics that have to be compared to neuroimaging data (typically a constant

factor, and several categorical ore real-valued factors). Contrasts of the parameter $\bar{\beta}$ can then be tested, i.e. they assess the impact of the contrasts of inter-individual variables on brain activity.

Before going further, its should be pointed out that Eq. (6.2) is a simplification of a more general model:

$$\beta_s = X_g \bar{\beta} \circ T_s + \eta_s, \tag{6.3}$$

where T_s is a spatial deformation that warps a certain template $\bar{\beta}$ to the data of subject s. A more ambitious formulation of this problem might use several template maps $\bar{\beta}$ to model different subgroups in the population, to take into account the structural variability between subgroups of subjects [6, 3]. In this work, we resort to the simplest possibility, by considering that one single template is defined, and that T_s is the identity.

Under the hypothesis that between-subject variability is normally distributed (i.e. $\eta \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_S)$ non-normal effects being part of the model), inference boils down to the set of procedures described in Chapter 3. Statistical inference has to face the same difficulties or shortcomings as those described in Chapter 3, i.e. the problem of handling multiple comparisons, and these are handled through the same solutions (use a family-wise error control or false discovery rate procedure, cluster-level thresholding). In many cases, the analysis is simplified by neglecting the uncertainty on the first-level analysis in Eq. (6.1), thus reducing the model to Eq. (6.2), where β_s is assumed to be directly observed. Given a certain contrast c on between subject characteristics, the decision statistic is:

$$F(v) = \frac{Tr(c\bar{\beta}(v)^T \bar{\beta}(v) c^T)}{\hat{\sigma}(v)^2 Tr(c^T (X_g^T X_g)^{-1} c)} = \frac{B(v)^T (P_{Xg} - P_R) B(v)}{B(v)^T (I - P_X) B(v)} \frac{S - R}{\text{rank}(c)}, \tag{6.4}$$

where $B = (\beta_s)_{s \in \{1..S\}}$ is the set of first-level parameters estimates, $P_{Xg} = X_g (X_g^T X_g)^{-1} X_g^T$, $P_R = X_g c (c^T X_g^T X_g c)^{-1} c^T X_g^T$ and $R = \text{rank}(X_g)$. In order to use more usual conventions, we will denote the set of individual maps Y instead of B to in the sequel of this chapter.

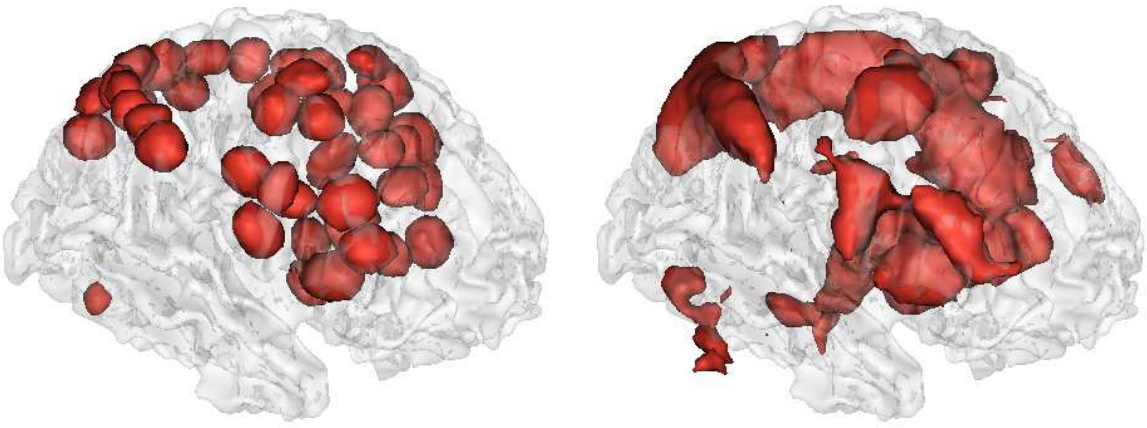
The main limitation of the approach is the the real effects are usually weak, and are further weakened by the various sources of variability in between-subjects analysis.

The main solution to these issues is to increase the size of cohorts, given that the sensitivity roughly increases as a function of $\sqrt{S}$.

What we propose here is to use the non-parametric (watershed-based) method described in section 4.2.3 to define regions of interest and perform the regression analyses in this regions (see [203, 206] for different versions of the present work).

6.1.1 Landmark-based analysis

The application of the Functional Landmark detection method discussed in Chapter 4 to a group of $S = 102$ subjects extracted from the dataset described in Sec. B.4.1 yielded $K = 45$ regions, with a corrected p-value of 0.05 of making one false detection. The confidence regions (CR) map is presented with the random effects map on these same subjects in Fig. 6.1. As is usual with large sample sizes, the



(a) (b)

Figure 6.1: Results of the group analysis that shows regions activated for a computation task across 102 subjects. (a) Confidence regions obtained with our approach, at $p < 0.05$; (b) Supra-threshold regions of the RFX map for this group of subjects, thresholded at the $p < 0.05$, corrected, voxel-level. The images are superimposed on a typical grey-white matter interface.

RFX map, thresholded at $p < 0.05$, corrected for multiple comparisons, shows very wide activated areas. This is is the result of blurring process inherent to the inter-subject variability.

The list of confidence regions readily indicates possible asymmetries in the spatial repartition of activations across subjects: activation in the Inferior Frontal cortex are found in the right hemisphere only, activations in the Supramarginal cortex, the Angular cortex are found in the right hemisphere only and activations in the Precentral regions are more systematic in the right hemisphere (4 regions) than in the left hemisphere (1 region).

Next, we computed the average signal per region per subject, and tried to characterize the population by unsupervised classification techniques. We call the average signal per region for each subject the *profile*. Based on simple Euclidean distance between profiles, we have performed an average-linkage agglomerative clustering of the population. The results are shown as a dendrogram of the subject's profiles in Fig. 6.2. In this case, it clearly shows that the population is mainly divided into one group of 97 subjects, and 5 isolated subjects. One can conclude that the population is rather unimodal, with a few outliers. A closer inspection of the outlier datasets reveals that four of them had no significant activations, and the last one had an odd pattern of activity, probably confounded by motion or another low-level artifact.

Finally, we regressed the voxel-based activity maps as well as the *profiles* against three regressors of interest defined in each subject: age, sex, and ability to perform a 3D mental rotation as in Eq. (6.2). Additionally, we also regressed the ROI coordinates in MNI space against these individual information.

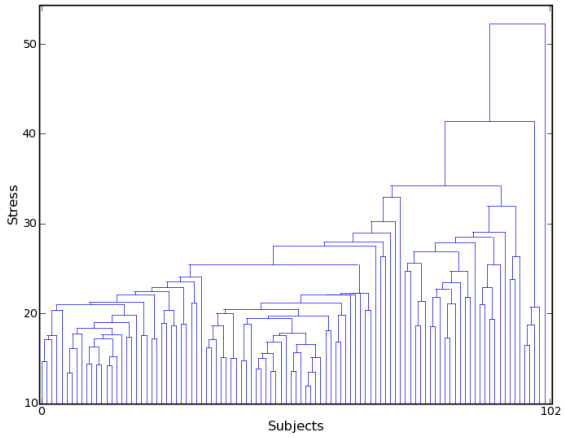


Figure 6.2: Unsupervised classification of the group of subjects, based on their *profile*. The dendrogram shows the organization of the population in terms of hierarchical clustering, based on an average linkage approach. The dendrogram shows that there is one main group, plus a few scattered subjects (on the right side).

We found no effect of age in either case[1].

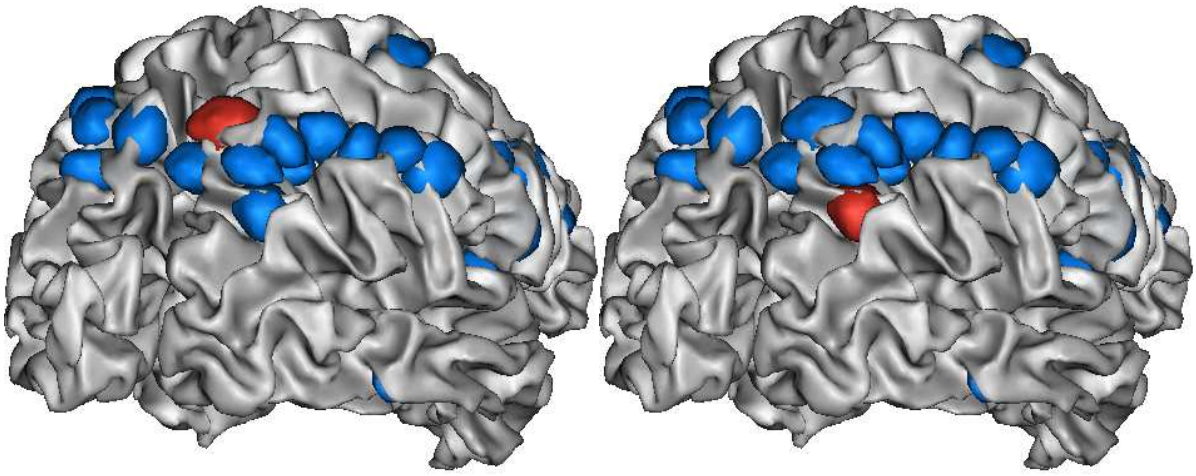
Concerning sex, we found in the voxel-based analysis a region where the magnitude of the activity is larger for males than for females [$z = 5.35, p < 0.05$, corrected at $(18, -68, 60)$mm]. As this place is on the posterior edge of the parietal lobe, and not in a significantly activated region, the interpretation of this result is quite unclear. The ROI-based analysis revealed that there was indeed a positive effect on one ROI [$z = 2.73, p < 0.01$, uncorrected at $(15, -68, 52)$mm], but moreover that there was a significant effect of the sex on the ROI coordinates in MNI space across subjects [$\chi^2_3 = 13.95, p < 0.01$, uncorrected at $(15, -68, 52)$mm], indicating that there might be some systematic shift effect between males and females. Importantly, no such effect can be observed using the voxel-based analysis.

Finally, we found an almost significant ($p < 0.06$, after correction for multiple comparisons) correlation for the 3D rotation score in the activity in a sub-region the left occipito-parietal boundary [$z = 4.03, p < 3.10^{-5}$, uncorrected at $(-33, -65, 45)$mm]. Using the ROI-based analysis, the result was also present, and significant [$z = 3.10, p < 0.05$, corrected at $(-27, -73, 36)$mm], while this -and only this- ROI had a significant effect of the score on its position across subjects [$\chi^2_3 = 34.5, p < 0.05$, corrected at $(-27, -73, 36)$mm]. The regions that exhibit significant correlation of their signal or their position with the subject's sex and score in the 3D task are shown in Fig. 6.3 (a) and (b) respectively.

In summary, while voxel-based and ROI-based analyses show similar effect (in the statistical sense) of the age, sex or 3D task performance on the fMRI signal, the ROI-based analysis clearly indicates that these differences could be spatial, thus possibly anatomical, rather than merely quantitative.

In the future, we plan to investigate more systematically the impact of the different spatial models proposed in this work(functional landmarks and parcels) on the group-level regression analysis described

[1]Note that the population is quite homogeneous, mean age=23.9 years and std=3.8 years. In this condition, the absence of an age effect at the group level is not surprising.



(a) (b)

Figure 6.3: ROI whose activity and position is significantly modulated by the sex of the subjects (a) or their ability to perform a 3D task (b). Voxel-based analyses yield similar regions, but ROI-based increases the significance in case (b) and enables us to study the effect of regressors of interest on the position of the ROIs.

here. This includes the interesting question of the optimal resolution of neuroimaging datasets for the inference of behavioural and genetics factors of interest.

6.2 Classification and feature selection

In general, the regression approach described in the previous section is not very sensitive, and not powerful for diagnostic. Moreover, it fails to extract interactions between regions that would be correlated to the inter-individual characteristics under study. A converse approach has thus been set, in which the targets embodied in the design matrix X_g have to be predicted from the neuroimaging data. This is basically the inverse of Eq. (6.2). Note that in the remainder of this chapter, we will deliberately omit first-level variance or uncertainty implied by Eq. (6.1).

Let x be the characteristic that has to be predicted from the neuroimaging data, typically a column of X_g using the formalism of Eq. (6.2). The question is thus to find a certain function Φ of the data Y (typically an image) that predicts x (typically a scalar or categorical value). Y is N_v-dimensional. Learning this mapping will be based on a S couples (x_s, Y_s).

$$x = \Phi(Y) + \epsilon \tag{6.5}$$

If x is categorical, Φ is a classifier: more generally Φ is a function. Standard choice for Φ is a Support

Vector Machine, Relevance Vector Machine, or Linear Discriminant Classifier (LDA) (see Section 6.2.2).

The difficulties of this study are related to the high dimension and weak contrast-to-noise ratio of fMRI data. Moreover, fMRI signals are uncalibrated, which makes quantitative inter-subject comparison difficult in general. Anatomical variability results in shifts in the location of corresponding regions; nevertheless, stereotactic normalization is the most widely used technique and therefore the one we choose for this study.

Given the large size of the data (about $N_v = 65000$ voxels at 3mm resolution), which is much larger than the number of samples (about $S = 200$ subjects), the classifiers tend to overfit, i.e. they use non-specific information to make predictions. To make sure that the classification rule will generalize to unobserved data, we need to base the classification on a restricted number of features/voxels.

6.2.1 Salience map and Features extraction

Feature selection is important for efficient classification, but it is also essential for interpretation purposes: it is generally expected that a few specific regions of the brain show a specific effect related to the information of interest. An adapted selection of relevant regions conveys thus an important information to neuroscientists.

Each feature (voxel-based information y) can be evaluated independently using the F test in Eq. (6.4), see also [86, p.11-13]. The significance of F can be assessed with a Fisher statistic with $(\text{rank}(c), S - R)$ degrees of freedom. Here we interpret it as a features salience map.

Typically, the features y are ranked according to their F values, and only the $N'_v < N_v$ features. Another possibility consists in retaining only its peaks, i.e. the local maxima of the F map. Assuming a spatial neighboring relationship in the dataset $\sim$, we can generalize it recursively, thus creating the $\sim^{(k)}$ neighboring relationship, i.e. the set of points that can be joined through a path of topological length $l \leq k$. v is a $k-$ local maxima iff

$$F(v) = max_{w \sim^{(k)} v} F(w) \tag{6.6}$$

This k parameter may be learnt from the data to maximize the correct classification rate. Note that such a procedure yields a reduction of the number of features of a factor greater than $f_k \propto k^3$ corresponding to the volume of a ball of radius k. The reduced $(N'_v < N_v)$-dimensional dataset is still denoted Y.

Variable selection in multiple regression In multiple regression, many features $Y(v), v \in [1, N'_v]$ are used conjointly to fit the target. A good subset $\mathcal{S}_v \subset [1, N'_v]$ of variables should be at the same time small in terms of cardinality N'_v, and fit the target correctly, as explained by the explained variance of the target, which is defined as

$$R^2 = 1 - \frac{\sum_{s=1}^{S} (x_s - \sum_{v \in \mathcal{S}_v} h_v Y_s(v))^2}{\sum_{s=1}^{S} (x_s - \bar{x})^2}, \tag{6.7}$$

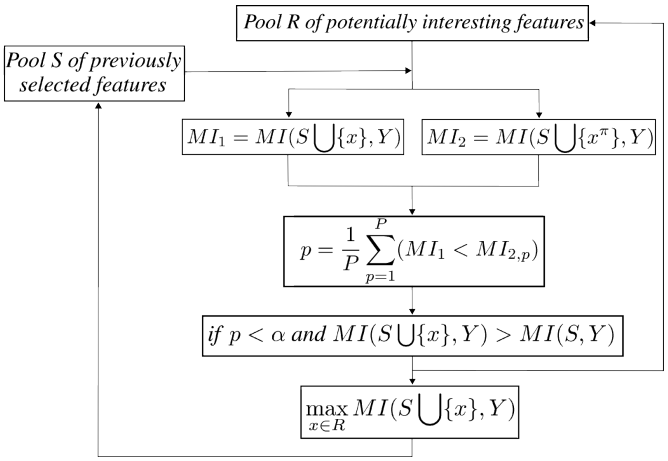


Figure 6.4: Flowchart of the MISF algorithm: a forward selection procedure adds at each step the most informative feature given the previously selected ones, while the backward procedure removes from Ω all the features which are not informative at this step.

where $\bar{x}$ is the sample mean of $(x_s)_{s \in [1,S]}$. R^2 can be readily translated to a F score.

$$F = \frac{R^2}{1 - R^2} \frac{S - N'_v - 1}{N'_v} \sim \mathcal{F}_{N'_v, S - N'_v - 1} \tag{6.8}$$

To determine the features included in the multivariate model, we used the simplest data-driven model building approach, i.e. forward selection. The most significant feature is recursively added to the model until no additional feature improves the Fscore [86, p.55]. When the target is discrete x, the F test associated with regression models is called Anova (Manova in the case of multiple regression).

Finally, let us notice that this procedure, and in particular the probabilistic assessment based on the F test is correct (and optimal) if the residuals are actually normally distributed ($\eta \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_S)$).

Mutual Information Mutual information-based feature selection is similar to Multiple regression, but makes no assumption on the distribution of the data. The Goodness of fit of the features is measured through mutual information $\mathcal{I}(x, Y(v)_{v \in \mathcal{S}_v})$ instead of the F-test.

$\mathcal{I}$ is estimated with the k-nearest-neighbour approach inspired from [114]; in [135] we have proposed a new approach in which is well suited when x takes discrete values and Y takes continuous values. The significance of the $\mathcal{I}$ value can be assessed through a non-parametric (permutation-based) approach. The selection of the features takes place within a feed-forward loop similar to [177], called Mutual Information-based feature selection (MISF) which is described in Fig. 6.4.

Firstly, let S and Ω be the sets of selected and candidate features: we start with $S = \emptyset$ and $\Omega = \{Y(v)\}, v = 1..N_v$ and the algorithm will stop when Ω is empty. This algorithm uses an hybrid stepwise selection. The forward strategy adds at each step the most informative feature given the previously selected ones. The backward strategy removes from Ω all the features which are not informative at this step : we indeed assume that those features will not be informative in the next steps. In order to select a feature, we compute at each step, for each dimension y in Ω, the value $MI_1 = MI(S \bigcup \{y\}, x)$, which

yields the amount of information about x present in S and y. Let y^π be a permutation of the values of y across samples, and let $MI_2 = MI(S \bigcup \{y^\pi\}, x)$. The distribution of MI_2 is computed by drawing randomly P permutations. We obtain the following approximate p-value : $p = \frac{1}{P} \sum_{k=1}^{P} (MI_1 < MI_{2,k})$. If this p-value is below a pre-defined threshold α, one can consider that this dimension is informative; otherwise we can remove it from Ω. In order to avoid redundancy of information, we also remove all the previously pre-selected features y for which $MI(S \bigcup \{y\}, x) < MI(S, x)$

Finally, we select the dimension with the highest value of MI_1 and keep the other pre-selected ones in Ω.

Comments on feature selection

- First, it should be noticed that features selection techniques can be cascaded. This may be advantageous for the sake of performance, but may require the problematic definition of many parameters that are not independent.

- Second, multivariate selection techniques (multiple regression, Manova, mutual information) are generally more powerful than univariate techniques (Anova, regression), because they assess the features conjointly (as the ensuing classifiers, that also take use the whole set of features). However, these approaches are computationally more demanding, and select generally very few features $N'_v \ll S \ll N_v$.

- The selection of local maxima is an intermediate case, given that the decision of including some features is not performed independently across sites. On the other hand, this technique does not evaluate the features conjointly, and may be suboptimal in this respect.

6.2.2 Standard classifiers

In this section, we briefly recall some basic definitions of the main classifier used in fMRI based diagnostic. This is of course not an original contribution.

Support Vector Machines SVM are the standard tool for large-margin classification [181]. Let $(x_l, Y_l)_{l \in [1,L]}$ be the available training dataset. The prediction of the regression-SVM model for a new instance Y is:

$$\Phi(Y) = \alpha_0 + \sum_{l=1}^{L} \alpha_l x_l < \psi(Y_l), \psi(Y) >, \tag{6.9}$$

where $<,>$ denotes some scalar product, and ψ is a non-linear embedding function. ψ does not have to be explicit: as long as the quantities $k(Y_l, Y) =< \psi(Y_l), \psi(Y) >$ have some positivity properties, the kernel k is admissible, and is in fact the only necessary object in the model. This is known as the *kernel trick* [182]. Note however that the standard bilinear form is the most efficient for fMRI datasets. For

classification purposes the training points $(x_l)_{l=1..L}$ belong to $\{-1,1\}^L$, and the classification response can be read out from Eq. (6.9) as $sign(\Phi(Y))$.

Training the SVM in a classification problem consists in estimating the $(\alpha_l)_{l\in[1,L]}$ so that

$$\hat{\alpha} = (\hat{\alpha}_l)_{l\in[1,L]} = \text{argmin}_\alpha \sum_{l=1}^{L} \alpha_l - \frac{1}{2}\sum_{l=1}^{L}\sum_{j=1}^{L} \alpha_l\alpha_j x_l x_j k(Y_l, Y_j) \tag{6.10}$$

under the constraints $0 < \alpha_l < C \; \forall l \in [1, L]$ and $\sum_{l=1}^{L} \alpha_l x_l = 0$, where $C > 0$ is a pre-defined constant that control the robustness of the SVM to overfit.

The training samples for which $\alpha_l > 0$ are known as support vectors; these are close to the boundary of systematically misclassified training samples. The SVM decision implied by Eq. (6.9) will not depend on the other training samples.

Choosing the parameter C is a difficult task, for which there is no closed-form results. In practice, grid search has to be performed to optimize the classification rate or regression score on the validation set. Fortunately, the sensitivity of the SVM with respect to C is not very strong in our experiments; we set $C = 1$ and consider that changing it moderately this parameter has little or no effect on the classifier performance.

Relevance Vector Machines (RVM) RVM are an alternative to Support Vector Machines that provide sometimes a more interpretable output [218]. Unlike SVM, RVM do not maximize the classification margin. Rather is is based on sparse Bayesian learning principle:

$$\Phi(Y) = \alpha_0 + \sum_{l=1}^{L} \alpha_l < \psi(Y_l), \psi(Y) >, \tag{6.11}$$

Note that this formulation is by construction compatible with the kernel trick. Training the RVM consists in estimating the weight α_l from some data. The RVM model includes an automatic relevance determination (ARD) prior on (α_l); the entails the introduction of an hyper-parameter vector ω:

$$p(\alpha|\omega) = \prod_{l=1}^{L} \mathcal{N}(\alpha_l; 0, \omega_l^{-1}) \tag{6.12}$$

During the training process, the evidence of the data, which is a function of ω, is optimized, which induces the shrinkage effect: many ω_l coefficients tend to infinity, thus imposing $\alpha_l = 0$.

For classification purpose, the output of the RVM is generally post-processed through a sigmoid function. The interest of RVMs is in classification is that the identification is based on class prototypes rather than boundary points as in SVM; the interpretation is thus clearer. However, in many practical cases, SVMs are more efficient in terms of prediction accuracy.

Elastic Net Elastic Net [247] is simply a sparse regression technique, that simply aims at obtaining the parameters $H = (h_v)_{v=1..V}$ such that

$$x = \sum_{v=1}^{V} h_v Y(v) + \epsilon \tag{6.13}$$

under some penalty on the h coefficients both in L_1 norm to enforce sparsity (this is known as the LASSO procedure [86]) and L^2 norm (to remove some pathological behaviour of the L_1 only regularization). Training Elastic net consists in solving the following problem

$$min_H \sum_{l=1}^{L}(x_l - \sum_{v=1}^{V} h_v Y_l(v))^2 + \gamma_1 \sum_{v=1}^{V} |h_v| + \gamma_2 \sum_{v=1}^{V} h_v^2, \tag{6.14}$$

where $\gamma_1 > 0$ and $\gamma_2 > 0$. The solution is found efficiently for all values of γ_1 through the least angle regression (LARS) algorithm [48]. Elastic Net can be viewed as a procedure that performs feature selection and classification (in fact, regression) simultaneously. Nevertheless, for computation reasons, it requires a limited number of input features (< 2000 typically).

Linear Discriminant Analysis (LDA) Let $c \in \{-1, 1\}$ be the two classes of the binary target, $Y^c = \{Y_{i|x_i=c}\}$ the class-specific datasets, $m_c = card(Y^c)$ and $\mu_c = \frac{1}{m_c}\sum_i Y_{i|x_i=c}$ their cardinals and means, $\mu = mean(X)$ the overall mean, $S_b = \sum_{c=0}^{1} m_c(\mu_c - \mu)^T(\mu_c - \mu)$ the between class covariance and S_t the total covariance. LDA searches the projection axes on which the data from different classes are best separated from each other while data from the same class are close to each other. Since the targets are binary, there is only one projection axis. The optimal projection axis is defined by Equation (6.15) and found in practice by the generalized eigenvalue problem (6.16):

$$a^* = \text{argmax}_a \frac{a^T S_b a}{a^T S_t a} \tag{6.15}$$

$$S_b a = \lambda S_t a \tag{6.16}$$

The problem is that the matrix S_t is rank-deficient as soon as the number is greater than the number of subjects $N'_v > S$, and is ill conditioned as soon as $N'_v \sim S$, making LDA an invalid method as soon as the number of features is of the order of the number of individuals.

Spectral Regression Discriminant Analysis (SRDA) SRDA is an algorithm for large scale discriminant analysis for classification problems [19] which generalizes Linear Discriminant Analysis (LDA) classification in high dimensional spaces. To get a stable and single solution of LDA, the between scatter matrix Σ_b and the total scatter matrix Σ_t are required to be non-singular, which may not hold for Σ_t. SRDA is a novel algorithm that overcomes these difficulties through the combination of spectral graph analysis and regularized regression.

Let W be the (S, S) matrix defined by $W_{i,j} = \frac{1}{m_c}$ if $x_i = x_j = c$ and $W_{i,j} = 0$ otherwise. Then W is re-arranged as a block-diagonal matrix. Let $\bar{Y}^c = Y^c - \mu_c,\ c \in \{-1, 1\}$ be the centered data matrix of class c and $\bar{Y} = [\bar{Y}^c]_{c\in\{-1,1\}}$. Σ_b and Σ_t can be rewritten: $\Sigma_b = \bar{Y}^T W \bar{Y}$ and $\Sigma_t = \bar{Y}^T \bar{Y}$. Equation (6.16) becomes:

$$\bar{Y}^T W \bar{Y} a = \lambda \bar{Y}^T \bar{Y} a \tag{6.17}$$

Let $\bar{x}$ be such that *i)* $\exists a : \bar{x} = \bar{Y}a$ *ii)* $W\bar{x} = \lambda\bar{x}$. Then a clearly solves (6.17). SRDA thus proceeds in two steps:

- Step 1: Solve the eigen-problem $W\bar{x} = \lambda\bar{x}$ such that $\bar{x} \in Im(\bar{Y})$. The solution is explicit in the case of a binary target: $\bar{x} = x - mean(x)$
- Step 2: Solve the regularized least squares problem (6.18) where $\alpha \geq 0$ controls the amounts of shrinkage. In our method, we learn this parameter to maximize the correct classification rate.

$$a = \text{argmin}_a \left((\bar{Y}a - \bar{x})^T(\bar{Y}a - \bar{x}) + \alpha a^T a\right) = (\bar{Y}^T\bar{Y} + \alpha I)^{-1}\bar{Y}^T\bar{x} \tag{6.18}$$

Prediction: Once we get the projection axis a, we embed a new sample x to one dimensional space $x \to z = x^T a$. If $z > 0$, the sample belongs to the first class, otherwise it belongs to the second class.

6.2.3 Cross-validation method

The (x, Y) dataset is separated into two datasets, a training dataset and a test dataset. The training dataset is used for an optimization procedure where a reduced number of features are extracted from Y and classifier parameters are optimized. For instance for LMSRDA method, we optimize both the α parameter of SRDA classifier and k in the Local Maxima features selection procedure. The classifier is trained on this dataset and evaluated on the test dataset, which is subject to the same dimension reduction as the training dataset. This is repeated for $P = 10$ different splits, i.e. features selection and parameter optimization are performed in each split. The ten splits of the dataset are balanced in order to have the same proportion of samples from each class.

6.3 A classification experiment

6.3.1 Experimental data

Simulated Data: Let Y be a set of S samples of a binary random variable. Y_ν are S randomly generated images with a zero-mean, unit-variance Gaussian distribution. Let Y_p be a binary image with $N_f = 100$ sparsely distributed non-zero elements. The surrogate data is defined as $Y = Y_\nu + \gamma Y_p x$, where γ controls the Contrast-to Noise Ratio (CNR) of the data, and is spatially smoothed to mimic fMRI correlation structure (kernel width is 1 voxel).

Real Data: The database includes about 200 subjects who performed the same experimental protocol described in detail in Sec. B.4.1. Five types of data have been collected from these subjects: Anatomical and functional MRI data (3T MRI scanner, SHFJ Orsay), demographic data, behavioral data from psychological tests and genetic data. Here we concentrate only on the fMRI, demographic and behavioral data. Standard preprocessing and first-level analysis have been performed on the data, which results is $S = 200$ fMRI maps of $N_v = 65000$ voxels for several functional contrasts.

From the database, we take seven couples (fMRI contrast, binary target). Let c_1 be the functional contrast that yields the areas involved in *computation versus sentences reading*, c_2 the contrast that yields the areas involved in *left hand button press versus right hand button press*, c_3 the contrast that yields the areas involved in *sentences reading versus viewing checkerboard* and c_4 the contrast that yields the areas involved in *motor tasks versus cognitive tasks*.

Let t_1 bet the demographic information *sex*, t_2 the behavioral information *lateralization* (all the subjects are right handed but a right handedness index characterizes subtle differences, and is thresholded to yield a binary target), t_3 a behavioral information which measures a possible dyslexia, t_4 a measure of reading fluency and t_5 an index of the difficulty of the subjects to identify right and left. The seven couples (fMRI contrast, binary target) are: $(c_1, t_1), (c_1, t_2), (c_2, t_3), (c_3, t_3), (c_3, t_4), (c_1, t_5), (c_4, t_5)$.

6.3.2 Results

We compare the performance of the LMSRDA method (SRDA classification based on Local Maxima of the salience map) to: (E_1) Linear SVM and Linear RVM classifiers based on local maxima of the features salience map; (E_2) SRDA, Linear SVM and Linear RVM classifiers based on features selected with the Local Maxima (LM) approach followed by multiple regression or mutual information-based selection; (E_3) SRDA, Linear SVM and Linear RVM classifiers based on features selected with Anova, Manova or MI only.

Results on simulated Data: We tested different values for $CNR = \{0.1, \ldots, 1.0\}$ but significant results only occur when $CNR \geq 0.4$. In Table 6.1, we can see that the LMSRDA method outperforms Linear SVM and Linear RVM classifiers combined to Local Maxima (E_1) and SRDA, Linear SVM and Linear RVM classifiers combined to Anova, Manova and MI features selection techniques (E_2) . We obtained similar results for E_3 (data not shown). The superiority of LMSRDA on the other methods is more significant for small CNR.

Results on real Data. The results of the method LMSRDA for the seven couples (contrast, target) are presented in Table 6.2. The number of subjects varies due to some missing information. The couple (c_3, t_4) has been created from the database but contains only 60 subjects, 30 very quick readers and 30 very slow readers. The results of experiment (E_1) are shown in Figure 6.5.a: For all the (contrast, target) couples, SRDA classifier outperforms Linear SVM and Linear RVM classifiers when based on local maxima. Part of the E_2 experiment is shown in Figure 6.5.b: SRDA, Linear SVM and Linear RVM classification performance for the seven datasets based on Local Maxima plus further Anova selection,

Rate of correct classification												
Classifiers	SRDA				Linear SVM				Linear RVM			
FS	LM	Anova	Manova	MI	LM	Anova	Manova	MI	LM	Anova	Manova	MI
$CNR = 0.4$	**57.7**	46.7	45.9	45.3	48.8	50.9	47.1	43.1	50.0	50.0	31.5	49.3
$CNR = 1.0$	**90.8**	82.9	62.9	79.9	89.2	79.8	66.9	75.2	50.0	50.0	23.2	32.9

Table 6.1: Rate of correct classification of LMSRDA, LM+Linear SVM, LM+Linear RVM, Anova+SRDA, Anova+Linear SVM and Anova+Linear RVM methods for a low and a high $CNR = \{0.4, 1.0\}$. For $CNR = 0.4$, the optimal couple of parameters $(\alpha, k) = (11, 0)$ and for $CNR = 1.0$, we use $(\alpha, k) = (76, 4)$.

LMSRDA: SRDA classifier based on Local Maxima							
(contrast, target)	(c_1, t_1)	(c_1, t_2)	(c_2, t_3)	(c_3, t_3)	(c_3, t_4)	(c_1, t_5)	(c_4, t_5)
S	181	180	179	179	60	180	180
% of correct classification	69.2	67.6	61.9	60.4	65.0	73.4	65.4
p-value	6.10^{-8}	1.10^{-6}	8.10^{-4}	2.10^{-3}	6.10^{-3}	5.10^{-11}	1.10^{-5}
N'_v	1548	113	231	32	920	1566	109

Table 6.2: Rate of correct classification, number of subjects, p-value and number of selected features of SRDA classifier based on the Local Maxima of the features salience map of seven (contrast, target) couples.

compared to LMSRDA classification performance. LMSRDA outperforms all the other techniques. We obtained similar results with Manova and MI features selection (data not shown). Only for the couple (c_3, t_4), SRDA classifier based on some Local Maxima selected with Manova achieves higher scores. Part of the E_3 experiment is shown in Figure 6.5.c: SRDA, Linear SVM and Linear RVM classification performance based on features selected with Anova compared to LMSRDA classification performance. LMSRDA outperforms the Linear SVM and RVM classifiers combined to Anova. We obtained similar results with Manova and MI features selection (data not shown).

Local Maxima of features salience map correspond to several isolated voxels. From this distributed localization of the selected and relevant information, regions of interest emerge. For instance in the case of the couple (c_1, t_5), the fronto-parietal network known to be involved in the spatial representation obtains the largest weights [29] (Figure (6.5.d)).

6.3.3 Discussion

Predicting relevant behavioral information and demographic information from neuroimaging data is a difficult problem, which is especially challenged by the small Contrast-to-Noise Ratio (CNR) of fMRI data and the spatial variability between subjects. Moreover, the population studied here is rather homogeneous, so that the differences under study may be quite subtle. Our results show that on this problem,

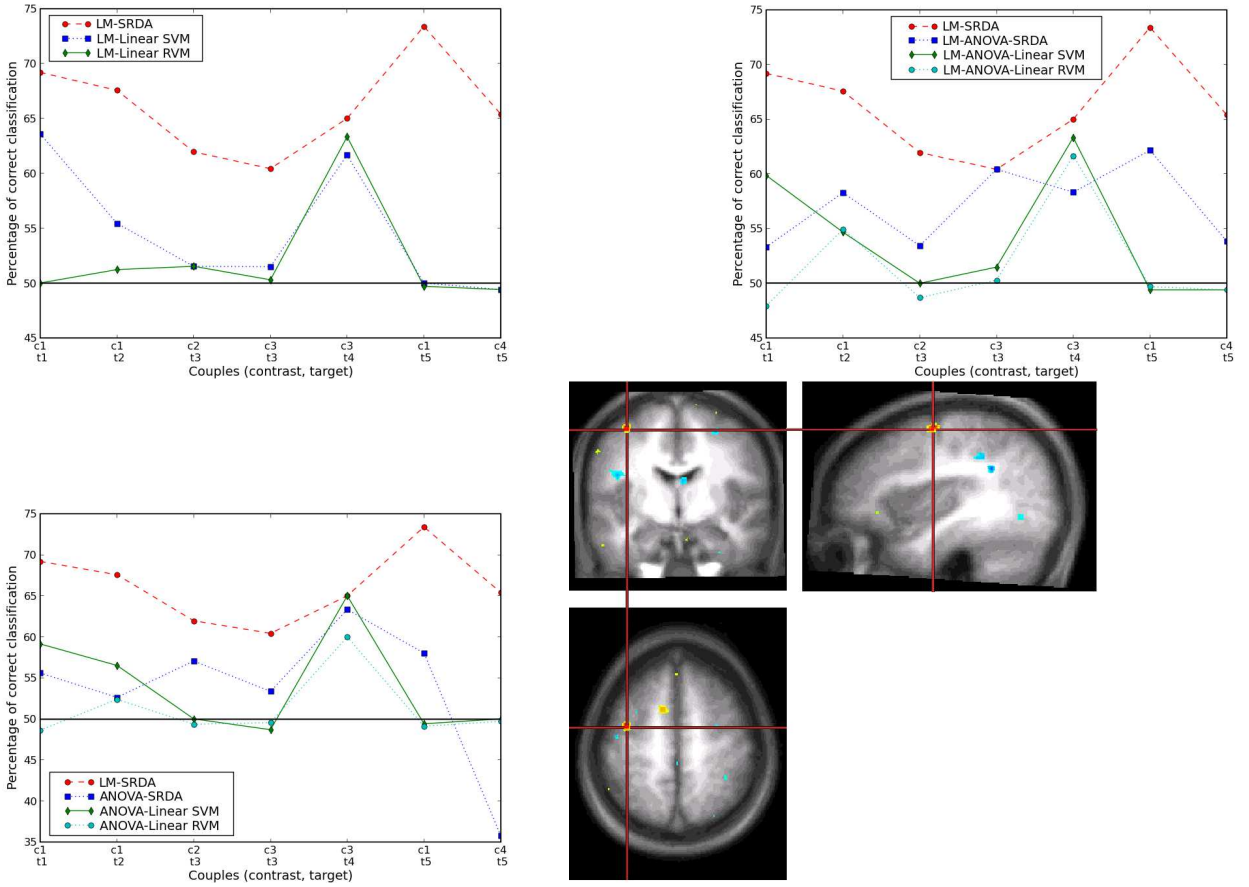


Figure 6.5: a) Results of E_1 experiment: Linear SVM (blue line) and Linear RVM (green line) classification performance for the seven couples (contrast, target) based on Local Maxima fMRI data compared to LMSRDA (red line) classification performance. b) Results of E_2 experiment: SRDA, Linear SVM/RVM classification performance for the seven couples (contrast, target) based on some Local Maxima plus Anova compared to LMSRDA (red line). c) Results of E_3 experiment: SRDA, Linear SVM and Linear RVM classification performance for the same data based on Anova selected features compared to LMSRDA (red line). d) Map of the final feature weights of the LMSRDA classifier for the (c_1, t_5) couple. The peak weight located at (-24, -18, 65)mm in MNI space belongs to the prefrontal cortex.

the combination of local maxima of the feature salience map and SRDA classifier achieves better results than other combinations of features selection (Anova, Manova and MI) and classifiers (Linear SVM and Linear RVM).

SRDA explicitly down-weights non-informative features and fully takes advantage of the sparsity of the salience map. This technique uses all the available samples for prediction while RVM and SVM techniques are based on a subset of the training set. SVM is associated to critical samples close to the decision boundary of the separated large-margin hyperplane (support vectors) and similarly RVM uses only prototypical samples of classes, called relevant vectors. These two approaches cannot explicitly down-weight noisy features, but only combinations thereof.

The selection of the local maxima of the feature salience (Anova statistic) map is a method that simply eliminates the redundancy of the features, and very unlikely discards informative features. It may also provide spurious features, but their influence is easily down-weighted by the SRDA classifier. By contrast, Anova selects redundant features which are not necessarily jointly informative. Manova- and MI-based forward selection techniques avoid feature redundancy but keep too few features. Results obtained with simulated data confirm that LMSRDA method outperforms significantly the other combinations of methods for smoothed data with small CNR, i.e. realistic conditions for our fMRI data.

The last advantage of LMSRDA is that it readily provides a weighting of the features which can be interpreted quite easily, and thus enables neuroscientists to understand the impact of different regions on the prediction of individual characteristics.

Extracting the local maxima of the feature salience map and using them in SRDA classifier yields better results than standard approaches in datasets with small CNR, which is typically the case for fMRI data. One of the reasons is that this technique fully exploits the image structure, which is not the case for general purpose heuristics. Further extensions of the present work include the study of genetic and non-binary target information.

Conclusion

Finding fMRI correlates of behavioural variables of interest in fMRI data is a difficult task, which is especially challenging when the number of available subjects is not very large. A compromise will have to be found between the paucity of features to be tested or included in a regression model, an the necessary of not biasing the classification toward priors, which may result in a systematical bias, and ultimately reduce the power of the analysis to discover significant correlations between brain activation and behavioural or genetics variables.

The next step in that respect to use multi-resolution approaches that may focus on particular regions, where information is encoded at a fine spatial case, without defining them a priori. We plan to contribute efficient clustering procedures to that end.

Chapter 7

Conclusion: Databasing the brain

In this conclusion, we discuss the question of data meta-analysis, which provides a new point of view on our contributions as well as a vision of our future work. This is written in non-technical terms and takes the form of a manifest.

7.1 Population-level analyses, databasing and knowledge management in neuroimaging

The common approach of neuroscientists using neuroimaging is to design an experiment based on their current knowledge given by the literature and their own previous experiments. Typically, one or several groups of 15-20 subjects undergo an experimental protocol (with behavioural and imaging data acquisition), and the knowledge obtained will be stored and shared with the community through publications. Re-analysis of old data is not very frequent except for methodological reasons. Data acquired for a specific experiment are rarely used for further experiment, except through the informal knowledge of the position of the activity found for some functional contrasts. With the conjoint advance of storage hardware, databasing systems and statistical methodology, it should be possible to take into account previously acquired data for designing experiments and analysing new data.

Current meta-analyses of multi-group or multi-protocol data are based on the position of activity in the standard coordinate system. The most advanced and ambitious resource in this field is the Brainmap project (see [63, 117, 223]) which contains about 1600 papers, 7300 experiments, 80 paradigm classes and 58000 reported brain locations. While a powerful framework, meta analyses using the standard coordinate system rely on *i)* an approximate warping of individual subject anatomy to a weakly contrasted template image, *ii)* the correspondence of the template with (most often) the Talairach neurosurgical atlas, and *iii)* a group analysis that uses a spatial smoothing filter of 8 to 12mm and the reporting of the local maxima of statistical maps. It is therefore an approximate system that stores group results but looses information at the subject level. The information extracted from a study is not generally stored in a formal system within a laboratory, so that the only existing repository is either the publication repository

system (e.g. Pubmed) or the scientist notes or memory. Meta analyses and accumulation of knowledge are weakened by the absence of IT system to represent, store, organize, and make inference on the information contained in several neuroimaging datasets. We observe that:

- Meta analyses are usually performed at a very abstract level (the regions cited in the papers that describe the activity, Brodmann areas or the coordinates in standard space), and are by nature coarse in the spatial domain.
- The standard brain referential system only approximately deals with the well-known brain variability, and introduces inherently a blurring of spatial descriptions of brain activity. Current MRI acquisition improvement in data spatial resolution may be completely lost.
- Knowledge on brain structure and function is rapidly increasing, but it is often partly subjective as it is based on the priors that neuroscientists have acquired along their experiments. Moreover, it is generally difficult to come back to the original data that have produced this knowledge. A system that would objectively present a synthetic inference through a collection of datasets is still lacking.
- The limited description of the functional paradigms precludes a deep understanding of the particular role of regions or the influence they can exert on each other in different experimental contexts.

7.2 Object-oriented and ontology-based analysis framework are necessary to model multi-modal data

Different designs have been proposed in order to better take into account inter-subject activity: structural models for fMRI (see chap. 4) parcellations (see chap. 5), clustering and subdivision of some Regions of Interest. These are promising approaches allowing for a description at the subject level but their results consisting in more complex objects still have to be accumulated in an appropriate system.

A crucial aspect is thus to find the most robust and reliable ways to describe individual activity patterns given the usually available data: functional and anatomical MRI, sometimes diffusion MRI. Using multi-modal data is indeed essential in order to provide a complete picture of brain organization [39]: anatomical MRI provides the definition of the basic anatomical compartments and interfaces but also 2D and 3D coordinate systems and some essential landmarks to divide the cortex into sub-regions (e.g. the gyri). FMRI provides important information for the subdivision of the broad cortical entities (lobes and gyri) into smaller scale modules (i.e. homogeneous functional regions), and also provides some landmark points. Diffusion-weighted MRI (Dw-MRI) provides useful complementary information: *i)* A local diffusion model, which is a fingerprint of the local white matter fiber orientation and therefore might be used to subdivide certain regions of the brain [8]. *ii)* A global diffusion model, obtained by using tractography algorithms in order to compute the main fibre tracts of the cortex (among others, see [50]), which provides a connectivity model of the cortex ([21, 95]). *iii)* Diffusion MRI has more recently been used in combination with fMRI to better describe the network structure of some regions defined functionally ([97, 9]).

In general, there exists currently no solution to handle systematically multi-modal and multi-subject neuroimaging data. More specifically, there is no attempt to produce consistent representations of both anatomical and functional data, at the individual and the group level.

From a bio-informatics point of view, several neuroimaging multi-modal databases and associated ontologies projects have been proposed. Among those, let us cite the BIRN initiative and its xml based ontology (Xcede, see `www.nbirn.net/research/function/xcede.shtm`), the Xnat (`www.xnat.org`) and other associated projects such as Brainscape, the fMRIDC database. However these projects and ontologies are not associated with dedicated inference tools, and thus have not been used so far to bring expertise in the representation of specific multi-modal datasets. Moreover, the particular issue of inter-subject variability has not been modelled systematically.

7.3 What we propose as a next step

We would like to introduce a new paradigm, in which statistical analysis of brain data is now embedded in a knowledge system that accumulates and organizes experimental evidence to refine a persistent brain model. In order to fulfil this objective, we need to address one main challenge (C1), which can be further elaborated in three challenges (C2-C4).

C1 (Main challenge): Inter-subject anatomical and functional variability prevents any straightforward brain averaging strategy.

Unlike most current approaches, we consider that image warping and averaging is not enough to build a comprehensive brain model. Instead, we plan to set a new approach, where image-based features commonly used in neuroimaging (contrast or t maps, individual regions of interest, segmented grey/white matter), should be considered as part of a certain model, which becomes the natural framework for inference on these data. This implies that not only feature averages should be considered, but also all the distribution at the population level for each of this features. Briefly, we propose to build a system to go from multi-subject data to brain models and from these models to individual information. To do so, we might i) take advantage of powerful segmentation, coregistration and pattern matching strategies to relate individual data to the template or group model ii) explicitly store a derived group model as a probability distribution of the individually extracted information, such as the brain activity in a given region of interest, its position, the size of a particular brain structure, the functional or anatomical connectivity between pairs of regions. In this way, a comprehensive model of the brain neuroimaging data will be built gradually without loosing spatial resolution. Specific subjects or groups of subjects could then be compared easily to already stored populations. This should help both cognitive neuroscientists concerned with general model of the brain and clinicians concerned with specific subjects for diagnosis.

C2: No observation modality provides a complete view on brain structure and function.

Indeed, identifying brain structures requires the joint use of multi-modal and multi-protocol data. What we propose is to build a system to store, augment, and interrogate heterogeneous data on a series of experiments. A key aspect is to allow the definition, labelling, storage of neuroimaging objects (e.g. regions, links between regions, networks) at the subject and group level, relating them to either imaging, behavioural or genetic data. We thus need to manage and organise large amounts of data using up-to-date database technologies but also make the use of previously acquired data easy when analysing a given dataset or when analysing several experiments conjointly. More specifically, it should be straightforward to define regions of interest or networks from previous experiments and use those in further analyses. The system should easily be extensible and simple to use. This implies the conception of a data management system with strong performances but a very intuitive and easy user (and software developer) interface.

C3: Traditional representations (maps, images) of brain activity have to be reconciled with objects-based representations.

To this end, we need to build a system to go from maps or 3D images to a hierarchical description with regions or brain structures, and their connectivity, and from brain structures to template maps. If neuroimaging methods are generally dealing with 3-4D images, neuroscientists are in fact much more often working with object oriented concepts. Typically, results of a functional neuroimaging experiment are described in terms of brain areas (”the anterior cingulate gyrus, the dorso lateral prefrontal, the FFA”) and the description is complemented by a position in a stereotactic atlas. Our objective here is to make the relation from maps to objects and from objects to maps easier and systematic when analysing data. Extracting features from images and constructing specific objects has to be facilitated, as well as the storage, organization and retrieval of those objects. One key aspect of this is that both representations (objects or maps) have a hierarchical structure, such that the level of description of the results can be adapted to the amount of information that is extracted from the data. A specific goal will be to refine the description whenever it is possible, e.g. to specify the localization of brain activity with respect to known brain landmarks. To a given level of the hierarchy at the object level will correspond a template map. Typically, a probabilistic template can be obtained for gyri or lobes or functional activity patterns from BOLD data seen at different resolutions.

C4: To date, there is no well-defined approach to explore effectively the variability of brain structures.

The main variations in brain structure, shape, size and organization may be explicitly accounted for through the use of multiple atlases and templates that need to be organized consistently. Understanding the variations is fundamental, and this will be made possible through the comparison of neuroimaging data with inter-subject genetic and behavioural information. To perform such comparisons, we also need to build an interface to help neuroscientists to easily and interactively put questions to the data entered in the databasing system. Those questions can be about brain activity (”Is this area likely not

activated by this experimental condition ?" "Are those two regions activating with similar strength ?"), brain connectivity, signal localization, and the modulation of these quantities by various inter-subject factors. To this aim, we have to link the available data and related knowledge (database, ontology and built objects) to a series of statistical methods to model distributions, allow robust inference on brain activity in a Bayesian framework, and incorporate classification methods and cross-validation techniques. In this perspective, it is useful to define a list of question types that neuroscientists are putting to the data. This list will include standard statistical comparisons and prediction but in the course of time will include open questions that have to resort to non supervised clustering techniques.

Appendix A

Reverse inference

In this appendix, we develop a question different from the multi-subject fMRI data analysis techniques discussed in the main part of this document: we focus on reverse inference in intra-subject data analysis. This part can be viewed as a generalization of traditional voxel-based statistics discussed in chapter 2, where inference is no longer performed on voxel-based information, but on more general objects (brain regions or brain maps). The question becomes: what information is carried by the brain data on a certain stimulation or cognitive state of the scanned subject ?

This chapter is organized as follows: in Sec. A.1, we discuss the general approach to solve this inference problem, part of which has been published in [135, 211] then we detail an original experiment that we have performed in Sec. A.2, and published as [211]. Finally, we give a few perspectives on this topic in Sec. A.3.

A.1 The classification approach to brain reading

A.1.1 Introduction

Validation of anatomo-functional knowledge produced from neuroimaging data is a difficult task. While statistical significance, reproducibility and multi-modal coherence are well-accepted proofs of consistency, neuroscientists lack a gold standard to assess the significance of their findings.

A possible way to solve this issue is to reason as follows: understanding a cognitive subsystem of the brain means that the stimulus-to-activation chain has been identified. More precisely, although the detailed mechanisms of neural and hemodynamic activation are not fully understood, we can expect that a controlled stimulus (e.g. a flashing checkerboard) will produce a known pattern of activation. When this holds, the processing chain can be inverted, leading to activation-to-stimulus inference. When possible, this inverse inference allows good performance characterization, since the results are expressed in terms of predicted versus true stimulus, in the well-known (and controlled) stimulus space.

This point of view has already been investigated in the case of motor experiments [35], mental imagery [150], counting/subitizing [166], the notion of object categories [87, 20, 31], the orientation of visual stimuli

[89, 100], and lie detection [34]. It has been popularized under the concept of *brain reading*. This novel approach in neuroimaging has been facilitated by the use of data classification techniques such as Linear Discriminant Analysis (LDA) [20] and more recently, Support Vector Machines (SVM) [31, 116] that can take functional images as input and classify them into categories (supervised classification). But in that case the activation-to-stimulus function remains implicit, i.e. it is embedded in a set of learning samples, each one being associated with a known stimulus.

Assume that we are given a set of brain activation images $\phi^1, .., \phi^n$ associated with a set of known stimuli $x^1, .., x^n$. In the *supervised classification* framework, a subset of the images $\phi^1, .., \phi^r, r < n$ associated with stimulus labels $x^1, .., x^r$ are used as a *learning set*. The learning algorithm (SVM typically) learns how to predict the stimulus label given the functional image. The test set, that consists of the remaining images $\phi^{r+1}, .., \phi^n$ is used to predict labels $\hat{x}^{r+1}, .., \hat{x}^n$. The performance of the classifier is given by the rate of correct predictions. The advantage of such an approach is that it works efficiently, without requiring prior knowledge on the precise activation mechanisms (functional architecture, connectivity, hemodynamic phenomena). In that sense it is *universal*. The disadvantage is that it is hard to diagnose a failure in the system. Moreover the interpretation is not straightforward (see e.g. [84]). Last, the ability to discriminate between activation patterns and to associate correct labels is restricted to the dataset used in the learning procedure.

A.1.2 Technical aspects

Performing reverse inference through the use of classification and regression techniques in high-dimensional spaces exactly imposes the same challenges as the question discussed in Section 6.2 in multi-subject studies: problem of overfit, and control of this overfit through cross-validation procedures, use of sparse classification/regression techniques to control the overfit, use of feature selection to optimize the performance in generalization (see below). The only difference is that inference does not suffer from anatomical variability or spatial mismatch, since the individual data is generally well coregistered. In this section, we make a short recall of this approach.

With the notations defined above, a given classifier proceeds by learning to discriminate between the images associated with a given label $\Phi_s = \{\phi^i | x^i = s\}$ and the other images. Each image is defined by its values on a number N_v of voxels.

Multivariate analysis and the risk of overfit

The discrimination function is efficient if it combines the information from many of these voxels: classifiers are thus inherently multivariate methods. Henceforth, we call *feature* a voxel-based information, *sample* an image of the learning or test set, and *label* the indicator of the stimulus associated with an image. Many features, i.e. non-specific voxels, do not carry any discriminant information, and thus they do not improve the classifier performance. When the proportion of such useless features increases, some of them are simply correlated with the associated label within the training set by chance, and their information cannot generalize to another set of samples (test samples). Those *fake positives* may dramatically harm

the classifier performance. This problem, known as the *curse of dimensionality*, is illustrated in Figure A.1. The number N_v of features increases along the horizontal axis, so that the ratio $\frac{n}{N_v}$ decreases: the training space becomes sparser. As shown by the blue line, the classifier rapidly reaches 100% of correct recognition on the training samples. In parallel the performance of the classifier on an independent test set of images increases until it reaches a maximum of 86% recognition for $N_v = 150$ as shown by the red line. Then it starts to decrease as N_v further increases.

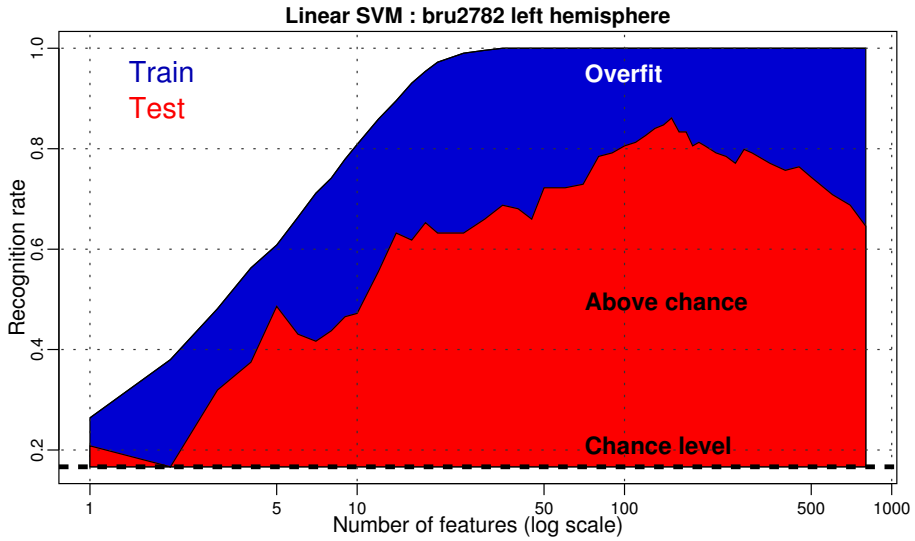


Figure A.1: Recognition rate (evaluated by a Leave-One-Out scheme) as a function of the number of input features (voxels) for the subject bru2782, left hemisphere (see the data in Section A.2). We measure the recognition rate of the classifier on the training set and on an independent test set when the number N_v of input features varies from 1 to 800.

To overcome this problem, we first apply a feature selection (described in Section A.1.2); the selected features are then given as input to a linear Support Vector Machines (SVM) classifier, presented in Section A.1.2.

Classification step one : feature (voxels) selection

Feature selection is a crucial step of classification: it improves the generalization power of a classifier and it is also useful to select a *small* subset of discriminant features, which is a requirement to interpret results in biological applications such as neuroscience or even genomics. Among feature selection methods [82], we choose a supervised univariate method based on the computation of an Anova (Analysis of Variance, which is a particular case of multiple regression performed on a categorical variable, see Eq. 6.4) F and p-values. This simple feature selection approach belongs to the family of methods called *filters*. Filters

are supervised univariate methods that rank features independently of the context of others features, according to their ability to separate the populations. Such methods are computationally efficient, which makes them tractable even on thousands of features as in our case. Unlike PCA, filters select the features in the original feature space which eases the interpretation in terms of discriminant information; moreover, these methods are less prone to overfitting than multivariate selection methods in general.

We first perform an Anova to compute to which extent the features are label-related ; this yields an F and a p-value. We select the features with two different methods for the control of false positives.

In the first case, we select significant features ($p < 0.05$) after a Bonferroni correction. When doing so, we have a strong control of the type I error (the number of false positives) selecting few but reliable voxels which minimize the risk of overfit. We use this very stringent method in difficult problems like mental imagery (Section A.2.4).

In the case of visual stimulation (trial-specific images of the domino experiment, see the results in Section A.2.4), the risk of overfit is lower. Hence we want to reduce the type II error in order to grab more discriminant features as input of the classifier. Thus we select significant features ($P < 0.1$) after a False Discovery Rate (FDR) [10] correction.

Classification step two : Linear SVM

Support Vector Machines (SVMs) [181] have recently been successfully used in fMRI applications [31, 116, 100]. Briefly speaking SVMs build their discriminant model as a linear combination of critical training samples. Those samples called *Support Vectors* (SVs) are either samples that lie close to the boundary of the two classes or samples that cannot be correctly classified. The success of SVMs on real data may be explained by their design which properly deals with few samples in high dimensional spaces: In a N_v-dimensional space with n samples, SVMs are fully parametrized with $n + 1$ parameters, while e.g. Linear Discriminant Analysis requires the estimation of $N_v(N_v + 3)/2$ parameters. This simple fact may explain the good behaviour of SVMs in high dimensional spaces. Another argument is that the SVM model enhances the *parsimony* of the discriminant model : SVMs not only attempt to perform a good classification of the training samples, as a perceptron algorithm does, but also constrain the discriminant model to be as simple as possible, i.e. a model in which the number of SVs is minimal. The choice of a linear SVM instead of a radial SVM [181] has been done after simple experiments conducted on one of the subjects (bru2782, left hemisphere, see the data in Section A.2), without any feature selection procedure. The linear SVM reaches 62% of correct classification while the radial SVM only reaches 22%. It is noticeable that the superiority of linear SVM has also been reported in [31]. The *cost* parameter ([181]) of the linear SVM has been set to 1 and the implementation comes from LIBSVM (`http://www.csie.ntu.edu.tw/~cjlin/libsvm`).

Validation

Validation is a simple but crucial point that must be carefully conducted in order to assess the quality of a discriminant model without any methodological bias. The classical way is to perform an out-of-sample

validation which consists of: (i) setting aside an independent set of subjects (the test set), (ii) learning on the remaining subjects (the training set) and (iii) testing the discriminant model on the test set. Cross-validation or bootstrap validation repeat the previous procedure and average the errors on test sets. The limit case of cross-validation is the Leave-One-Out procedure where only one subject is set aside. It should be noticed that feature selection is the first step of the discriminant model, thus it must be performed within the cross-validation loop, only on the training samples, and not as a pre-processing on all samples before the cross-validation.

A.1.3 The searchlight approach

Some approaches have been set to reconcile the classification approach with brain mapping: assuming that the discriminative information is supported by compact brain regions, it is possible to instantiate a classifier on small arbitrary regions of the brain volume. Such a region (typically a ball in the brain 3D space) can vary across the brain volume; one classification score will then be produced for each site where the small region has been set. This approach is known as searchlight [115], and finally yields a new kind of statistical map, which is the classification accuracy of a classifier instantiated in the neighborhood of the tested brain sites. This in turn can be compared to the chance level.

This technique is well suited to describe some relatively fine-grained organization of brain responses (although this does by no way improve the resolution of data analysis [111]), like the representation of categories in the infero-temporal cortex [115]. On the other hand, it is blind to widely distributed brain networks, which may carry some relevant information [87]. Technically, this technique nicely overcomes the curse of dimensionality by focusing on a limited number of voxels. Another important feature is that it allows group analysis by making inference on the classification maps obtained in multiple subjects.

For instance, we have performed a regression experiment in multiple subjects and derived an group-level prediction map. We use real data from an experience on quantity representations, i.e. numerotopy (data is a courtesy from Evely Eger). During the experiment, ten healthy volunteers (6 male and 4 female, mean age 21.2 ± 3.0 years) view dot patterns for different quantities of dots ($x = 2, 4, 6$ or 8), then with 8 repetitions of each stimulation, this yields a total of $m = 32$ images by subjects. Functional images were acquired on a 3 Tesla MR system with 12-channel head coil (Siemens Trio TIM) as T2* weighted echo-planar image (EPI) volumes using a high-resolution EPI-sequence. 26 oblique-transverse slices covering parietal and superior parts of frontal lobes were obtained in interleaved acquisition order with a TR of 2.5 s (FOV 192 mm, fat suppression, TE 30 ms, flip angle 78 , $1.5 \times 1.5 \times 1.5$ mm voxels). Standard pre-processing and the fit of the general linear model have been performed with the SPM5 software.

Trial-specific maps were kept for regression analysis. Searchlight was instantiated in each voxel location with 4-voxel radius balls. The performance was characterized by the ratio of explained variance ζ of the regression model (Elastic net or support vector regression [SVR], see Sec. 6.2.2) in a cross-validation loop:

$$\zeta(v) = \frac{var(x) - var(x - \hat{x}(v))}{var(x)}, \tag{A.1}$$

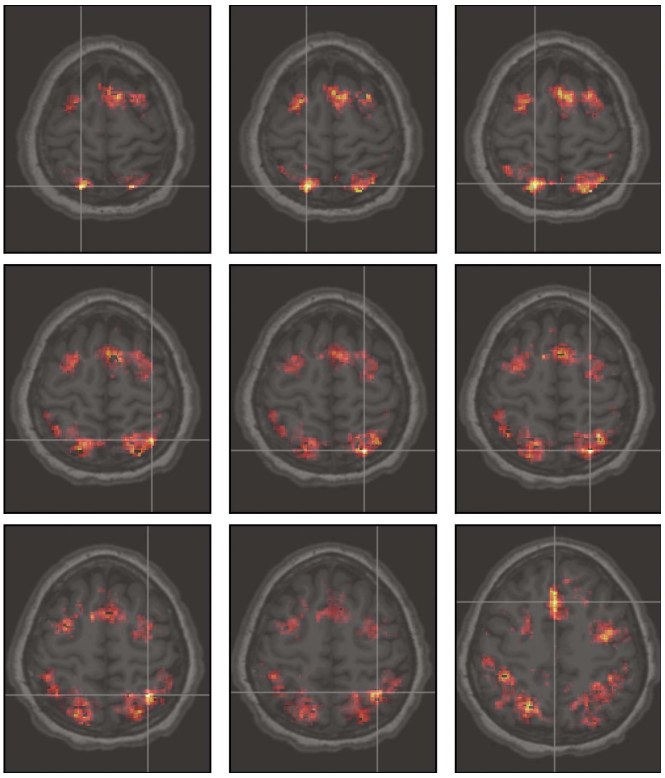


Figure A.2: Result of a multi-subject searchlight analysis in a numerotopy experiment. A group-level statistical score can be computed to assess the presence of discriminative activity around each voxel of the brain volume. This map is thresholded at the $p < 10^{-3}$ uncorrected, level. It clearly shows the fronto-parietal network, which is classically involved in number manipulation. This fact nicely illustrates how classification-based analysis can be reconciled with traditional brain mapping.

where $\hat{x}(v)$ is the prediction of x produced by elastic net or SVR, when trained on different acquisition sessions. This quantity is voxel-dependent.

The explained variance ratios were transformed to $p-values$ by using a permutation approach, then to z scores, and a between-subject random effect map was derived. It is illustrated in Fig. A.2, where cross-subjects significant regions are displayed on an average anatomical image. This clearly shows the involvement of the intra-parietal sulcus region in the internal representation of quantities, but also of regions of the frontal-eye field and the supplementary motor area.

This experiment can be seen as an assessment of brain regions for the representation quantities, and suggest the existence of a spatial coding of this representation: another code to be explicitly uncovered. It nicely illustrates how classification-based analysis can be reconciled with traditional brain mapping.

A.2 Making the link from brain activity to fMRI maps explicit: inverse retinotopy

A.2.1 Retinotopy of the Human visual cortex

There is at least one system in which the stimulus-to-activation coding is known explicitly: this is the case of retinotopy, where the spatial layout of an image is in the visual field also spatially encoded in the primary visual cortex [183]. The inverse problem consists in predicting the spatial layout of an activation pattern (stimulus) given a functional activation image. We address this problem with two kinds of analysis tools : Supervised classification (based on SVMs) and an explicit inversion of the stimulus-to-activation function (inverse retinotopy).

In this work, we study two different situations: a visual stimulation experiment, in which the subject passively views a sequence of stimuli chosen among a discrete set and a mental imagery experiment in which the subject is asked to imagine a self-selected pattern chosen among the presented stimuli.

It is well known that the human visual cortex is retinotopically organized, at least in early areas. Retinotopic mapping, based on a travelling wave paradigm, is a standard procedure in the fMRI literature (see e.g. [183, 40, 219, 51, 235, 221, 43, 244]). It is frequently performed in order to delineate the early visual areas (V1,V2,V3,V3a,VP), which can be characterized by a visual field sign (VFS) [183]. By contrast, we interpret here the retinotopic information as a forward mapping from the visual field to the visual cortex: We assume that there exists a transfer function that maps visual stimulation patterns to the primary visual cortex. In this work, the retinotopic data are used to estimate the transfer function. This kind of model has been suggested for V1 [220], and is supported by recent experiments [83]. In our setting, we take into account the receptive field structure [189] that characterizes the responsivity of cortical neurons to retinal stimulation. Let us note however that such a model ignores some parts of the response (non-linear and/or negative components) [186, 185].

In a recent paper [228], a direct estimation of the transfer function has been proposed based on randomized visual stimulation in an event-related design. However, this procedure is not as generic as the phase-encoded retinotopic experiments and it can only delineate pre-defined regions of the visual field.

A.2.2 Data acquisition and pre-processing

Subjects

Nine subjects participated to the study. One dataset was discarded due to poor fixation during the experiment (see below). This provided us with a total of 16 datasets, each hemisphere being analysed independently. The subjects gave written informed consent and the protocol was approved by the local ethics committee.

Stimuli

The experimental protocol consisted in three parts: (i) a retinotopic mapping of the subjects, (ii) a passive viewing experiment, in which the subjects were viewing so-called *domino* stimuli (iii) an imagery experiment, in which the subjects had to *imagine* one of the domino stimuli when prompted to. Next we describe the stimuli used in these three parts.

(i) The retinotopic experiment consisted in rotating wedges and expanding/contracting rings that flickered at a rate of 7.5 Hz. The checkerboard pattern was superimposed on a uniform grey field. The stimuli were projected onto a rear-projection viewing screen mounted within the scanner. Subjects were supine and viewed the display by means of a mirror placed above their eyes and housed in a custom-designed head piece. The duration of a complete stimulus movement was 32s, and it was repeated eight times for either condition. The wedge stimuli had one single lobe, with a maximal eccentricity of 10.5 degrees and an angular width of 40 degrees (see Figure A.3(a)). The ring had an eccentricity between 0.8 and 10.5 degrees. The size of the display, which matched the red circle in Fig A.3(b), was 21 degrees diameter. The subjects were instructed to fixate a central cross, and fixation was controlled using an eye-tracker system.

(ii) In the *domino* experiment, two grids, situated on the left and right parts of the visual field, and a central fixation cross were presented to the subjects. The grid was surrounded by a disk of 9.5 degrees diameter. Every 8s, a flickering pattern appeared in several sectors of the grid. These patterns belonged to a set of 6 possible shapes (see Figure A.3(c)). The patterns were presented simultaneously in the left and right visual field for a total of 36 combinations which were all presented once per fMRI run, in a randomized order. Each subject performed four sessions of this domino experiment.

(iii) Then the subjects were asked to choose one of the six patterns. During the last session, the subjects viewed the same grid, but without any pattern presentation. The central cross was changed to a small left/right arrow of the same size (0.8 degrees), prompting the subject to imagine the selected pattern on the left or right side (see Figure A.3(d)). The arrow occurred four seconds every ten seconds interval and appeared a total of 36 times: 18 times on the left side, 18 times on the right side. Once the experiment was finished, the subject reported which pattern he or she had chosen for the imagery experiment.

During all the scanning sessions, the subjects were instructed to fixate the center of the screen. Eye movement were registered with an ISCAN eye-tracker system, in order to ensure that fixation was maintained. One subject did not fixate adequately, and the dataset was eliminated from the analysis.

Acquisition parameters, pre-processing

Functional images were acquired on a 3T Bruker scanner using an EPI sequence (TR = $2000ms$, TE = $40ms$, matrix size=64×64, FOV = $19.2cm \times 19.2cm$). Each volume consisted of 35 $3mm$-thick axial slices without gap. The first four functional scans were discarded in order to allow the MR signal to reach steady state. Anatomical T1 images were acquired on the same scanner, with a spatial resolution of $1 \times 1 \times 1.2\ mm^3$.

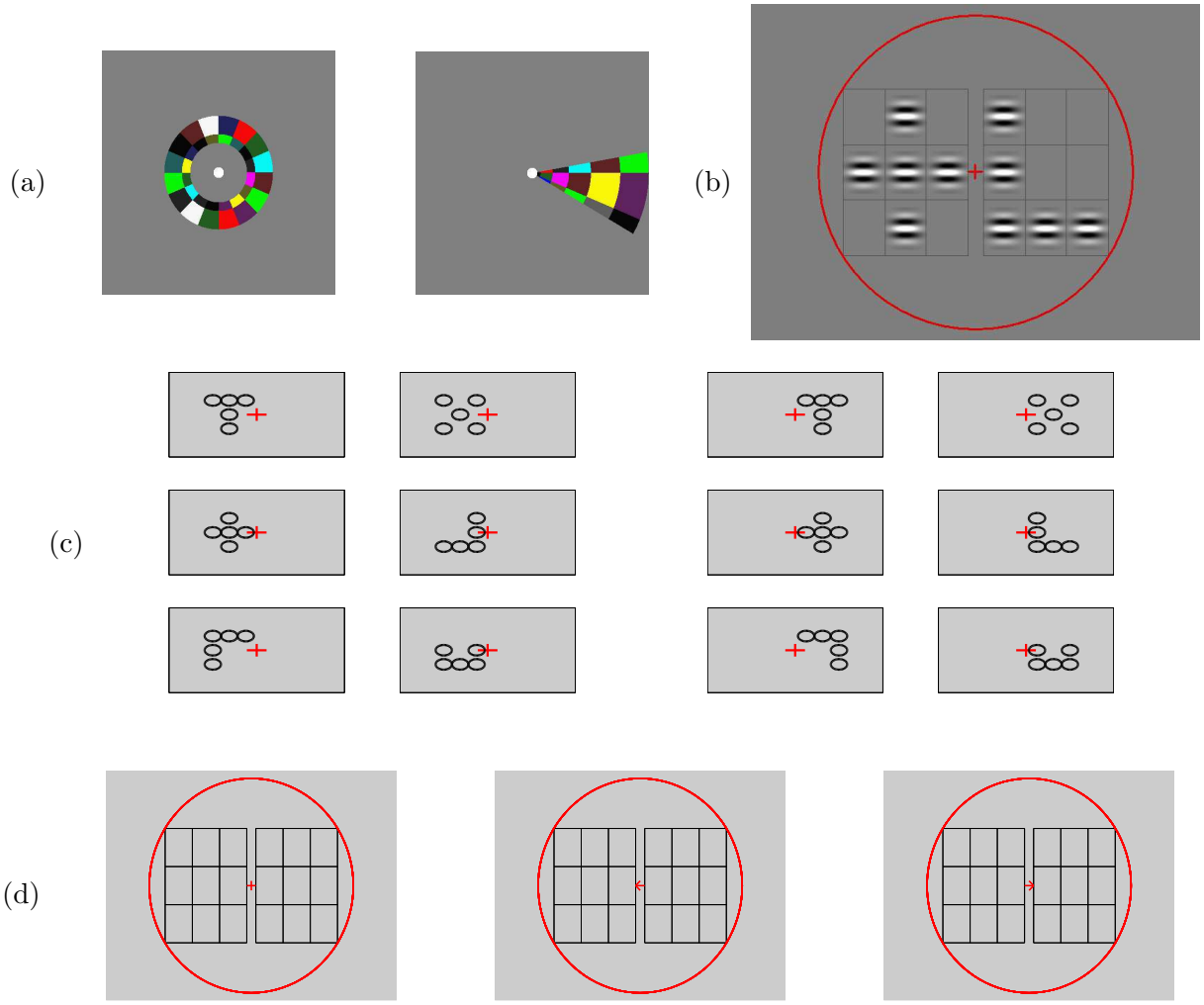


Figure A.3: Visual stimuli used in our experiments. (a) First the subject was involved in a classical retinotopic mapping experiment, in which he viewed flickering rotating wedges and expanding/contracting rings. (b) In the domino experiment, the subject viewed groups of quickly rotating Gabor filters in an event-related design. These disks appeared simultaneously on the left and right side of the visual field, superimposed on a low-contrast grid and a fixation cross.(c) There were 6 different patterns in each hemifield. (d) In a last session, the subject was presented with the same grid. When the central fixation cross (left) became a right arrow (middle) or a left arrow (right), the subject had to imagine one of the six patterns presented previously, either in the left or right hemifield.

Motion estimation was performed on each dataset using SPM2 software (see e.g. [5]). The anatomical images were then normalized to the MNI template of the SPM2 software, and resampled. The interpolation of the functional data took into account motion estimates, so that the normalized images were also realigned. Resolution after interpolation was $2 \times 2 \times 2\ mm^3$. No other pre-processing was performed.

First-level analysis of the data

All datasets were analyzed using the General Linear Model (GLM) implemented in the SPM2 software: retinotopic sessions were analyzed using sinusoidal regressors at the stimulus frequency; the other sessions were analyzed by convolving the activation onset vectors with a standard hemodynamic response; standard high pass filtering ($hf_{cut} = 80s$) and AR(1) noise whitening were used. Activation maps were produced for each experiment. In the retinotopic mapping experiment, these maps show regions with significant activity at the stimulus frequency, hence retinotopic regions. By contrast, the statistical images resulting from the analyses of the domino and imagery experiments were associated with occurrences of the stimuli. They could thus be readily interpreted as stimulus-induced activation patterns.

The parameter maps of the retinotopic experiments were further processed as indicated in [183] in order to yield polar and eccentricity maps. False positives were discarded by retaining only the main connected component of supra-threshold voxels, after thresholding at $p < 10^{-3}$ uncorrected. This systematically corresponded to a symmetric occipital cluster. This yielded $N_v (\sim 10000 - 15000)$ voxels, according to the subject. The retinotopic regions were divided into left and right hemispheres using the segmentation of the anatomical image by the Brainvisa analysis pipeline [174].

The domino experiment was analyzed on a trial-by-trial basis, yielding trial-specific *(ts)* activation maps. Condition-specific *(cs)* contrasts and activation maps were also estimated. For further processing, both *cs* and *ts* maps were masked by the retinotopic regions. Similarly, the imagery experiment was analyzed in order to yield trial- and condition-specific images. All maps were masked as the images of the domino experiment.

A.2.3 Explicit solution of the inverse problem

The explicit reconstruction of images in the visual field requires the solution of a forward problem (definition of a mapping from retina to cortical activity) and then the solution of an inverse problem (visual image associated with a given activation image). The global setting is described in Figure A.4.

Solution of the forward problem

We define a visual image as a function ρ that associates an activity value $\rho(p)$ with any point p on the retina R. In practice the retina will be discretized on a grid of size $\mathcal{P}$. In our setting, $\mathcal{P}$ is 100×100 to balance the competing demands of computational efficiency and resolution. An activation image is a function ϕ that associates an activation value $\phi(v)$ with any voxel v of the brain volume. In practice the brain volume is restricted to a set of V voxels that have retinotopically specific responses.

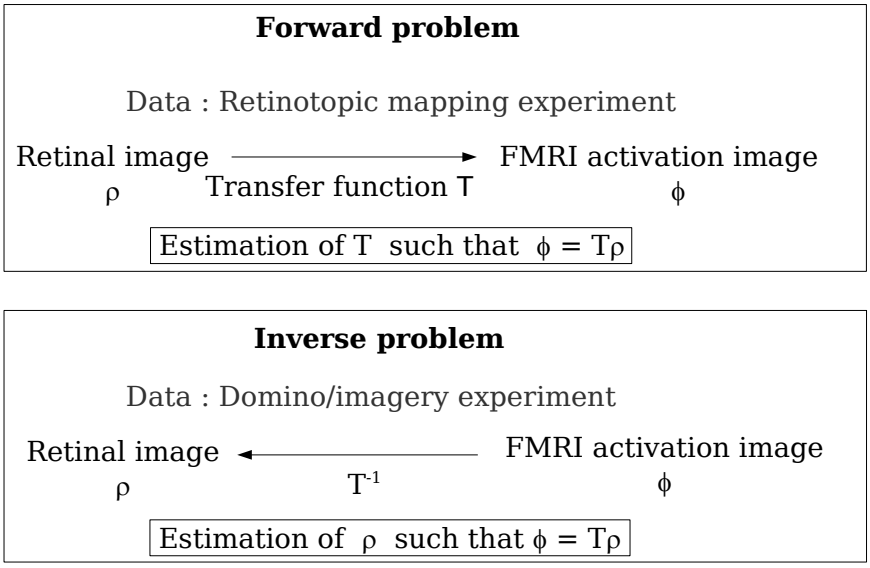


Figure A.4: Illustration of the forward and inverse problem in an *inverse retinotopy* framework. The forward problem consists in estimating explicitly a transfer operator that maps a stimulus into an activation image. The inverse problem consists in predicting the stimulus associated with an activation image, given the transfer operator.

We use the following generative model : the visual stimulation ρ is mapped to a functional image ϕ through a transfer operator T, i.e., $\phi = T(\rho)$. The forward problem consists in estimating T. Since the travelling wave paradigm used in the retinotopic mapping experiment performs a complete sweep of the visual field, we use the corresponding data to estimate T: let $\rho^1, .., \rho^n$ be the visual images of the retinotopic mapping paradigm, and $\phi^1, .., \phi^n$ the associated functional images, we search T such that

$$\phi^i = T(\rho^i) + \epsilon^i, \ \forall \ i \in [1..n] \tag{A.2}$$

where ϵ^i is an additive (measurement) noise that models possible mismatch. This noise will be assumed to be independently identically distributed Gaussian and centered.

A priori T is a -possibly nonlinear- operator from $\mathbb{R}^{\mathcal{P}}$ to $\mathbb{R}^{N_v}$. Given that the sizes $\mathcal{P}$ and N_V are well above 10^3, the direct estimation of T from Eq. (A.2) is impossible. Thus, we first assume that T is linear, which is equivalent to a spatial superposition principle of visual activations; this hypothesis is supported by recent experiments [83]. If T is linear, it is fully specified by its behavior on spatial Dirac functions ($\delta_p, \ p \in [1..\mathcal{P}]$) in the input space. At this point, we use physiological prior knowledge to estimate T. For each voxel $v \in [1..V]$, we assume that there exist a point p_v of the retina, a positive real

number (radius) λ_v and a real number γ_v (gain) so that

$$[T\delta_p](v) = \gamma_v \exp\left(-\frac{\|p-p_v\|^2}{2\lambda_v^2}\right) + \epsilon(v) \tag{A.3}$$

where δ_p is a Dirac function on p, and $[T\delta_p](v)$ is the associated functional image evaluated at voxel v. This simply means that voxel v is associated with a receptive field, i.e. a Gaussian kernel centered on p_v, with width λ_v, and that the gain of the filter is γ_v. The receptive fields are assumed to be isotropic. This model is illustrated in Figure A.5.

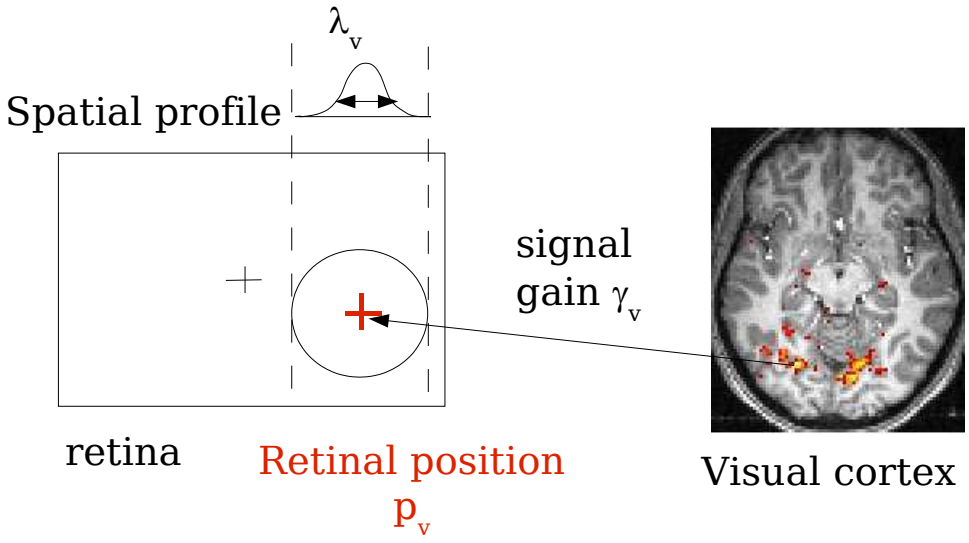


Figure A.5: Receptive field model that is implemented in the forward model. Any voxel v of the visual cortex is associated with a kernel centered on a retinal point p_v, with a width λ_v. The gain that maps the magnitude of visual activity to BOLD activity is modeled by a parameter γ_v.

Given the model (A.3), the estimation of T boils down to the estimation of the parameters $(p_v, \lambda_v, \gamma_v)$. Given Eq. (A.2), this amounts to solving the following equations

$$(p_v, \lambda_v, \gamma_v) = argmin_{p,\lambda,\gamma} \sum_{i=1}^{n} \|\phi^i(v) - \gamma \int_R \rho^i(r) \exp^{-\frac{\|r-p\|^2}{2\lambda^2}} dr\|^2 \tag{A.4}$$

However, due to the non-linear nature of the estimation problem, we find an approximate solution by estimating (i) p_v first, then (ii) λ_v, then (iii) γ_v.

(i) The estimation of p_v is standard in retinotopic mapping experiments [183].

(ii) The size λ_v of the receptive field could be determined from the retinotopic data [189, 47] ; however, here we prefer to rely on two models:

$$(M_1)\ \lambda_v = l_0 \tag{A.5}$$
$$(M_2)\ \lambda_v = l_1\|p_v\| \tag{A.6}$$

In the first model the width of the receptive field is constant. In the second model, the width is proportional to the eccentricity of its center. These two models are two possible simplifications of the current physiological knowledge about receptive field size, which corresponds to an increasing affine function whose characteristics depend on the visual area considered [189]. Model (M_1) might be a more robust choice on real data, given the strong non-linear dependence of the model (A.4) on λ_v. An illustration of the results of the inverse problem using (M_1) or (M_2) is given in Figure A.6. Thereafter we retain the model (M_1), where the constant l_0 is 0.75 degrees in the visual field.

(iii) Last, the estimation of γ_v from Eq. (A.4) is now straightforward and is performed by linear regression. Our estimation procedure thus yields

$$[\hat{T}\rho](v) = \hat{\gamma_v} \int_R \rho(r) \exp^{-\frac{\|r-\hat{\rho_v}\|^2}{2\hat{\lambda_v}^2}} dr \tag{A.7}$$

Solution of the inverse problem

Once the operator T has been estimated, it can be used to infer the visual image ρ associated with any activation map ϕ. In our setting, the images ϕ are now those obtained from the domino or imagery experiment.

If $\hat{T}$ was invertible, the straightforward estimate of ρ would be

$$\hat{\rho} = \hat{T}^{-1}\phi \tag{A.8}$$

However, $\hat{T}$ might not be invertible -neither in theory nor in practice. The estimation of ρ must be regularized. This can be simply cast in a Bayesian framework

$$P(\rho|\phi) \propto P(\rho)P(\phi|\rho) \tag{A.9}$$

Given our Gaussian noise model hypothesis, the likelihood writes

$$P(\phi|\rho) \propto \exp\left(-\frac{1}{2}(\phi - \hat{T}\rho)'\Delta^{-1}(\phi - \hat{T}\rho)\right), \tag{A.10}$$

where Δ models the uncertainty about the measurement ϕ. Note that, assuming that Δ is diagonal [1], an estimate of this uncertainty is provided by the GLM analysis when ϕ is a parametric image.

[1] This amounts to assuming that the errors in the forward model are uncorrelated. This oversimplification is necessary for computational reasons.

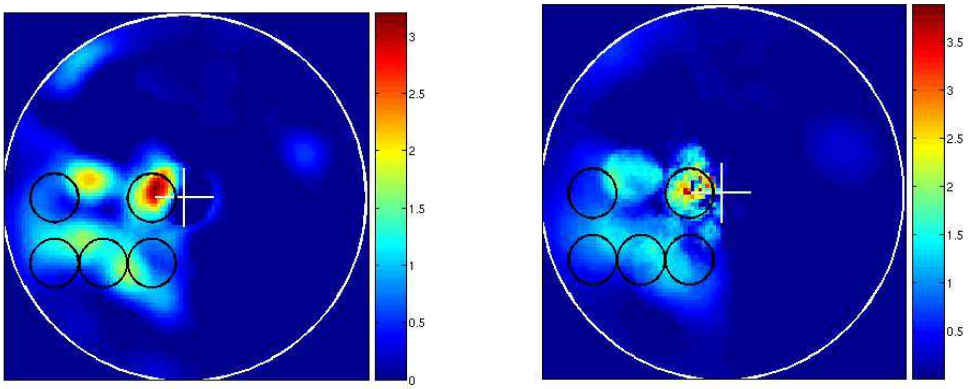


Figure A.6: Comparison of receptive fields models M_1 and M_2 when applied in the solution of inverse problem described in section A.2.3. Left: the model M_1 is used, resulting in a spatially stationary smoothness of the reconstructed image. Right: using model M_2, the reconstructed images are rougher in the foveal region and smoother at the periphery. The true stimulus is represented by five circles in both cases. Note that these differences have little or no impact on pattern identification.

Given the model (A.10) for the likelihood, it is natural to choose the conjugate, hence normal prior

$$P(\rho) = \exp\left(-\frac{1}{2}\rho' K^{-1}\rho\right) \tag{A.11}$$

This means that visual activations are expected to be zero, with a spatial correlation structure provided by K. The prior can be a simple shrinkage prior ($K = \mu^{-1} I_{\mathcal{P}}$), $I_{\mathcal{P}}$ being the $\mathcal{P} \times \mathcal{P}$ identity matrix and μ a positive constant, or it may involve some spatial modeling (e.g. $K_{ij} = k(p(i) - p(j))$, where k is some decreasing function of the distance $|p(i) - p(j)|$).

The solution of the inverse problem consists in minimizing the following functional

$$\Psi(\rho) = (\phi - \hat{T}\rho)'\Delta^{-1}(\phi - \hat{T}\rho) + \rho' K^{-1}\rho \tag{A.12}$$

Note that the covariance of the estimator $\hat{\rho}$ can be estimated as

$$\hat{\Lambda}_\rho = \left(K^{-1} + \hat{T}'\Delta^{-1}\hat{T}\right)^{-1} \tag{A.13}$$

This allows to estimate the likelihood of an activation at a given point r of the retina through the statistic

$$\tau(r) = \frac{\hat{\rho}(r)}{\sqrt{\hat{\Lambda}_\rho(r,r)}} \tag{A.14}$$

parameter	value
p_v	polar coordinates estimated from the retinotopy data (voxel-based)
λ_v	0.75 degrees
γ_v	estimated by linear regression (voxel-based)
μ	$0.0001 \sum_{i,j} \lvert T_{ij} \rvert$
η	0.01

Table A.1: Values of the parameters used in the forward/inverse retinotopy model.

This neglects the covariance between neighboring points, but allows for an easy interpretation, since it yields the probability that $\rho(r)$ is indeed positive given our observation.

In practice, we initialize ρ to 0, and iterate the update rule

$$\rho^{(i+1)} = \rho^{(i)} - \eta \nabla \Psi(\rho^{(i)}), \tag{A.15}$$

η being small enough to ensure convergence. We have tried two possible alternatives for K^{-1}, namely $K^{-1} = \mu I_{\mathcal{P}}$ and

$$K^{-1}_{ij} = \begin{cases} 1 & \text{if } i = j \\ -0.25 & \text{if } p(i) \text{ and } p(j) \text{ are four-neighbors} \\ 0 & \text{otherwise} \end{cases} \tag{A.16}$$

The final difference was not very important but the model A.16 performed slightly better and was used in our experiments. The factor $\mu > 0$ - which characterizes the amount of regularization - has to be set a priori. It was chosen to be proportional to the norm of $\hat{T}$. We noticed that halving it had little impact on the resulting image.

The set of parameters used in the forward/inverse problem is summarized in Table A.1. Last, we have approximated Λ_ρ (see Eq. (A.13)) through the inverse of the diagonal part of $K^{-1} + \hat{T}' \Delta^{-1} \hat{T}$.

Evaluation of the inverse reconstruction

We have computed the correlation of the reconstructed pattern with the different candidate patterns $corr(i, s) =< \tau^i | s > \forall s \in \mathcal{S}$; the predicted stimulus is then the best correlating one $s^\star(i) = argmax_s corr(i, s)$. The performance of the stimulus decoding can then be assessed as the correct prediction rate $P(s^\star(i) = x^i)$.

A.2.4 Results

Explicit reconstruction of the visual stimuli

The reconstructed visual images $\tau(r)$(see section A.2.3) were correlated with the true stimuli, so that the prediction was the best correlated input image. Note that these are *ts* images, i.e. one for each trial.

subject	correct classification rate	
	left hemisphere	right hemisphere
bru2773	41%	65%
bru2774	71%	51%
bru2782	69%	53%
bru2783	60%	47%
bru2784	42%	43%
bru3070	63%	61%
bru3071	69%	56%
bru3072	50%	56%

Table A.2: Correct classification rate of the trial-specific functional patterns across subjects and hemispheres after explicit reconstruction of the stimuli in the domino experiment. The chance level is 1/6, and a score of 27% is above chance level with a P-value of 10^{-3}.

The rate of correct responses is given in Table A.2 for each subject and hemisphere. This rate varies between 41% and 71%, hence is significantly ($P < 10^{-11}$) above chance level (1/6) in all cases, in spite of significant between subject variability.

An example of the condition-specific reconstructed maps $\tau(r)$ is given for the two hemispheres of subject bru3070 in Figure A.7. The correlation of the reconstructed visual patterns with the different candidate shapes is given in Table A.3 for this subject. Although the most lateral part of the stimulus was imperfectly inferred, this reconstruction allows for an unambiguous recognition of the true stimulus in both hemispheres.

In general the images reconstructed from the other datasets have similar quality and correlation scores. The average of the correlation scores across all subjects is given in Table A.4.

We have also performed the reconstruction of the stimulus using only voxels from area V1, which has been delineated from the retinotopy experiment. This gives quite similar results as the reconstruction from all retinotopic voxels, in terms of visual appearance and in terms of correlation. Reconstructed images are provided in Figure A.8 for subject bru3070.

Classification of the trial-specific activation images

In the analysis of the domino experiment we have selected the voxels based on their Anova score, keeping only voxels with an activity significantly modulated by the domino category ($P < 0.1$, FDR corrected), see Appendix A.1.2. The activity of the selected voxels is the input to a linear SVM classifier. Table A.5 presents results with two different cross-validation schemes: on the left part of the table, three sessions (108 trials, Section A.2.2) are used as the training samples and the fourth session (36 trials) as the independent test set. This procedure is repeated four times and classification rate are averaged across the four runs. On the right part of the table, we performed a Leave-One-Out (LOO) procedure

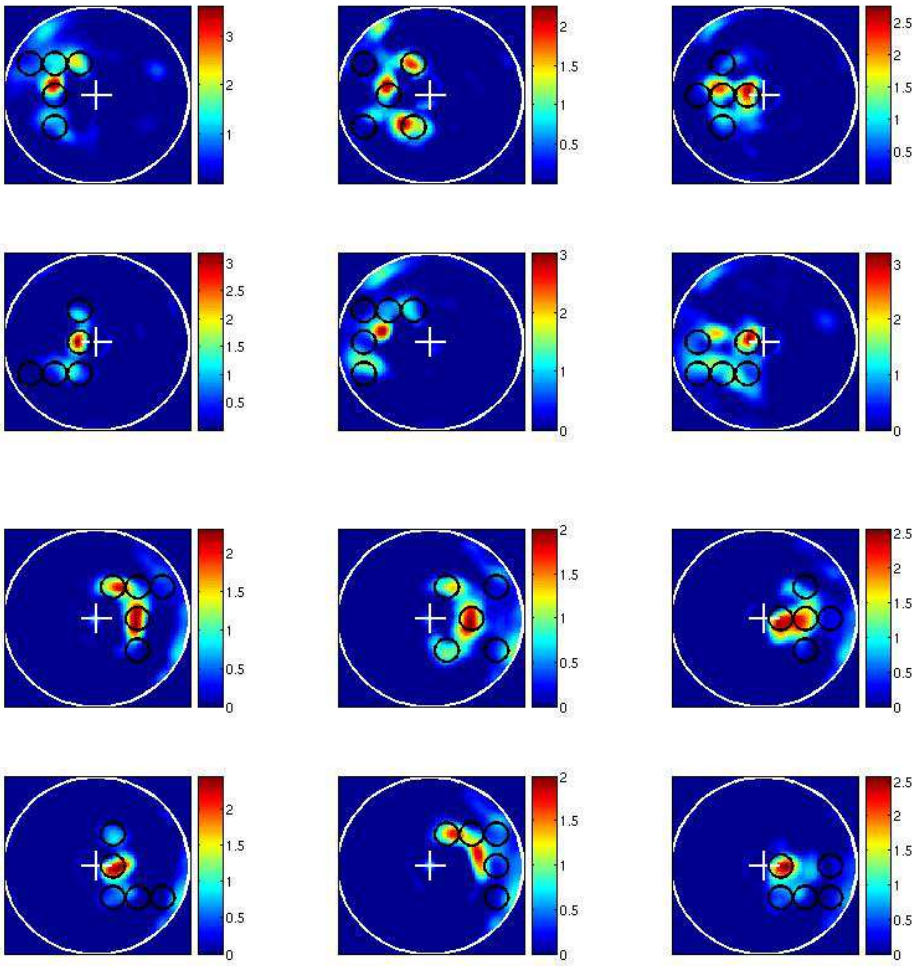


Figure A.7: Explicit reconstruction of the condition-specific visual patterns τ obtained for subject bru3070: (top) left part of the visual field; (bottom): right half of the visual field. The true stimulus is defined by five disks whose contour is superimposed on the reconstructed visual image. In spite of the imperfections of the reconstruction, all patterns can be identified with the true stimulus for both hemispheres by correlation analysis.

	reconstructed pattern	candidate pattern					
							
left hemi-sphere		**0.6055**	0.3904	0.3442	0.2102	0.3941	0.0736
		0.4259	**0.4411**	0.2489	0.3897	0.2417	0.2368
		0.2722	0.1674	**0.5217**	0.3280	0.0116	0.3211
		0.1227	-0.0766	0.0590	**0.3182**	-0.2499	0.1973
		0.2379	0.1198	0.1296	0.0503	**0.3653**	0.0327
		0.2075	0.3084	0.4839	0.5154	0.1406	**0.6000**
right hemi-sphere		**0.5634**	0.4418	0.2522	0.1395	0.2668	-0.0771
		0.5289	**0.5803**	0.3946	0.3531	0.3231	0.2651
		0.3385	0.2886	**0.5412**	0.2869	0.1399	0.2888
		0.0620	0.1483	0.2369	**0.4620**	-0.0576	0.3053
		0.4219	0.2483	0.2177	0.0700	**0.4618**	-0.0305
		0.1135	0.2035	0.4869	0.4376	-0.0260	**0.4929**

Table A.3: Correlation of the reconstructed pattern with the different candidate patterns for subject 3070. Ideally, the diagonal value should be 1, and the off-diagonal values should be between 0 and 0.8, reflecting the correlation between the true stimuli. In the present case, the maximal values of each row , indicated in bold font, are actually in the diagonal, within the [0.3 0.8] interval.

	reconstructed pattern	candidate pattern					
Left hemisphere		**0.5346**	0.3622	0.4258	0.2872	0.4181	0.2121
		0.4802	0.4449	0.4279	0.4179	0.3973	0.3579
		0.3809	0.2871	**0.5613**	0.3910	0.2788	0.4069
		0.3019	0.3000	0.4188	**0.5148**	0.2049	0.4247
		0.4819	0.3391	0.4043	0.2939	**0.5094**	0.2557
		0.2769	0.3018	0.5169	0.5050	0.2064	**0.5398**
Right hemisphere		**0.5279**	0.3725	0.4311	0.2915	0.3881	0.2116
		0.4872	0.4430	0.4694	0.4199	0.3871	0.3638
		0.3601	0.2792	**0.5784**	0.4343	0.2568	0.4382
		0.2563	0.2600	0.4127	**0.5059**	0.1671	0.4236
		0.4487	0.3047	0.3486	0.1995	**0.4661**	0.1657
		0.2581	0.2658	**0.5002**	0.4869	0.1633	0.4937

Table A.4: Correlation of the reconstructed pattern with the different candidate patterns, averaged across all subjects. Ideally, the diagonal value should be 1, and the off-diagonal values should be between 0 and 0.8, reflecting the correlation between the true stimuli. The maximal values of each row are indicated in bold font.

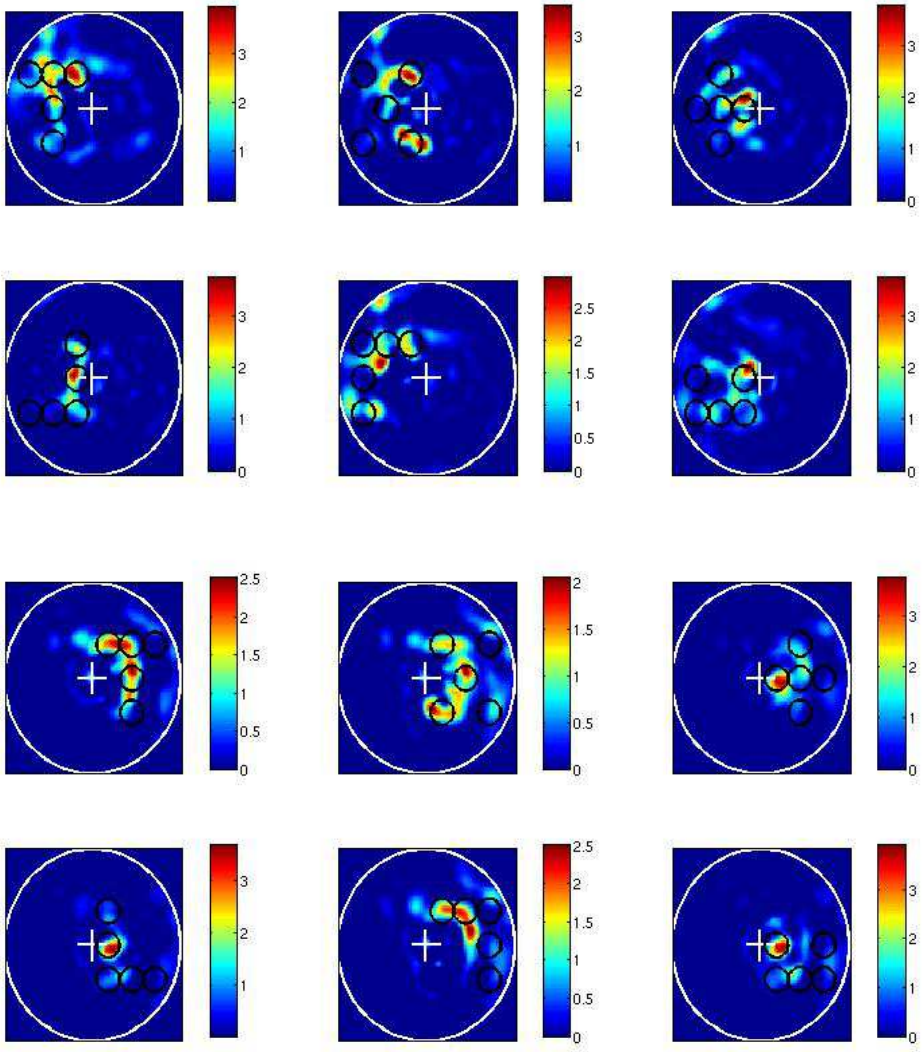


Figure A.8: Explicit reconstruction of the condition-specific visual patterns τ obtained for subject bru3070, using V1 voxels only: (top) left part of the visual field; (bottom): right half of the visual field. The true stimulus is defined by five disks whose contour is superimposed on the reconstructed visual image. Compare with Fig. A.7.

where all samples except one are used to train the discriminant model, which is then tested against the left-out sample. We obtained between 70 and 96% correct classification, according to the subject and the hemisphere. All 16 datasets were classified significantly ($P < 10^{-11}$) above the chance level (1/6 or 16.7% correct responses).

	correct classification rate			
	4 folds cross-validation		LOO cross-validation	
subject	left hemisphere	right hemisphere	left hemisphere	right hemisphere
bru2773	81%	81%	70%	74%
bru2774	77%	73%	85%	70%
bru2782	78%	80%	85%	86%
bru2783	92%	96%	91%	94%
bru2784	83%	75%	85%	76%
bru3070	81%	88%	86%	90%
bru3071	93%	83%	96%	83%
bru3072	72%	87%	75%	88%
means	82.5% ($P < 10^{-15}$)		83.4% ($P < 10^{-15}$)	

Table A.5: Correct classification rate of the trial-specific functional patterns across subjects and hemispheres in the visual stimulation experiment. The selection of significant features ($P < 0.1$, FDR corrected), is followed by a linear-SVM analysis. Two cross-validation methods are used, left: learn on 3 sessions, then test on the fourth; right learn on all samples except one then test on the left out sample. The chance level is 1/6, and a score of 27% is above chance level with a P-value of 10^{-3}.

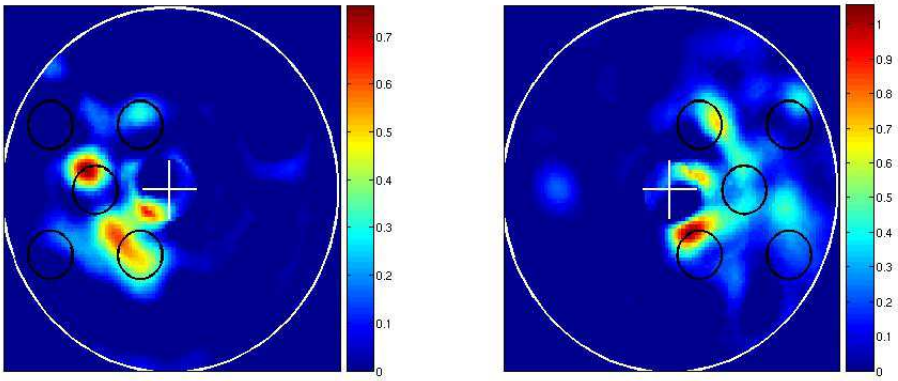
Across subjects and hemispheres, we found that 50-60% of the most discriminative voxels were in V1, while only 20% were in V2 (ventral and dorsal). We did not try to study other visual areas, since their delineation was not reliable enough from our retinotopic maps.

Mental imagery : explicit reconstruction of the patterns

The imagery activation images were also submitted to the inverse reconstruction procedure. We have tried to identify the pattern that was imagined by the subject using separately the data from the left and right hemisphere using condition-specific activation images, i.e. the images being averaged across trials, and trial-specific activation images. An example of condition-specific reconstructed pattern is given in Figure A.9, together with its correlation with the candidate shapes.

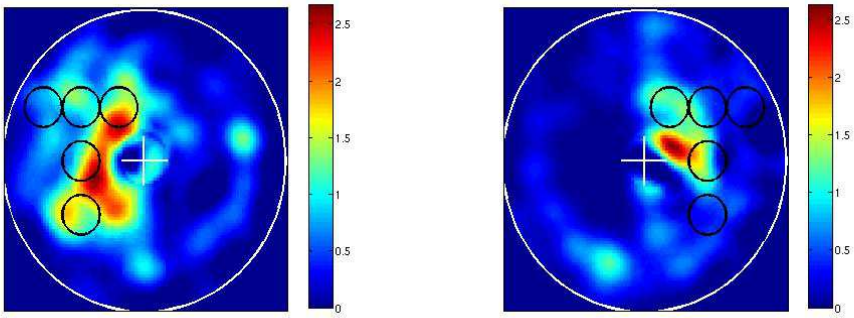
In five out sixteen hemispheres (bru2774, right, bru2783, right, bru3070, right, bru3071, left, bru3072, left), we were able to predict the stimulus that the subject had imagined -or reported to imagine- by correlation of the reconstructed pattern with the candidate patterns.

On a trial-by trial basis, we were able i) to identify the laterality (left or right hemifield) of the imagined pattern with significant accuracy, in five out of eight subjects ii) to identify the imagined



(a)

	candidate pattern					
hemisphere						
Left	0.1154	0.1367	0.2471	**0.2978**	-0.0571	0.2651
right	0.3269	**0.3841**	0.3728	0.3553	0.3126	0.3713



(b)

	candidate pattern					
hemisphere						
Left	0.4960	0.4808	0.4340	**0.5003**	0.3674	0.4218
right	**0.4232**	0.3667	0.4157	0.3792	0.2912	0.2724

Figure A.9: (a) Explicit reconstruction of the condition-specific imagery pattern obtained for subject bru3070, in the left and right hemifields. The true pattern (as stated by the subject) is given by the set of circles superimposed on the figure. Note from the color scale that the activation magnitude is clearly smaller than in Figure A.7. (Bottom) correlation with the candidate shapes. The left hemifield is not identified correctly, while the right hemifield is. (b) The same data, for subject bru3072. In that case, the imagined stimulus is also identified correctly in the right hemifield.

	correct classification rate computed by LOO (p-value)							
	Bonferroni $P < 0.05$				FDR $P < 0.1$			
subject	left hemisphere		right hemisphere		left hemisphere		right hemisphere	
bru2773	11%	(0.83)	**44%**	(**0.005**)	11%	(0.83)	**83%**	(**1.04e−9**)
bru2774	28%	(0.17)	*0%*	(0.96)	17%	(0.6)	*0%*	(0.96)
bru2782	33%	(0.07)	17%	(0.6)	17%	(0.6)	17%	(0.6)
bru2783	*0%*	(0.96)	6%	(0.96)	*0%*	(0.96)	11%	(0.83)
bru3070	**67%**	(**2.19e−5**)	**40%**	(**0.03**)	**73%**	(**1.94e−6**)	20%	(0.47)
bru3071	*0%*	(0.96)	11%	(0.83)	*0%*	(0.96)	6%	(0.96)
bru3072	**50%**	(**1.13e−3**)	**38%**	(**0.02**)	**44%**	(**5.3e−3**)	**39%**	(**0.02**)

Table A.6: Correct classification rate of the 18 trials of the imagery experiment, using a discriminant model built on the 144 trials of the domino experiment of the corresponding subject and hemisphere. The left and right parts of the Table differ on the feature selection method. Each rate is given with P-values computed relative to the null hypothesis that the classifier was operating at chance level. Significantly greater than chance results are emphasized with a bold font. Non-significant results are in grey. The chance level is 1/6, and a score of 33% is above chance level with a P-value of 0.05.

stimulus with significant accuracy in three hemispheres: bru2783, right (78% accuracy, $P < 10^{-8}$), bru2784, left (44% accuracy, $P < 0.006$) and bru3071, left (39% accuracy, $P < 0.03$).

Mental imagery: identification of trial-specific patterns through classification

In this experiment we tried to identify the pattern that was imagined by the subject using the data from the left and right hemispheres separately. The method is similar to the one used for the classification of the trial-specific domino activation maps (Section A.2.4), except for two points.

(i) We compare a stringent feature selection scheme ($P < 0.05$ after Bonferroni correction, see the left part of Table A.6) with the initial scheme ($P < 0.1$ FDR corrected, see the right part of Table A.6).

(ii) The second difference lies in the validation procedure: for each subject, we used the 144 visual trials of the domino experiment as training samples, and the 18 trials of the imagery experiment as test samples. The correct classification rate thus indicate the correspondence between the predicted shape and the shape reported by the subjects.

In the first experiment (Bonferroni $P < 0.05$), we obtain significantly above chance recognition rates in five of out of sixteen hemispheres. Using the less stringent feature selection method (FDR $P < 0.1$), four cases remain above chance, but it should be noticed that it dramatically improves (up to 83%) the recognition rate in subject bru2773, right hemisphere. A possible reason is that lenient feature selection schemes yield high performance in some datasets with high signal levels, and poor performance in datasets with lower signal levels.

A.2.5 Discussion

Inverse retinotopy

We have presented a model-based scheme to decode the information carried by the occipital retinotopic cortex that was successful in identifying the presented stimuli, significantly for all subjects and hemispheres (see Table A.2).

These results confirm that *i)* retinotopic activations in the primary visual cortex are reproducible across trials and sessions, *ii)* the retinotopic information obtained with the now classical *traveling wave* paradigm [51] can be used as a code, as suggested in [220, 228], *iii)* a linear filter model for $V1$ [83], that we use in our forward model, holds as a first approximation.

An important new feature of our approach compared to recent contributions [89, 100] is that we compare classification techniques that model implicitly the stimulus/activation relationship with the explicit resolution of an inverse problem. Importantly, these techniques are based on different hypotheses:

- Supervised classification assumes that reproducible differences might be found between functional images, so that the associated stimulus can be inferred. The important issue is to identify and select the discriminating information and to assess its reliability. Cross-validation and heuristic arguments are used to solve the problem. The acquired knowledge is restricted to the categories presented in the learning set.
- Inverse reconstruction builds on a model of the activation process, with explicit simplifying assumptions. The main issue is to find a simplified model that remains consistent with the data. The inverse problem can then be solved in a rather systematic way, and with any kind of input data: the initial retinotopic mapping is assumed here to yield a generalizable model of *any* visual activation.

As we have noticed, SVM-based classification yields more accurate results than the inverse problem; the price to pay is that it is not as general. But a key point is that the high performance of classifiers indicates that sufficient discriminant information is indeed present in the data, even if it was not explicitly decoded: some identification failures in the inverse problem can thus be attributed to shortcomings of the model rather than insufficient information in the data (which might e.g. be related to the performance or attention of the subjects).

Mental imagery

Moreover, we were able to extrapolate our predictions from passive viewing experiments to mental imagery in some of the subjects, with particularly strong evidence when using classification tools. These latter findings are consistent with the hypothesis that mental imagery involves activation in the primary visual areas, and that the spatial structure of these activations is accounted for by standard retinotopic mapping [113, 110, 188].

If inter-subject differences play a mild role in the performance of inverse retinotopy algorithms applied to actual visual stimuli, they might have more impact on the results of the imagery experiment. In particular, the different subjects reported more or less subjective difficulty in the task performance, as already noticed in the literature [112]. This is clear in Table A.6, where the performance in the prediction of the laterality varies strongly across subjects. One can also notice that in [150], mental imagery activations were decoded in three out of eight subjects. In view of this, the good performance achieved in five hemispheres is thus an important result (if responses were random, the probability of obtaining significant values in five hemispheres would be $P < 0.00043$). The explicit identification of the imagined pattern is a challenging task, due to the weak signals that are obtained (see Figure A.9). Note that, besides the well-known weakness of retinotopic activations in imagery experiments [110], the subjects had to keep their eyes open and fixate the grid during this experiment, which might have reduced the level of activation in primary visual areas.

Technical aspects

Care should be taken when evoking fMRI-based brain reading experiments. In particular, any method rests on a deconvolution of the hemodynamic responses on a voxel-by-voxel basis. We have performed this using a standard GLM procedure, which is reasonable given that the temporal linearity hypothesis might be fulfilled given our inter-stimulus intervals (6s) [13, 191]. One might think, however, that unmodeled spatio/temporal interactions may be present in the data.

Another set of simplifications were introduced in our formulation of the forward model: linearity of the transfer operator, isotropic Gaussian receptive field (RF) structure at the voxel level, constant RF size, linear gain. While this model might be partially supported by current knowledge about V1 [220, 83], it is obviously an over-simplification [152], especially if one considers higher visual areas. However, it is important to keep in mind that fMRI signals represent in each voxel the average of the activity of thousands of neurons, so that some hypotheses, e.g. the spatial linearity or superposition principle used here, that are known to be violated at a microscopic level, may hold approximately at the much lower spatial resolution and/or using standard field strength. Clearly, our forward problem framework may be a good benchmark to test violations of different hypotheses e.g. the spatial linearity of visual activity as seen in fMRI (see e.g. [186, 185]). For instance, we have also implemented *Mexican hat* filters (Laplacian of the Gaussian), but did not find significant improvements in the results. Our interpretation is that the main bottleneck in the forward/inverse problem is the correctness of the estimate of p_v in each voxel [see Eq. (A.3)]. Assuming that this parameter is perfectly none, modelling non-linear effects or non-isotropic and non-Gaussian Receptive fields would become worthwhile.

We found that a majority (50 to 60 %) of the most discriminant voxels used in the classifier were in V1, while a much smaller proportion (around 20 %) were in V2. Since this delineation was not the primary goal of our analysis, we were not sure to obtain reliable boundaries for other visual areas and did make further identification of the discriminative information. One could indeed expect that most of the information on the spatial layout of the stimulus would be encoded in V1. In Fig. A.8, we show that

performing the inverse reconstruction based only on V1 voxels yields a very good approximation of what is achieved when considering all the retinotopic voxels (see Fig. A.8).

The inverse problem is also limited by the possibility of evaluating correctly the precision of the results (see e.g. Eq. (A.13). Approximations must be performed, so that the probabilistic interpretation of the visual patterns is not fully assessed. For this reason, we based our test procedure on the correlation of the reconstructed pattern with the possible candidates, rather than on a fully probabilistic interpretation of the reconstructed maps.

Power and limits of inverse retinotopy

It should first be noticed that the retinotopic mapping was performed in less that 20 minutes in each subject, so that the limited accuracy of the retinotopic information may be the main limit in this experiment. Not unexpectedly, many confusions occurred between spatial patterns that overlapped. For instance, the  -shaped and  -shaped patterns in Figure A.3(c) were often confounded. Moreover, the  -shaped pattern was rarely identified in the framework of the inverse problem: it is interesting to note that this pattern is the least compact. This might be attributable to the low-pass filtering inherent to the inverse problem. This effect is particularly evident from Figure A.7, and affects the results at the group level (see Table A.4). Interestingly, the  -shaped stimulus was not confused with other patterns when using classification tools. Thus the problem described here might be an intrinsic shortcoming of the forward/inverse problem solution. Another important effect is that the portion of the stimulus closest to the centre of the visual field is apparently much better reconstructed than the activity in the peripheral regions, which is often smoothed out in Figure A.7 -and similarly for all the datasets studied. This weakness is apparently not related to the receptive field size model (see Figure A.6), and might be related to spatial/attentional modulations. For instance, the fact that left and right patterns were presented simultaneously facilitates fixation, but possibly increases foveal attention at the expense of the periphery. Note that this foveal emphasis effect is apparently also present in the imagery data (see Figure A.9).

A.3 Conclusion: perspectives on brain reading

The brain reading approach has gained much popularity in the functional neuroimaging community due to its ability to show that brain activation conveys information about perceptual or cognitive states, even when this information is hardly detectable at the subject level [100]. Ongoing developments include the following:

i. Development of spatial models in conjunction with classification.

ii. Predicting wider classes of perceptual and cognitive variables.

iii. Cross-task generalization of brain classifiers.

Developing Spatial models for classification purpose

As we suggested in Section 6.3.3, one of the promising approaches is the development of multi-resolution methods that may focus on particular regions, where information is encoded at a fine spatial case, without defining them a priori. We also mentioned that we plan to contribute efficient clustering procedures to that end. The spatially constrained Ward's clustering algorithm (see Fig. 3.9) is very promising in this perspective, since it yields a hierarchical subdivision of the brain volume, where subdivisions are maximally homogeneous. Moreover, it seems important to use the classification objective to drive the parcellation process itself, i.e. to combine the spatial model with the different criteria described in Section 6.2.

Predicting wider classes of perceptual and cognitive variables

The prediction of visual scenes perceived by a subject described in Sec. A.2 has elicited important follow-up experiments. In particular, Gallant and colleagues have reused the a similar forward model methodology to model the response of the visual cortex to the representation of images [101]. This model is in fact much richer, since it also includes explicit parameters for the receptive field width, spatial frequency and orientation, as fully parametrized Gabor filters. The resulting bank of filters being defined, the authors first fit them to some training data, here natural images: they find which filter best match the activity at each voxel of the visual cortex.

This model cannot be inverted, thus the authors do not proceed with an explicit reconstruction of the stimuli. Instead, they simulate the activation elicited by one hundred candidate images, and find among them which one has been presented: they use it to *identify* the presented image, which can be seen as an intermediate step between the percept reconstruction that we performed and standard classification experiments, where inference is limited to categories.

One of the most fascinating aspects is that natural images are used, instead of geometric patterns as those used in Sec. A.2. Although the inverse problem approach cannot always be feasible, using parallel classifiers in charge of characterizing one portion of the visual field is an interesting alternative, that has been explored in [137]. It is possible that the combination of explicit prediction and *local classifier* is a powerful combination to go one step further in brain activity decoding.

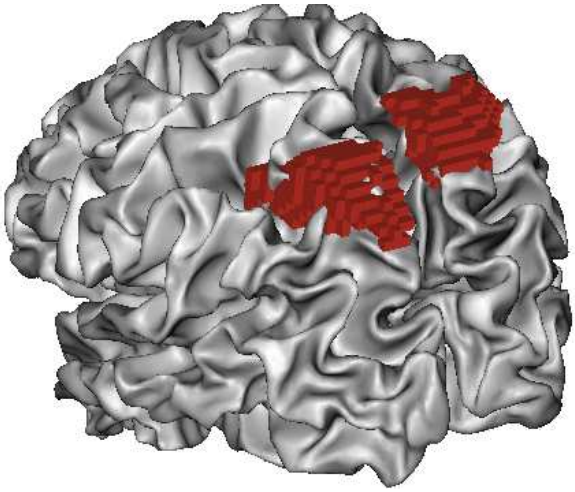
Cross-task generalization of brain classifiers

Last, standard classifier can be subverted to unexpected use, and provide that way, unexpected but essential information on brain function. To simplify our presentation of this *subversion*, let us briefly describe an experiment recently performed at Neurospin by André Knops and Edward Hubbard [submitted]: 15 subjects underwent a scanning session where they first performed left and right saccadic eye movement when prompted to. The stimuli were sufficiently slow and sparse to allow trial-by-trail analysis of the data. Then the same subjects had to perform mental arithmetic tasks, such addition and subtractions, where the operands were clouds of points presented a few seconds apart. In this second part of the experiment, the period of interest is the instant where the subjects evaluate mentally the -necessarily

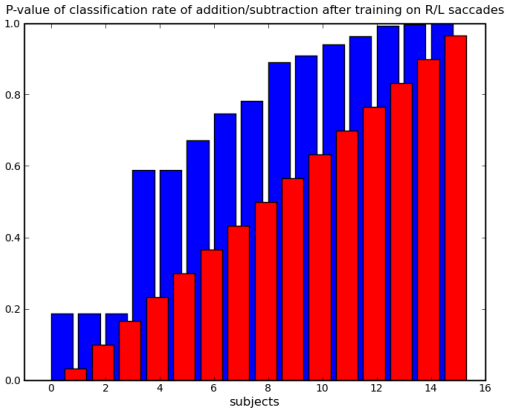
approximate- result of the operation before choosing one solution over six. Functional images were acquired at Neurospin Center on a 3T MR system (Siemens TrioTim Syngo) as T2*-weighted echo-planar image (EPI) volumes. Forty transverse slices covering the whole brain were obtained with a TR of $2.4s$ (TE: $30ms$; flip angle, 81°; $3 \times 3 \times 3$ mm voxels; no gap).

We first defined the saccade-responsive regions from a separate localizer task: the LIP-homologue region (see Fig. A.10,(a)), the frontal eye field, and a more anterior pre-frontal region, bilaterally. With different classifiers, we checked that we could predict if a given saccade was towards left or right: we achieve about 70% accuracy across subjects (std=9%, range:[56%-84%]) in a 10-fold cross-validation using bilateral LIP, but the prediction was also consistently accurate with other brain regions (FEF, IPS, primary motor). Then we used the classifier trained on saccades on the arithmetics trials. It turned out that, while subtraction were equally classified as left or right saccades across subjects, additions were more systematically classified as right saccades (about 60% across subjects). Altogether, the interpretation of subtractions (respectively addition) as left (resp. right) saccades yielded a classification accuracy of about 55% across subjects ($p < 0.008$). This generalization from saccades to mental arithmetics occurred when considering the LIP region only. The interpretation is that the same neural populations would be involved in the performance of mental arithmetics and ocular saccades, so that higher cognitive operations recruit neural networks that are specialized for low-level tasks common to other species.

This experiment can be viewed as a new instance of the use of classifiers as a new kind of statistic that provides region-level or information -level information, and can be used to address new questions, concerning the joint selectivity of neuronal populations to different kinds of stimuli. With the current increase of brain data resolution, we might expect that brain decoding approaches will become more and more popular in the cognitive neuroscience community.



(a)



(b)

(c)

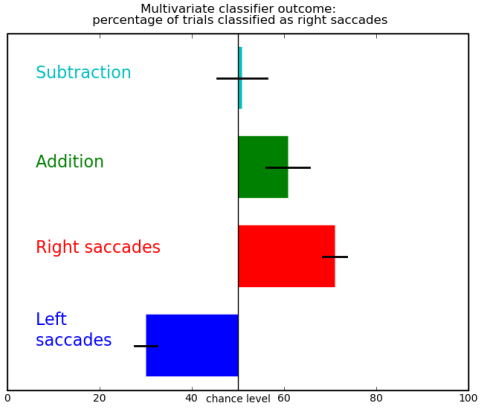


Figure A.10: Cross-task generalization of the prediction. (a) the two LIP-homologue regions are outlined as saccade-responsive regions in a separate localizer task. (b) Across subjects, the classification of subtractions as left saccades and of the additions as right saccades is above chance: the ordered classification p-value across individuals is above the diagonal, i.e. the p-value expected by chance. (c) In fact the classification of addition as right saccades is more successful than the classification of subtraction trials as left saccades.

Appendix B

Technical appendix

In this we chapter, we simply provide some developments and formulas that are needed for the completeness of this document, but are not original developments.

B.1 On the equivalence of the Likelihood ratio and the F statistic for a Gaussian i.i.d. model

In this section, we simply show that the Likelihood Ratio is equivalent to an F test in a non-hierarchical model, assuming that the noise is i.i.d.. This is meant to justify the use of the summary statistics t, F instead of the Likelihood Ratio (LR) in the inference problems encountered in fMRI. Let us recall that the generative model for a given voxel-based time course is

$$y = X\beta + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{B.1}$$

Given a contrast c, the two alternative hypotheses are

$$\begin{array}{ll} (H_0) & c^T\beta = 0 \\ (H_1) & c^T\beta \neq 0 \end{array}$$

Note that this corresponds to an unsigned or two-sided test. Without loss of generality, this contrast can be assume to be of the form $c = [1, .., 1, ..., 0]$, where the number of 1 corresponds to the dimension of c; in this case the design matrix can be partitioned as $X = [X_i, X_r]$ and similarly β is partitioned in $\begin{bmatrix} \beta_i \\ \beta_r \end{bmatrix}$. The model B.1 becomes

$$y = X_i\beta_i + X_r\beta_r + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{B.2}$$

The livelihood ratio simply writes:

$$\Lambda = \frac{sup_{\{\beta_i, \beta_r, \sigma^2\}}\mathcal{N}(y; X_i\beta_i + X_r\beta_r, \sigma^2 I)}{sup_{\{\beta_r, \sigma^2\}}\mathcal{N}(y; X_r\beta_r, \sigma^2 I)} \tag{B.3}$$

Letting $P_X = X(X^TX)^{-1}X^T$ be the projection matrix to the column space of X and similarly $P_r = X_r(X_r^TX_r)^{-1}X_r^T$, it can simply be shown that

$$\log(\Lambda) = \frac{N_t}{2}\log\frac{y^T(I-P_R)y}{y^T(I-P_X)y} \tag{B.4}$$

On this same problem, the F statistic used for testing this hypothesis is given by

$$F = \frac{y^T(P_X-P_R)y}{y^T(I-P_X)y}\frac{N_t - rank(X)}{rank(X)-rank(X_i)} \tag{B.5}$$

It follows that

$$F = \frac{N_t - rank(X)}{rank(X)-rank(X_i)}(\Lambda^{\frac{2}{N_t}}-1) \tag{B.6}$$

and finally,

$$\Lambda = \left(\frac{rank(X)-rank(X_i)}{N_t - rank(X)}F + 1\right)^{\frac{N_t}{2}} \tag{B.7}$$

This shows that making a test on the likelihood ratio Λ or the F statistic is equivalent in that case. This can be easily be generalized in the case of known covariance V (possibly different from the identity). However, the strict equivalence is broken in the case where V is unknown [36].

While Λ can be assessed only asymptotically by using the property that $\Lambda \propto \chi^2_\nu$, where $\nu = \text{rank}(X_i)$ when the number of samples goes to infinity, the F test can be assessed whatever the number of degrees of freedom.

In the case of the one-sided test, the equivalence of the LR with the t test generalizes readily the equivalence between the LR and the F test.

B.2 Bayesian derivation of the t test

Let us consider the model defined in Eq. (2.4). In a Bayesian approach, this model can be shown to yield a student posterior distribution for the effect β [66]: The likelihood model writes simply:

$$p(Y|\beta,\sigma) = \mathcal{N}(Y; X\beta, \sigma^2 V) \tag{B.8}$$

Uninformative (and improper) priors are then chosen on the (β,σ) parameters:

$$p(\beta,\sigma) \propto \frac{1}{\sigma} \tag{B.9}$$

Then

$$p(\beta|Y) = \int p(\beta,\sigma|Y)d\sigma \propto \int \mathcal{N}(Y; X\beta,\sigma^2V)\frac{d\sigma}{\sigma} = \frac{1}{Q(\beta)^{N_t/2}}, \tag{B.10}$$

where $Q(\beta) = (Y - X\beta)V^{-1}(Y - X\beta)$ can be rewritten using the least squares estimates in (2.5-2.6):

$$Q(\beta) = \hat{\varepsilon}^T V^{-1}\hat{\varepsilon} + (\beta - \hat{\beta})^T X^T V^{-1} X(\beta - \hat{\beta}), \tag{B.11}$$

where $\hat{\varepsilon} = Y - X\hat{\beta}$. Finally, we obtain:

$$p(\beta|Y) \propto \frac{1}{\left(1 + \frac{(\beta-\hat{\beta})^T X^T V^{-1} X(\beta-\hat{\beta})}{\hat{\varepsilon}^T V^{-1}\hat{\varepsilon}}\right)^{N_t/2}} \tag{B.12}$$

It follows naturally that β follows a multivariate student distribution with mean $\hat{\beta}$, variance $\frac{\hat{\varepsilon}^T V^{-1}\hat{\varepsilon}}{N_t - rank(X)}(X^T V^{-1} X)^{-1}$ and $N_t - rank(X)$ degrees of freedom. The application of contrasts yields posteriori distributions for $c^T\beta$ that correspond to the test found in Eq. (2.13).

Obviously, slightly different results would arise if informative priors were used instead. The main advantage of the Bayesian approach is that it allows an interpretation of the parameters $\hat{\beta}, \hat{\sigma}$ as a parametrization of the distribution of the true effects. For instance, the probability $p(\beta > \theta|Y)$ for thresholds θ that are different from 0 [69].

B.3 D'Agostino-Pearson normality test

Let x be a sample of size n whose normality needs to be tested. The D'Agostino-Pearson test first analyzes data to determine skewness $\sqrt{b_1}$ (to quantify the asymmetry of the distribution) and kurtosis b_2 (to quantify the shape of the distribution).

$$\sqrt{b_1} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3\right)^{\frac{3}{2}}} \tag{B.13}$$

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \tag{B.14}$$

It then calculates how far each of these values differs from the value expected with a normal distribution, and computes a single p-value from the sum of the squares of these discrepancies. This test is a combination of the D'Agostino skewness test and Anscombe-Glynn kurtosis test. The test statistic is

$$K^2 = \Xi^2(\sqrt{b_1}) + \psi^2(b_2), \text{where} \tag{B.15}$$

where $\Xi^2(\sqrt{b_1})$ and $\psi^2(b_2)$ are the standard normal deviates equivalent to observing $\sqrt{b_1}$(skewness) and

b_2 (kurtosis).

$$\begin{aligned}
\Xi(\sqrt{b_1}) &= \delta \log\left(\frac{Y}{\alpha} + \sqrt{\frac{Y^2}{\alpha^2}+1}\right), \\
\alpha &= \sqrt{\frac{2}{W^2-1}}, \\
\delta &= \frac{1}{\sqrt{\log(W)}}, \\
W^2 &= -1 + \sqrt{2\beta_2(\sqrt{b_1})-1}, \\
\beta_2(\sqrt{b_1}) &= \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)} \\
Y &= \sqrt{b_1}\sqrt{\frac{(n+1)(n+3)}{6(n-2)}}
\end{aligned}$$

and

$$\begin{aligned}
\psi^2(b_2) &= \left(\left(1-\frac{2}{9A}\right) - \sqrt[3]{\frac{1-2/A}{1+\chi\sqrt{2/(A-4)}}}\right)\sqrt{\frac{9A}{2}}, \\
A &= 6 + \frac{8}{\sqrt{\beta_1(b_2)}}\left[\frac{2}{\sqrt{\beta_1(b_2)}} + \sqrt{1+\frac{4}{\sqrt{\beta_1(b_2)}}}\right], \\
\sqrt{\beta_1(b_2)} &= \frac{6(n^2-5n+2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}, \\
\chi &= \frac{1}{\sigma_b}\left(b_2 - \frac{3(n-1)}{n+1}\right), \\
\sigma_b^2 &= \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}
\end{aligned}$$

The K^2 statistic in Eq. (B.15) has approximately a chi-squared distribution, with 2 degrees of freedom when the population is normally distributed. For sample sizes $n > 8$, a normal approximation that is easily computerized is available [32].

B.4 description of the datasets used in our work

B.4.1 The localizer dataset

We use a Localizer dataset, which is a short (5 minutes) experiment used as a routine since 2003 at SHFJ, Orsay and Neurospin.

This experiment is based on an event-related experimental paradigm consisting of ten conditions. Subjects were presented with a series of stimuli and were engaged in tasks such as passive viewing of horizontal or vertical checkerboards, left or right button press after audio or video instruction, computation (subtraction) after video or audio instruction and sentence processing, from the audio or visual modality. Events were randomly occurring in time (mean inter stimulus interval: 3s), with ten occurrences per event type (except button presses for which there are only five trials per session). The experimental paradigm is described with more detail in [167]. Note that contrasts of experimental conditions rely in fact on the sum of number of trials of each condition. For instance, the *left-right button press* contrast combines four experimental conditions (left/ right button press after audio/video instruction) and relies on twenty trials. Similarly, the *audio-video* and *computation-sentences* contrasts rely on sixty and forty trials respectively.

Eighty-one right-handed subjects participated in the study. The subjects gave informed consent and the protocol was approved by the local ethics committee. Functional images were acquired either on a 3T Bruker (2003-2006) scanner using an EPI sequence (TR = $2400ms$, TE = $60ms$, matrix size=64×64, FOV = $24cm \times 24cm$), or on the 1.5 GE Signa scanner (2005-2007), or on Neurospin Siemens Trio scanner (2007-2008). Each volume consisted of n_a $3mm$- or $4mm$-thick axial slices without gap, where n_a varied from 26 to 40 according to the session. A session comprised 130 scans. The first four functional scans were discarded to allow the MR signal to reach steady state. Anatomical T1 images were acquired on the same scanner, with a spatial resolution of $1 \times 1 \times 1.2\ mm^3$.

fMRI data processing consisted in 1) temporal Fourier interpolation to correct for between-slice timing, 2) motion estimation. For all subjects, motion estimates were smaller than 1mm and 1 degree, 3) spatial normalization of the functional images, re-interpolation to $3 \times 3 \times 3\ mm^3$, and 4) smoothing (5mm FWHM). This pre-processing was performed with the SPM5 software (see e.g. [5]). First-level statistical analysis was carried out on these images using the SPM5 software, using standard high-pass filtering and AR(1) whitening. For further analysis, the voxel-based estimated effects, as well as the corresponding statistical maps for several contrasts of interest were retained.

In group studies (for any of the proposed approaches), we need to determine a global brain mask for the group by considering all the voxels that belong to at least half of the individual brain masks defined with SPM5. It comprises approximately 60000 voxels (this is the average size of individual brain masks).

Bibliography

[1] G. K. Aguirre, E. Zarahn, and M. D'Esposito. Empirical analysis of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *Neuroimage*, 5(3):199–212, April 1997.

[2] G. K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human BOLD hemodynamic responses. *Neuroimage*, 7:574, 1998.

[3] S. Allassonnière, Y. Amit, and Trouv A. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):3–29, January 2007.

[4] J. Ashburner and K.J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–66, 1999.

[5] J. Ashburner, K.J. Friston, and W. Penny, editors. *Human Brain Function, 2nd Edition*. Academic press, 2004.

[6] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, October 2007.

[7] C.F. Beckmann and S.M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2):137–152, 2004.

[8] T. E J Behrens, M. Jenkinson, M. D. Robson, S. M. Smith, and H. Johansen-Berg. A consistent relationship between local white matter architecture and functional specialisation in medial frontal cortex. *Neuroimage*, 30(1):220–227, Mar 2006.

[9] Timothy E J Behrens, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. Learning the value of information in an uncertain world. *Nat Neurosci*, 10(9):1214–1221, Sep 2007.

[10] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1):289–300, 1995.

[11] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34:537–541, 1995.

[12] Nathalie Boddaert, Catherine Barthélémy, J.-B. Poline, Yves Samson, Francis Brunelle, and M. Zilbovicius. Autism: functional brain mapping of exceptional calendar capacity. *Br J Psychiatry*, 187:83–86, July 2005.

[13] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, 16:4207–4221, 1996.

[14] M Brett, IS Johnsrude, and AM. Owen. The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, 3(3):243–249, March 2002.

[15] E Bullmore, B Horwitz, G Honey, M Brammer, S Williams, and T Sharma. How good is good enough in path analysis of fMRI data? *Neuroimage*, 11(4):289–301, April 2000.

[16] R. B. Buxton, E. C. Wong, and Frank. L. R. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine*, 39:855–864, June 1998.

[17] A. Cachia, J.-F. Mangin, D. Rivière, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch, and J. Régis. A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams. *Medical Image Analysis*, 7(4):403–416, 2003.

[18] A. Cachia, M.-L. Paillère-Martinot, A. Galinowski, D. Januel, R. De Beaurepaire, F. Bellivier, E. Artiges, T. Gallarda, J. Andoh, D. Bartrés-Faz, E. Duchesnay, D. Rivière, Y. Cointepas, M. Plaze, J.-F. Mangin, and J.-L. Martinot. Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *Neuroimage*, 39(3):927–935, 2008.

[19] Deng Cai, Deng Cai, Xiaofei He, and Jiawei Han. Srda: An efficient algorithm for large-scale discriminant analysis. *IEEE TKDE*, 20(1):1–12, 2008.

[20] Thomas A Carlson, Paul Schrater, and Sheng He. Patterns of activity in the categorical representations of objects. *J Cogn Neurosci*, 15(5):704–717, July 2003.

[21] P. Cathier and J.-F. Mangin. Registration of cortical connectivity matrices. In *Proceedings MMBIA2006*, New York, USA, 2006.

[22] G. A. Christensen, R. D. Rabbit, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, October 1996.

[23] Justin R Chumbley, Karl J Friston, Tom Fearn, and Stefan J Kiebel. A metropolis-hastings algorithm for dynamic causal models. *Neuroimage*, 38(3):478–487, Nov 2007.

[24] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging*, 22(10):1235–1251, October 2003.

[25] P. Ciuciu, T. Vincent, A.-L. Fouque, and A. Roche. Improved fMRI group studies based on spatially varying non-parametric bold signal modeling. In *5th International Symposium on Biomedical Imaging*, pages 1263–1266, Paris, France, May 2008.

[26] C. Clouchoux, O. Coulon, J.-L Anton, , J.-F. Mangin, and J. Régis. A new cortical surface parcellation model and its automatic implementation. In *Proceedings 9th International Conference on Medical Image Computing and Computer Assised Intervention*, LNCS 4191, pages 193–200, Copenhaegen, Denmark, October 2006. Springer Verlag.

[27] C. Clouchoux, O. Coulon, D. Rivière, A. Cachia, J.-F. Mangin, and J. Régis. Anatomically constrained surface parameterization for cortical localization. In *MICCAI'05*, pages 344–351, 2005.

[28] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. 24(5):603–619, May 2002.

[29] Maurizio Corbetta, Michelle J Kincade, Chris Lewis, Abraham Z Snyder, and Ayelet Sapir. Neural basis and recovery of spatial attention deficits in spatial neglect. *Nat Neurosci*, 8(11):1603–1610, Nov 2005.

[30] O. Coulon, J.-F. Mangin, J.-B. Poline, M. Zilbovicius, D. Roumenov, Y. Samson, V. Frouin, and I. Bloch. Structural group analysis of functional activation maps. *Neuroimage*, 11:767–782, 2000.

[31] D. D. Cox and R. L. Savoy. Functional magnetic resonance imaging (fmri) "brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19(2):261–270, June 2003.

[32] Ralph B. D'Agostino and Albert Belanger. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321, November 1990.

[33] C. Damon, P. Pinel, M. Perrot, V. Michel, E. Duchesnay, J.-B. Poline, and B. Thirion. Discriminating between populations of subjects based on FMRI data using sparse features selection and SRDA classifier. In *MICCAI 2008 Workshop on Analysis of Functional Medical Images*, New York, Sept 2008.

[34] C. Davatzikos, K. Ruparel, Y. Fan, D. G. Shen, M. Acharyya, J. W. Loughead, R. C. Gur, and D. D. Langleben. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage*, 28(3):663–668, November 2005.

[35] S. Dehaene, G. Le Clec'H, L. Cohen, J.-B. Poline, P.-F. Van de Moortele, and D. Le Bihan. Inferring behavior from functional images. *Nature Neuroscience*, 1(7):549–550, 1998.

[36] A. J. den Dekker, D. H. J. Poot, R. Bos, and J. Sijbers. Likelihood based hypothesis tests for brain activation detection from mri data distributed by colored noise: a simulation study. *IEEE Transactions on*, pages 1–1, Volume PP,&a href='/xpl/tocpreprint.jsp?isnumber=4359023&punumber=42'>Forthcoming& 2003. Accepted for future publication Medical Imaging.

[37] Thomas Deneux and Olivier Faugeras. Using nonlinear models in fmri data analysis: model selection and activation detection. *Neuroimage*, 32(4):1669–1689, Oct 2006.

[38] John E Desmond and Gary H Glover. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J Neurosci Methods*, 118(2):115–128, August 2002.

[39] Joseph T Devlin and Russell A Poldrack. In praise of tedious anatomy. *Neuroimage*, 37(4):1033–41; discussion 1050–8, Oct 2007.

[40] E.A. DeYoe, G. Carman, and P. Bandettini et al. Mapping striate and extrastriate visual areas in human cerebral cortex. *Neurobiology*, 93:2382–2386, March 1996.

[41] S. Dodel, J.-B. Poline, J.-L. Anton, and M. Brett. The influence of heart beat and respiration on functional connectivity networks. In *Proceedings 2th International Symposium on Biomedical Imaging*, pages 380–383, Arlington, VA, April 2004.

[42] G. Douaud, C. Poupon, Y. Cointepas, J.-F. Mangin, V. Gaura, N. Golestani, P. Krystkowiak, C. Verny, P. Damier, A.-C. Bachoud-Lévi, P. Hantraye, and P. Remy. Diffusion tensor imaging (DTI) in Huntington's disease patients: analyses of fractional anisotropy (FA) maps and apparent diffusion coefficient (ADC) maps. In *ISMRM Workshop on Methods for Quantitative Diffusion MRI of Human Brain, Lake Louise, Canada*, 2005.

[43] Robert F Dougherty, Volker M Koch, Alyssa A Brewer, Bernd Fischer, Jan Modersitzki, and Brian A Wandell. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J Vis*, 3(10):586–598, 2003.

[44] JR Duann, TP Jung, WJ Kuo, TC Yeh, S Makeig, Hsieh JC, and Sejnowski TJ. Single-trial variability in event-related BOLD signals. *Neuroimage*, 15(4):823–35, April 2002.

[45] J. Dubois, L. Hertz-Pannier, G. Dehaene-Lambertz, Y. Cointepas, and D. Le Bihan. Potential of DTI and fiber tracking to evaluate the organization and maturation of white matter bundles in healthy infants. In *ISMRM Workshop on Methods for Quantitative Diffusion MRI of Human Brain, Lake Louise, Canada*, 2005.

[46] E. Duchesnay, A. Cachia, A. Roche, D. Rivière, Y. Cointepas, D. Papadopoulos-Orfanos, M. Zilbovicius, J.-L. Martinot, and J.-F. Mangin. Classification from cortical folding patterns. *IEEE Transactions on Medical Imaging*, 26(4):553–565, 2007.

[47] Robert O Duncan and Geoffrey M Boynton. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, 38(4):659–671, May 2003.

[48] Bradley Efron, Trevor Hastie, Lain Johnstone, and Robert Tibshirani. Least angle regression. *Annals of Statistics*, 32:407–499, 2004.

[49] Simon B Eickhoff, Stefan Heim, Karl Zilles, and Katrin Amunts. Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage*, 32(2):570–582, Aug 2006.

[50] V. El Kouby, Y. Cointepas, C. Poupon, D. Rivière, N. Golestani, J.-B. Poline, D. Le Bihan, and J.-F. Mangin. MR diffusion-based inference of a fiber bundle model from a population of subjects. In *MICCAI'05*, pages 196–204, 2005.

[51] S. A. Engel, G. H. Glover, and B. A. Wandell. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, 7(2):181–192, March 1997.

[52] David C Van Essen. A population-average, landmark- and surface-based (pals) atlas of human cerebral cortex. *Neuroimage*, 28(3):635–662, Nov 2005.

[53] B. Fischl and A. M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci USA*, 97(20):11050–5, 2000.

[54] B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis II: inflation, flattening and a surface-based coordinate system. *Neuroimage*, 9:195–207, 1999.

[55] B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284, 1999.

[56] Bruce Fischl, Niranjini Rajendran, Evelina Busa, Jean Augustinack, Oliver Hinds, B. T Thomas Yeo, Hartmut Mohlberg, Katrin Amunts, and Karl Zilles. Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, Dec 2007.

[57] Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, Verne Caviness, Nikos Makris, Bruce Rosen, and Anders M Dale. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1):11–22, January 2004.

[58] G. Flandin, F. Kherif, X. Pennec, G. Malandain, N. Ayache, and J.-B. Poline. Improved detection sensitivity of functional MRI data using a brain parcellation technique. In *Proceedings 5th International Conference on Medical Image Computing and Computer Assised Intervention*, LNCS 2488 (Part I), pages 467–474, Tokyo, Japan, September 2002. Springer Verlag.

[59] G. Flandin, F. Kherif, X. Pennec, D. Rivière, N. Ayache, and J.-B. Poline. Parcellation of brain images with anatomical and functional constraints for fMRI data analysis. In *Proceedings 1st International Symposium on Biomedical Imaging*, pages 907–910, Washington, DC, July 2002.

[60] G. Flandin, W. Penny, X. Pennec, N. Ayache, and J.-B. Poline. A multisubject anatomo-functional parcellation of the brain. In *Proceedings 9th HBM*, New York,, June19–22 2003.

[61] G. Flandin and W. D. Penny. Bayesian fMRI data analysis with sparse spatial basis function priors. *Neuroimage*, 34(3):1108–1125, February 2007.

[62] J. Fodor. Let your brain alone. *London review of books*, 1999.

[63] Peter T Fox, Angela R Laird, and Jack L Lancaster. Coordinate-based voxel-wise meta-analysis: dividends of spatial normalization. report of a virtual workshop. *Human Brain Mapping*, 25(1):1–5, May 2005.

[64] R. S. J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, C. Price, J. Ashburner, W. Penny, and S. Zeki. *Human Brain Function, Second Edition*. Academic Press, San Diego,, 2003.

[65] C. Fraley and A.E. Raftery. Bayesian regularization for normal mixture estimation and model-based clustering. *Journal of Classification*, 24:155–181, 2007.

[66] L. R. Frank, R. B. Buxton, and E. C. Wong. Probabilistic analysis of functional magnetic resonance imaging data. *Magn Reson Med*, 39(1):132–148, Jan 1998.

[67] Ola Friman and Carl-Fredrik Westin. Resampling fmri time series. *Neuroimage*, 25(3):859–867, Apr 2005.

[68] K. J. Friston, A. P. Holmes, and K. J. Worsley. How many subjects constitute a study? *Neuroimage*, 10(1):1–5, Jul 1999.

[69] K. J. Friston and W. Penny. Posterior probability maps and SPMs. *Neuroimage*, 19(3):1240–1249, July 2003.

[70] K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, and R. N. Henson. A critique of functional localisers. *Neuroimage*, 30(4):1077–1087, May 2006.

[71] K.J. Friston. Statistical parametric mapping. In R.W. Thatcher, M. Hallet, T. Zeffiro, E.R. John, and M. Huerta, editors, *Functional Neuroimaging : Technical Foundations*, pages 79–93, 1994.

[72] K.J. Friston, J. Ashburner, C.D. Frith, J.-B. Poline, J.D. Heather, and R.S.J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189, 1995.

[73] KJ Friston, L Harrison, and W Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, August 2003.

[74] K.J. Friston, W. Penny, C. Phillips, S Kiebel, G. Hinton, and J. Ashburner. Classical and bayesian inference in neuroimaging: Theory. *Neuroimage*, 16(2):465–483, 2002.

[75] C. R. Genovese, D. C. Noll, and W. F. Eddy. Estimating test-retest reliability in functional MR imaging. I: Statistical methodology. *Magn Reson Med*, 38(3):497–507, September 1997.

[76] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, Apr 2002.

[77] G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9:416–429, 1999.

[78] P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9(1):69–86, February 2005.

[79] Polina Golland, Yulia Golland, and Rafael Malach. Detection of spatial activation patterns as unsupervised segmentation of fmri data. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 10(1):110–118, 2007.

[80] Alexandre Gramfort, Benoit Cottereau, Maureen Clerc, Bertrand Thirion, and Sylvain Baillet. Challenging the estimation of cortical activity from MEG with simulatedfmri-constrained retinotopic maps. In *Proc. of the 29th IEEE EMBS Annual international conference*, volume 1, pages 4945–4948, 2007.

[81] Kalanit Grill-Spector, Nicholas Knouf, and Nancy Kanwisher. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci*, 7(5):555–62, May 2004.

[82] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Special issue on variable and feature selection):1157–1182, March 2003.

[83] Kathleen A Hansen, Stephen V David, and Jack L Gallant. Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *Neuroimage*, 23(1):233–241, September 2004.

[84] Stephen Jos Hanson, Toshihiko Matsuka, and James V Haxby. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *Neuroimage*, 23(1):156–166, September 2004.

[85] L. M. Harrison, W. Penny, J. Daunizeau, and K. J. Friston. Diffusion-based spatial priors for functional magnetic resonance images. *Neuroimage*, 41(2):408–423, Jun 2008.

[86] T. Hastie, R. Tibshirani, and J. Friedman, editors. *The Elements of Statistical Learning*. Springer Series in Statistics, 2001.

[87] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, September 2001.

[88] S. Hayasaka and T.E. Nichols. Validating Cluster Size Inference: Random Field and Permutation Methods. *Neuroimage*, 20(4):2343–2356, 2003.

[89] J.-D. Haynes and G. Rees. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, 8:686–691, April 2005.

[90] Ruth Heller, Yulia Golland, Rafael Malach, and Yoav Benjamini. Conjunction group analysis: an alternative to mixed/random effect analysis. *Neuroimage*, 37(4):1178–1185, Oct 2007.

[91] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. Le Goualher., D. L. Collins, A. Evans, G. Malandain, N. Ayache, G. E. Christensen, and H. J. Johnson. Retrospective evaluation of intersubject brain registration. *IEEE Transactions on Medical Imaging*, 22(9):1120–1130, September 2003.

[92] M. Hollander and D.A. Wolfe. *Nonparametric statistical inference*. John Wiley & Sons, New York, USA, second edition edition, 1999.

[93] B. Horwitz, J.M. Rumsey, and BC Donohue. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc Natl Acad Sci*, 95(15):8939–44, June 1998.

[94] Daniel J Jacobsen, Lars Kai Hansen, and Kristoffer Hougaard Madsen. Bayesian model comparison in nonlinear bold fmri hemodynamics. *Neural Comput*, 20(3):738–755, Mar 2008.

[95] S. Jbabdi, M. W. Woolrich, J. L R Andersson, and T. E J Behrens. A Bayesian framework for global tractography. *Neuroimage*, 37(1):116–129, August 2007.

[96] P. Jezzard and R.S. Balaban. Correction for geometric distortion in echo planar images from b_0 field variations. *Magnetic Resonance in Medicine*, 34(1):65–73, July 1995.

[97] Heidi Johansen-Berg, Timothy E.J. Behrens, Emma Sillery, Olga Ciccarelli, Alan J. Thompson, Stephen M. Smith, and Paul M. Matthews. Functional-Anatomical Validation and Individual Variation of Diffusion Tractography-based Segmentation of the Human Thalamus. *Cerebral Cortex*, July 2004.

[98] Eric R. Cosman Jr., John W. Fisher III, and William M. Wells III. Exact map activity detection in f mri using a glm with an ising spatial prior. In *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2004*, pages 703–710, 2004.

[99] Marcel Adam Just, Vladimir L Cherkassky, Timothy A Keller, and Nancy J Minshew. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127(Pt 8):1811–21, August 2004.

[100] Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8:679–685, May 2005.

[101] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, Mar 2008.

[102] M. Keller, A. Roche, A. Tucholka, and B.Thirion. Dealing with spatial normalization errors in fMRI group inference using hierarchical modeling. *Statistica Sinica*, 18(4):1357–1374, 2008.

[103] F. Kherif, G. Flandin, P. Ciuciu, O. Simon, H. Benali, and J.-B. Poline. Model based spatial and temporal similarity measures between series of functional Magnetic Resonance Images. In *Proceedings 5th International Conference on Medical Image Computing and Computer Assised Intervention*, LNCS 2488 (Part II), pages 509–516, Tokyo, Japan, September 2002. Springer Verlag.

[104] F. Kherif, J.-B. Poline, S. Mériaux, H. Benali, G. Flandin, and M. Brett. Group analysis in functional neuroimaging: selecting subjects using similarity measures. *Neuroimage*, 20(4):2197–2208, January 2004.

[105] Ferath Kherif, Jean-Baptiste Poline, Guillaume Flandin, Habib Benali, Olivier Simon, Stanislas Dehaene, and Keith J Worsley. Multivariate model specification for fmri data. *Neuroimage*, 16(4):1068–1083, Aug 2002.

[106] S.J. Kiebel, J.-B. Poline, K.J. Friston, A. P. Holmes, and K. Worsley. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *Neuroimage*, 10(6):756–766, 1999.

[107] Seyoung Kim and Padhraic Smyth. Hierarchical Dirichlet processes with random effects. In *Advances in Neural Information Processing Systems*, Vancouver, 2006.

[108] Seyoung Kim, Padhraic Smyth, and Hal Stern. A nonparametric Bayesian approach to detecting spatial activation patterns in fMRI data. In *Proceedings 9th International Conference on Medical Image Computing and Computer Assised Intervention*, LNCS 4190, pages 217–224, Copenhaguen, 2006. Springer Verlag.

[109] Arno Klein, Jesper Andersson, Babak A Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E Christensen, Louis Collins, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P Woods, J. John Mann, and Ramin V Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, Jan 2009.

[110] Isabelle Klein, Jessica Dubois, J.-F. Mangin, F. Kherif, G. Flandin, J.-B. Poline, Michel Denis, Stephen M Kosslyn, and D. Le Bihan. Retinotopic organization of visual mental images as revealed by functional magnetic resonance imaging. *Brain Res Cogn Brain Res*, 22(1):26–31, December 2004.

[111] Andreas Kleinschmidt. Different analysis solutions for different spatial resolutions? moving towards a mesoscopic mapping of functional architecture in the human brain. *Neuroimage*, 38(4):663–665, Dec 2007.

[112] S. M. Kosslyn, J. Brunn, K. R. Cave, and R. W. Wallach. Individual differences in mental imagery ability: a computational analysis. *Cognition*, 18(1-3):195–243, Dec 1984.

[113] S. M. Kosslyn, A. Pascual-Leone, O. Felician, S. Camposano, J. P. Keenan, W. L. Thompson, G. Ganis, K. E. Sukel, and N. M. Alpert. The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, 284(5411):167–170, Apr 1999.

[114] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(6 Pt 2), June 2004.

[115] N. Kriegeskorte, R. Goebel, and P. Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103:3863–3868, 2006.

[116] Stephen LaConte, Stephen Strother, Vladimir Cherkassky, Jon Anderson, and Xiaoping Hu. Support vector machines for temporal classification of block design fMRI data. *Neuroimage*, 26(2):317–329, Jun 2005.

[117] Angela R Laird, Jack L Lancaster, and Peter T Fox. Brainmap: the social evolution of a human brain mapping database. *Neuroinformatics*, 3(1):65–78, 2005.

[118] T. Lindeberg. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *International Journal of Computer Vision*, 11(3):283–318, 1993.

[119] M. Liou, H.-R. Su, J.-D. Lee, P. E. Cheng, Huang C.-C., and C.-H. Tsai. Bridging functional MR images and scientific inference: Reproducibility maps. *Journal of Cognitive Neuroscience*, 15(7):935–945, 2003.

[120] Michelle Liou, Hong-Ren Su, Juin-Der Lee, John A D Aston, Arthur C Tsai, and Philip E Cheng. A method for generating reproducible evidence in fMRI studies. *Neuroimage*, Oct 2005.

[121] J.S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the Acoustical Society of America*, 89:958–966, 1994.

[122] G. Lohmann and D. Y. von Cramon. Automatic labelling of the human cortical surface using sulcal basins. *Medical Image Analysis*, 4(3):179–188, 2000.

[123] Gabriele Lohmann and Stefan Bohn. Using replicator dynamics for analyzing fMRI data of the human brain. *IEEE Trans Med Imaging*, 21(5):485–492, May 2002.

[124] E. Luders, K.L. Narr, P.M. Thompson, D.E. Rex, R.P. Woods, H. DeLuca, L. Jancke, and A.W. Toga. Gender effects on cortical thickness and the influence of scaling. *Human Brain Mapping*, 27:314–324, 2005.

[125] Oliver Lyttelton, Maxime Boucher, Steven Robbins, and Alan Evans. An unbiased iterative group registration template for cortical surface analysis. *Neuroimage*, 34(4):1535–1544, Feb 2007.

[126] Ranjan Maitra, Steven R Roys, and Rao P Gullapalli. Test-retest reliability estimation of functional MRI data. *mrm*, 48(1):62–70, Jul 2002.

[127] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969, July 2008.

[128] J.-F. Mangin, O. Coulon, and V. Frouin. Robust brain segmentation using histogram scale-space analysis and mathematical morphology. In W. M. Wells, A. Colchester, and S. Delp, editors, *Proceedings 1st International Conference on Medical Image Computing and Computer Assised Intervention*, LNCS-1496, pages 1230–1241, MIT, Boston, October 1998. Springer Verlag.

[129] J.-F. Mangin, V. Frouin, I. Bloch, J. Régis, and J. Lopez-Krahe. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision*, 5:297–318, 1995.

[130] J.-F. Mangin, D. Rivière, O. Coulon, C. Poupon, A. Cachia, Y. Cointepas, J.-B. Poline, D. Le Bihan, J. Régis, and D. Papadopoulos-Orfanos. Coordinate-based versus structural approaches to brain image analysis. *Artificial Intelligence in Medicine*, 30:177–197, 2004.

[131] A.R. McIntosh, F.L. Bookstein, J.V. Haxby, and C.L. Grady. Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*, 3:143–157, 1996.

[132] A.R. McIntosh, C.L. Grady, L.G. Ungerleider, J.V. Haxby, S.I. Rapoport, and B. Horwitz. Network analysis of cortical visual pathways mapped with pet. *J Neurosci.*, 2:655–666, 1994.

[133] MJ McKeown, TP Jung, S Makeig, G Brown, SS Kindermann, TW Lee, and TJ Sejnowski. Spatially independent activity patterns in functional mri data during the stroop color-naming task. *Proc Natl Acad Sci*, 95(3):803–10, February 1998.

[134] J. W. Meyer, N. Makris, J. F. Bates, V. S. Caviness, and D. N. Kennedy. MRI-based topographic parcellation of human brain cerebral white matter. 1. technical foundations. *Neuroimage*, 9:1–17, 1999.

[135] V. Michel, C. Damon, and B. Thirion. Mutual information-based feature selection enhances fMRI brain activity classification. In *5th International Symposium on Biomedical Imaging*, pages 592–595, Paris, France, May 2008.

[136] M.A. Mintun, B.N. Lundstrom, A.Z. Snyder, A.G. Vlassenko, G.L. Shulman, and M.E. Raichle. Blood flow and oxygen delivery to human brain during functional activity: theoretical modeling and experimental data. *Proceedings of the National Academy of Sciences of the United States of America*, 98(12):6859–6864, June 2001.

[137] Yoichi Miyawaki, Hajime Uchida, Okito Yamashita, Masa aki Sato, Yusuke Morito, Hiroki C Tanabe, Norihiro Sadato, and Yukiyasu Kamitani. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60(5):915–929, Dec 2008.

[138] N. Molko, A. Cachia, D. Rivière, J.-F. Mangin, M Bruandet, D. Le Bihan, L. Cohen, and S. Dehaene. Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron*, 40:847–858, Nov 2003.

[139] Kevin Murphy and Hugh Garavan. An empirical investigation into the number of subjects required for an event-related fMRI study. *Neuroimage*, 22(2):879–85, Jun 2004.

[140] S. Mériaux, A. Roche, G. Dehaene-Lambertz, and J.-B. Poline. When do mixed-effect models fail to improve detection sensitivity in fMRI group activation maps? In *Neuroimage (HBM'06)*, Florence, Italy, 2006.

[141] S. Mériaux, A. Roche, G. Dehaene-Lambertz, B. Thirion, and J.-B. Poline. Combined permutation test and mixed-effect model for group average analysis in fMRI. *Human Brain Mapping*, 27(5):402–410, May 2006.

[142] S. Mériaux, A. Roche, B. Thirion, and G. Dehaene-Lambertz. Robust statistics for nonparametric group analysis in fMRI. In *Proceedings 3th International Symposium on Biomedical Imaging*, pages 936–939, Arlington, VA, April 2006.

[143] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. 18(12):1163–1173, 1996.

[144] R. Natarajan and R. E. Kass. Reference bayesian methods for generalized linear mixed models. *Journal of American Statistical Association*, 95(449):227–237, 2000.

[145] R. M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

[146] J. Neumann and G. Lohmann. Bayesian second-level analysis of functional magnetic resonance images. *Neuroimage*, 20(2):1346–1355, 2003.

[147] Jane Neumann, Gabriele Lohmann, Jan Derrfuss, and D. Yves von Cramon. Meta-analysis of functional imaging data using replicator dynamics. *Human Brain Mapping*, 25(1):165–173, May 2005.

[148] Jane Neumann, D. Yves von Cramon, Birte U Forstmann, Stefan Zysset, and Gabriele Lohmann. The parcellation of cortical areas using replicator dynamics in fmri. *Neuroimage*, 32(1):208–219, Aug 2006.

[149] A. Nieto-Castanon, S.S. Ghosh, J.A. Tourville, and F.H. Guenther. Region of interest based analysis of functional imaging data. *Neuroimage*, 19(4):1303–1316, 2003.

[150] K. M. O'Craven and N. Kanwisher. Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *J Cogn Neurosci*, 12(6):1013–1023, Nov 2000.

[151] S. Ogawa, T. Lee, A. Kay, and D. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872, 1990.

[152] B. A. Olshausen and D. J. Field. How close are we to understanding V1? *Neural Computation*, 17(8):1665–99, August 2005.

[153] M. Ono, S. Kubik, and C. D. Abernethey. *Atlas of the Cerebral Sulci*. Georg Thieme Verlag, Stuttgart, 1990.

[154] G. Operto, R. Bulot, J-L. Anton, and O. Coulon. Projection of fmri data onto the cortical surface using anatomically-informed convolution kernels. *Neuroimage*, 39(1):127–135, Jan 2008.

[155] Wanmei Ou and Polina Golland. From spatial regularization to anatomical priors in fMRI analysis. In *IPMI, Glenwood Springs, Colorado*, July 2005.

[156] M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(1):167–172, Jan. 2007.

[157] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Francisco, second edition, 1988.

[158] M. Pelillo and A. Torsello. Payoff-monotonic game dynamics and the maximum clique problem. *Neural Computation*, 18(5):1215–58, May 2006.

[159] Scott J Peltier, Thad A Polk, and Douglas C Noll. Detecting low-frequency functional connectivity in fMRI using a self-organizing map (SOM) algorithm. *Human Brain Mapping*, 20(4):220–6, Dec 2003.

[160] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Comparing dynamic causal models. *Neuroimage*, 22(3):1157–72, July 2004.

[161] W.D. Penny and K.J. Friston. Mixtures of general linear models for functional neuroimaging. *IEEE Transactions on Medical Imaging*, 22(4):504–514, 2003.

[162] Will Penny and Karl Friston. Mixtures of general linear models for functional neuroimaging. *IEEE Transactions on Medical Imaging*, 22(4):504–514, April 2003.

[163] M. Perrin, Y. Cointepas, A. Cachia, C. Poupon, B. Thirion, D. Rivière, P. Cathier, V. El Kouby, A. Constantinesco, D. Le Bihan, and J.-F. Mangin. Connectivity-based parcellation of the cortical mantle using q-ball diffusion imaging. *Int J Biomed Imaging*, 2008.

[164] M. Perrot, D. Rivière, and J.-F. Mangin. Identifying cortical sulci from localizations, shape and local organization. In *5th International Symposium on Biomedical Imaging*, pages 420–423, Paris, France, May 2008.

[165] Mathias Pessiglione, Liane Schmidt, Bogdan Draganski, Raffael Kalisch, Hakwan Lau, Ray J Dolan, and Chris D Frith. How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science*, 316(5826):904–906, May 2007.

[166] M. Piazza, E. Giacomini, D. Le Bihan, and S. Dehaene. Single-trial classification of parallel pre-attentive and serial processes using functional magntic resonance imaging. *Phil. Trans. R. Soc. Lond. B*, 270:1237–1245, 2003.

[167] P. Pinel, B. Thirion, S. Mériaux, A. Jobert, J. Serres, D. Le Bihan, J.-B. Poline, and S. Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci*, 8(1):91, Oct 2007.

[168] J.-B. Poline, A. P. Holmes, K. Worsley, and K.J. Friston. *Making Statistical inferences, in Human Brain Function*, chapter 5, pages 85–106. Academic Press, 1997.

[169] C. Poupon, J.-F. Mangin, C. A. Clark, V. Frouin, J. Régis, D. Le Bihan, and I. Bloch. Towards inference of the human brain connectivity from MR diffusion tensor data. *Medical Image Analysis*, 5:1–15, 2001.

[170] F. Poupon, J.-F. Mangin, D. Hasboun, I. Magnin, and V. Frouin. Multi-object Deformable Templates Dedicated to the Segmentation of Brain Deep Structures. In *Proceedings 1st International Conference on Medical Image Computing and Computer Assised Intervention*, LNCS-1496, pages 1134–1143, MIT, Boston, October 1998. Springer Verlag.

[171] Carl Edward Rasmussen. A new technique for elastic registration of tomographic images. In *Advances in Neural Information Processing Systems 12 S.A. Solla, T.K. Leen and K.-R. Müller (eds.)*, pages 554–560, 2000.

[172] J.J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima. A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signal. *Neuroimage*, 21:547–567, 2004.

[173] D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin, and J. Régis. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical Image Analysis*, 6(2):77–92, 2002.

[174] D. Rivière, D. Papadopoulos-Orfanos, C. Poupon, F. Poupon, O. Coulon, J.-B. Poline, V. Frouin, J. Régis, and J.-F. Mangin. A structural browser for human brain mapping. In *Proceedings 6th HBM*, NeuroImage 11(5), page 912, San Antonio, Texas, 2000.

[175] A. Roche, S. Mériaux, M. Keller, and B. Thirion. Mixed-effects statistics for group analysis in fMRI: A nonparametric maximum likelihood approach. *Neuroimage*, 38:501–510, 2007.

[176] P. E. Roland, S. Geyer, K. Amunts, T. Schormann, A. Schleicher, A. Malikovic, and K. Zilles. Cytoarchitectural maps of the human brain in standard anatomical space. *Human Brain Mapping*, 5(4):222–227, 1997.

[177] F. Rossi, A. Lendasse, D. François, V. Wertz, and M. Verleysen. Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemometrics and intelligent laboratory systems*, 80:215–226, 2006.

[178] S. Sandor and R. Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Transactions on Medical Imaging*, 16(1):41–54, February 1997.

[179] Rebecca Saxe, Matthew Brett, and Nancy Kanwisher. Divide and conquer: A defense of functional localizers. *Neuroimage*, 30(4):1088–1096, May 2006.

[180] Armin Schwartzman, Robert F Dougherty, Jongho Lee, Dara Ghahremani, and Jonathan E Taylor. Empirical null and false discovery rate analysis in neuroimaging. *Neuroimage*, 44(1):71–82, Jan 2009.

[181] B. Schölkopf and A.J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, 2002.

[182] Bernard Schölkopf and Alexander J. Smola. *Learning with kernels. Support vector machines, regularization, optimzation and beyond*. Adaptive computation and machine learning. The MIT Press, Cambridge, Massachussets, 2002.

[183] M. I. Sereno, A. M. Dale, J. B. Reppas, K. K. Kwong, J. W. Belliveau, B. R. Rosen, and R. B. Tootell. Borders of multiple visual areas in humans revealed by fonctional magnetic resonance imaging. *Science*, 268:889–893, 1995.

[184] David W Shattuck, Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L Narr, Russell A Poldrack, Robert M Bilder, and Arthur W Toga. Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–1080, Feb 2008.

[185] Amir Shmuel, Mark Augath, Axel Oeltermann, and Nikos K Logothetis. Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. *Nat Neurosci*, 9(4):569–577, Apr 2006.

[186] Amir Shmuel, Essa Yacoub, Josef Pfeuffer, Pierre Francois Van de Moortele, Gregor Adriany, Xiaoping Hu, and Kamil Ugurbil. Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, 36(6):1195–1210, Dec 2002.

[187] Olivier Simon, F. Kherif, G. Flandin, J.-B. Poline, D. Rivière, J.-F. Mangin, D. Le Bihan, and S. Dehaene. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *Neuroimage*, 23(3):1192–1202, 11 2004.

[188] Scott D Slotnick, William L Thompson, and Stephen M Kosslyn. Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, 15(10):1570–1583, Oct 2005.

[189] A. T. Smith, K. D. Singh, A. L. Williams, and M. W. Greenlee. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, 11(12):1182–1190, Dec 2001.

[190] Stephen M Smith and Thomas E Nichols. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1):83–98, Jan 2009.

[191] David A Soltysik, Kyung K Peck, Keith D White, Bruce Crosson, and Richard W Briggs. Comparison of hemodynamic response nonlinearity across primary cortical areas. *Neuroimage*, 22(3):1117–1127, Jul 2004.

[192] Klaas Enno Stephan, Nikolaus Weiskopf, Peter M. Drysdale, Peter A. Robinson, and Karl J. Friston. Comparing hemodynamic models with dcm. *Neuroimage*, 38(3):387–401, Nov 2007.

[193] Peter Stiers, Ronald Peeters, Lieven Lagae, Paul Van Hecke, and Stefan Sunaert. Mapping multiple visual areas in the human brain with a short fMRI sequence. *Neuroimage*, 29(1):74–89, Jan 2006.

[194] John D. Storey. The positive false discovery rate: a bayesian interpretation and the q-value. *Ann. Statist.*, 31(6):2013–2035, 2003.

[195] Z. Y. Sun, D. Rivière, , E. Duchesnay, B. Thirion, F. Poupon, and J.-F. Mangin. Defining cortical sulcus patterns using partial clustering based on bootstrap and bagging. In *5th International Symposium on Biomedical Imaging*, pages 1629–1632, Paris, France, May 2008.

[196] J. Talairach and P. Tournoux. *Co–Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System : An Approach to Cerebral Imaging.* Thieme Medical Publishers, Inc., Georg Thieme Verlag, Stuttgart, New York, 1988.

[197] Xiaodong Tao, Xiao Han, Maryam E. Rettmann, Jerry L. Prince, and Christos Davatzikos. Statistical study on cortical sulci of human brains. In *IPMI*, pages 475–487, 2001.

[198] B. Thirion, B. Bascle, V. Ramesh, and N. Navab. Fusion of color, shading and boundary information for factory pipe segmentation. In B. Bascle, editor, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 349–356 vol.2, 2000.

[199] B. Thirion, S. Dodel, and J.-B. Poline. Detection of signal synchronizations in resting-state fMRI datasets. *Neuroimage*, 29(1):321–327, January 2006.

[200] B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J.-B. Poline, and S. Dehaene. Reading the brain visual system as an inverse problem. In *Proceedings 3th International Symposium on Biomedical Imaging*, Washington, USA, April6-9 2006.

[201] B. Thirion and O. Faugeras. Feature detection in fMRI data: The information bottleneck approach. In *Proceedings 6th International Conference on Medical Image Computing and Computer Assised Intervention*, pages 83–91, Montreal Canada, November 2003.

[202] B. Thirion and O. Faugeras. Nonlinear dimension reduction of fMRI data: the Laplacian embedding approach. In *Proceedings 2st International Symposium on Biomedical Imaging*, pages 372–375, Arlington, VA, April 2004.

[203] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Human Brain Mapping*, 27(8):678–693, August 2006.

[204] B. Thirion, P. Pinel, S. Mériaux, A. Roche, S. Dehaene, and J.-B. Poline. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35(1):105–120, 2007.

[205] B. Thirion, P. Pinel, and J.-B. Poline. Finding landmarks in the functional brain: detection and use for group characterization. In *MICCAI'05*, volume 8, pages 476–483, 2005.

[206] B. Thirion, P. Pinel, A. Tucholka, A. Roche, P. Ciuciu, J.-F. Mangin, and J.-B. Poline. Structural analysis of fMRI data revisited: Improving the sensitivity and reliability of fMRI group studies. *IEEE Transactions on Medical Imaging*, 26(9):1256–1269, September 2007.

[207] B. Thirion, A. Roche, P. Ciuciu, and J.-B. Poline. Improving sensitivity and reliability of fMRI group studies through high level combination of individual subjects results. In *Proceedings MMBIA2006*, New York, USA, June17-18 2006.

[208] B. Thirion, B. Thirion, and O. Faugeras. Revisiting non-parametric activation detection on fmri time series. In O. Faugeras, editor, *Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis MMBIA 2001*, pages 121–128, 2001.

[209] B. Thirion, B. Thirion, and O. Faugeras. Dynamical components analysis of fMRI data. In O. Faugeras, editor, *Proc. IEEE International Symposium on Biomedical Imaging*, pages 915–918, 2002.

[210] B. Thirion, A. Tucholka, M. Keller, P. Pinel, A. Roche, J.-F. Mangin, and J.-B. Poline. High level group analysis of FMRI data based on Dirichlet process mixture models. In *International Conference on Information Processing in Medical Imaging*, volume 4584 of *LNCS*, pages 482–494, 2007.

[211] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, J.-B. Poline, D. Le Bihan, and S. Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4):1104–1116, December 2006.

[212] Bertrand Thirion and Olivier Faugeras. Dynamical components analysis of fMRI data through kernel PCA. *Neuroimage*, 20(1):34–49, Sep 2003.

[213] Bertrand Thirion and Olivier Faugeras. Feature characterization in fMRI data: the information bottleneck approach. *Medical Image Analysis*, 8(4):403–419, December 2004.

[214] Bertrand Thirion and Olivier D. Faugeras. Dynamical components analysis of fmri data: A second order solution. In *EUROCAST*, pages 552–563, 2003.

[215] P.M. Thompson, K.M. Hayashi, G. de Zubicaray, A.L. Janke, S.E. Rose, J. Semple, D. Herman, M.S. Hong, S.S. Dittmer, and A.W. Doddrell, D.M. Toga. Dynamics of gray matter loss in alzheimer's disease. *Journal of neuroscience*, 23(3):994–1005, 2003.

[216] PM Thompson, C Schwartz, and AW Toga. High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain. *Neuroimage*, 3(1):19–34, Feb 1996.

[217] B. Thyreau, B. Thirion, G. Flandin, and J.-B. Poline. Anatomo-functional description of the brain: a probabilistic approach. In *Proceedings 31th Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, volume V, pages 1109–1112, Toulouse, France, May 2006.

[218] M. Tipping. The relevance vector machine. In *Advances in Neural Information Processing Systems, San Mateo, CA*. Morgan Kaufmann, 2000.

[219] R. B. Tootell, A. Dale, M. Sereno, and R. Malach. New images from human visual cortex. *Trends in NeuroScience*, 19:481–489, 1996.

[220] R. B. Tootell, N. K. Hadjikhani, W. Vanduffel, A. K. Liu, J. D. Mendola, M. I. Sereno, and A. M. Dale. Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3):811–817, Feb 1998.

[221] R. B. Tootell, D. Tsao, and W. Vanduffel. Neuroimaging weighs in: Humans meet macaques in "primate" visual cortex. *J. Neurosci.*, 23:3981–3989, 2003.

[222] R. Toro. *ontogenesis of the cortical organisation: modelling and analysis*. PhD thesis, Paris VI, 2003.

[223] Roberto Toro, Peter T Fox, and Tomás Paus. Functional coactivation map of the human brain. *Cereb Cortex*, Feb 2008.

[224] A. Tucholka, B. Thirion, M. Perrot, P. Pinel, J.-F. Mangin, and J.-B. Poline. Probabilistic anatomo-functional parcellation of the cortex: how many regions? In *11thProceedings MICCAI, LNCS Springer Verlag*, New-York, USA, 2008.

[225] A. Tucholka, B. Thirion, P. Pinel, J.-B. Poline, and J.-F. Mangin. Triangulating cortical functional networks with anatomical landmarks. In *5th International Symposium on Biomedical Imaging*, pages 612–615, Paris, France, May 2008.

[226] 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(1):273–89, Jan 2002.

[227] Dimitri Van De Ville, Thierry Blu, and Michael Unser. Integrated wavelet processing and spatial statistical testing of fMRI data. *Neuroimage*, 23(4):1472–1485, December 2004.

[228] S. Vanni, L. Henriksson, and A. C. James. Multifocal fMRI mapping of visual cortical areas. *Neuroimage*, 27(1):95–105, Aug 2005.

[229] Dimitri Van De Ville, Mohamed L Seghier, François Lazeyras, Thierry Blu, and Michael Unser. Wspm: wavelet-based statistical parametric mapping. *Neuroimage*, 37(4):1205–1217, Oct 2007.

[230] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):1583–598, 1991.

[231] T. Vincent, P. Ciuciu, and J. Idier. Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. In *32th Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, volume I, pages 325–328, Honolulu, Hawaii, April 2007.

[232] T. Vincent, P. Ciuciu, and B. Thirion. Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI. In *5th International Symposium on Biomedical Imaging*, pages 568–571, Paris, France, May 2008.

[233] T.D. Wager, M.C. Keller, S.C. Lacey, and J. Jonides. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage*, 26(1):99–113, 2005.

[234] Yalin Wang, Xianfeng Gu, Kiralee M Hayashi, Tony F Chan, Paul M Thompson, and Shing-Tung Yau. Brain surface parameterization using riemann surface structure. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 8(Pt 2):657–665, 2005.

[235] J. Warnking, M. Dojat, A. Guerin-Dugue, C. Delon-Martin, S. Olympieff, N. Richard, A. Chehikian, and C. Segebarth. fMRI retinotopic mapping - step by step. *Neuroimage*, 17:1665–1683, 2002.

[236] Xingchang Wei, Seung-Schik Yoo, Chandlee C Dickey, Kelly H Zou, Charles R G Guttmann, and Lawrence P Panych. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *Neuroimage*, 21(3):1000–8, Mar 2004.

[237] Christopher K. I. Williams and Matthias Seeger. Using the nystrm method to speed up kernel machines. In *Advances in Neural Information Processing Systems 13*, pages 682–688. MIT Press, 2001.

[238] M. Woolrich and T. Behrens. Variational Bayes inference of spatial mixture models for segmentation. *IEEE Transactions on Medical Imaging*, 25(10):1380–1391, October 2006.

[239] M. Woolrich, T. Behrens, Ch. Beckmann, and S. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Transactions on Medical Imaging*, 24(1):1–11, January 2005.

[240] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith. Fully Bayesian spatio-temporal modelling of fMRI data. *IEEE Transactions on Medical Imaging*, 23(2):213–231, February 2004.

[241] M. Woolrich, B. Ripley, M. Brady, and S. Smith. Temporal autocorrelation in univariate linear modelling of fMRI data. *Neuroimage*, 14(6):1370–1386, December 2001.

[242] K. J. Worsley. An improved theoretical P value for SPMs based on discrete local maxima. *Neuroimage*, 28(4):1056–1062, December 2005.

[243] K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fMRI data. *Neuroimage*, 15(1):1–15, January 2002.

[244] Nicolas Wotawa, Bertrand Thirion, Eric Castet, Jean-Luc Anton, and Olivier Faugeras. Human retinotopic mapping using fMRI. Technical Report 5472, INRIA Sophia-Antipolis, 2005.

[245] J.S. Yedidia, Yair Weiss, and William T. Freeman. Understanding belief propagation and its generalizations. In *International Joint Conference on Artificial Intelligence*, 2001.

[246] Jerrold H. Zar. *Biostatistical Analysis*. Prentice-Hall, Inc., Upper Saddle River, NJ., 1999.

[247] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.