Template estimation for arbitrary alignments : application to brain imaging

Thomas Bazeille

Thèse de doctorat

NNT : 2021UPASG071

université PARIS-SACLAY

Template estimation for arbitrary alignments: application to brain imaging

Estimation de modèle après des alignements arbitraires : application à l'imagerie médicale

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de l'Information et de la Communication (ED STIC)
Spécialité de doctorat: Mathématiques et Informatique
Unité de recherche : Université Paris-Saclay, Inria, Inria Saclay-Île-de-France, 91120, Palaiseau, France.
Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris-Saclay, le 20/10/2021, par

Thomas BAZEILLE

Composition du Jury

Michèle SEBAG Directrice de Recherche, CNRS (Univ. Paris-Saclay, INRIA, TAU)	Présidente
Oluwasanmi KOYEJO Professeur, University of Illinois at Urbana-Champaign (CS)	Rapporteur & Examinateur
Christophe PHILLIPS Professeur, Université de Liège (EECS, GIGA)	Rapporteur & Examinateur
Florent MEYNIEL Directeur de recherche, Univ. Paris-Saclay (INSERM, CEA)	Examinateur
Jonathan PILLOW Professeur, Princeton (Princeton Neuroscience Institute)	Examinateur

Direction de la thèse

Bertrand THIRION Directeur de Recherche, INRIA (Parietal)	Directeur de thèse

TEMPLATE ESTIMATION FOR ARBITRARY ALIGNMENTS: APPLICATION TO BRAIN IMAGING

THOMAS BAZEILLE

Dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Université Paris Saclay
Sciences et Technologies de l'Information et de la Communication

parietal team, inria saclay / neurospin, cea, france
2018–2021

ABSTRACT

An important goal of cognitive neurosciences is to understand the functional organization of the brain. It heavily relies on Functional Magnetic Resonance Imaging (fMRI), a powerful tool to investigate the link between brain function and anatomical structures at a high spatial-resolution. *Functional inter-individual variability* is a major obstacle limiting functional brain mapping precision and generalizability of results obtained in neuroimaging studies. This variability, observed across subjects performing the same task, goes far beyond anatomical variability in brain shape and size. In this work, we focus on a class of methods designed to address functional variability, namely *functional alignment*. These methods match subjects neural signals based on their functional similarity.

In a first part, we review standard functional brain mapping paradigms and techniques, as well as the challenges induced by functional variability. We additionally review existing functional alignment methods and related work, and discuss the current limitations of these approaches. In a second part, we develop a new functional alignment method, based on *optimal transport*—a mathematical theory interested in matching probability distributions while taking their geometry into account. Functional alignment methods are local, which means that many local alignments need to be aggregated to compose whole-brain alignments. Moreover, these methods derive pairwise matching and call for a *"functional template"*, a common functional representation to which all subjects of a study can be aligned. To overcome limitations of existing solutions, we additionally introduce a new aggregation scheme as well as a principled template design procedure. In a third part, we turn to empirical validation of alignment performance. Indeed, these methods are seldom used in applied studies, and it is unclear to what extent they can address functional variability in typical cognitive studies. We investigate their performance to improve generalization of predictive models to new subjects. In this *inter-subject decoding* set-up, spanning four different datasets, we show that alignment methods hold real potential to recover an important share of prediction accuracy lost due to inter-subject variability.

SYNTHÈSE EN FRANÇAIS

Un objectif important des neurosciences cognitives est de comprendre l'organisation fonctionnelle du cerveau. Pour cela, elles s'appuient sur l'Imagerie par Résonance Magnétique fonctionnelle (IRMf), un outil puissant pour étudier le lien entre les fonctions cérébrales et les structures anatomiques sous-jacentes à une haute résolution spatiale. La *variabilité fonctionnelle interindividuelle* est un obstacle majeur qui limite la précision de la cartographie fonctionnelle du cerveau et la généralisation des résultats obtenus par les études d'imagerie cérébrale. Cette variabilité importante, observable entre des sujets effectuant la même tâche cognitive, va bien au-delà des variations de taille et de forme des structures anatomiques. Elle crée des variations importantes dans la localisation, l'intensité et la géométrie des activations fonctionnelles. Dans cette thèse, nous nous intéressons à un ensemble de méthodes conçues pour traiter la variabilité fonctionnelle : *l'alignement fonctionnel*. Ces méthodes mettent en correspondances les signaux cérébraux de différents sujets sur la base de leur similarité fonctionnelle.

Dans un premier temps, nous présentons les concepts et les techniques usuels pour la cartographie fonctionnelle cérébrale, ainsi que les difficultés induites par la variabilité fonctionnelle. Nous passons également en revue les méthodes d'alignement fonctionnel existantes. Ces méthodes cherchent des transformations mettant en correspondances les pixels volumiques(voxels) des images d'un individu avec ceux d'un autre en fonction de la similarité leurs profils d'activité à travers différentes images. L'hyperalignment, méthode historique d'alignement fonctionnel, cherche une transformation orthogonale permettant de maximiser cette similarité. Une seconde méthode répandue est le Shared Response Model (SRM) qui trouve une décomposition des données de chaque sujet en une base orthogonale individuelle (des cartes spatiales orthogonales représentant divers *modes d'activation* du cerveau) et des décours temporels partagés à travers les sujets pour chacune des cartes de cette base. Nous passons également en revue les nombreuses méthodes proposées dans la littérature pour réaliser des alignements fonctionnels et soulignons les questions méthodologiques qui freinent leur utilisation dans la recherche en sciences cognitives.

Dans une deuxième partie, nous proposons une nouvelle méthode d'alignement fonctionnel, basée sur le *transport optimal* - une théorie mathématique qui s'intéresse à la mise en correspondance de distributions de probabilité prenant en compte leur géométrie. Cet algorithme permet de trouver un alignement parcimonieux qui transporte les voxels d'un sujet à l'autre en minimisant la dissimilarité induite par

ce transport. Ce *plan de transport* préserve la spécificité fonctionnelle des signaux après l'alignement.

Les méthodes d'alignement fonctionnel sont définies localement. Pour construire des alignements à l'échelle du cerveau entier, nous proposons un algorithme d'alignement *par parcelle* qui concatène de nombreux alignements appris sur des parcelles locales adjacentes mais disjointes. Par ailleurs, ces mises en correspondances sont définies entre deux sujets. Il est nécessaire, pour aligner plusieurs sujets, de recourir à un *"modèle fonctionnel de groupe"* : une représentation commune sur laquelle tous les sujets peuvent être alignés. Nous proposons un algorithme de minimisation pour inférer un optimum local de ce modèle fonctionnel, conjointement avec les alignements de chaque sujet vers cet espace. Nous évaluons la performance de ces méthodes en utilisant des tâches de prédiction de nouvelles images.

La troisième partie de ce travail est consacrée à une validation expérimentale approfondie de ces méthodes. En effet, l'alignement est rarement utilisé dans des études applicatives et sa capacité à compenser la variabilité fonctionnelle en pratique est peu documentée. Nous évaluons sa capacité à améliorer la généralisation de modèles prédictifs à de nouveaux sujets. Dans cette configuration— dite de *"décodage inter-sujet"*— appliquée à quatre jeux de données, nous montrons que les méthodes d'alignement ont le potentiel de compenser une part importante des effets de la variabilité interindividuelle, dans une grande variété de situations. Cette étude comparative nous permet également de tirer de nombreuses conclusions méthodologiques. Nous montrons par exemple la performance significative du Transport Optimal et du Shared Response Model pour réaliser des alignements ainsi que l'efficacité de l'alignement par parcelle pour pouvoir réaliser des alignements à l'échelle du cerveau entier. Nous avons publié des implémentations *open-source* optimisées de chaque méthode proposée lors de cette thèse.

ACKNOWLEDGMENTS

I warmly thank Pr. Christophe Phillips, Pr. Sanmi Koyejo, Pr. Michèle Sebag, Pr. Florent Meyniel and Pr. Jonathan Pillow, for accepting to review this manuscript and be part of my defense jury.

I feel really lucky to have conducted this work under the supervision of Bertrand Thirion. I'll stay deeply humbled and inspired by his scientific knowledge spanning so many complex topics, his impeccable work and scientific ethic, as well as his dedicated and kind guidance. I want to extend my warm thanks to all people with whom I collaborated during this PhD and made these works possible, including Elizabeth Dupre, Hugo Richard, Hicham Janati, Alexandre Gramfort and Jean-Baptiste Poline. Special thanks go to Elizabeth, who was a great collaborator and friend. We shared a great number of both exciting and frustrating projects, but she was always there to bring them forward with a smile. I also want to thank Hugo for sharing implementation problems as well as good times. I extend my thanks to people behind Nilearn and IBC, especially: Ana-Luisa Pinho, Swetha Shankar, Nicolas Gensollen and Gaël Varoquaux.

I have a grateful thought for all research friends that made this PhD enjoyable! It was a pleasure to spend (rare) conferences, Parietal retreats, weekly meetings, coffee breaks and after-works with such fun and diverse people. I especially want to thank Neurospin lab-mates who made the long transportation time to Saclay worth it (when offices were still a thing): Jérôme-Alexis Chevalier, Hamza Cherkaoui, Zaccharie Ramzi, Alexis Thual, Binh Nguyen and Alexandre Pasquiou as well as the former ones : Jérôme Dockès, Loubna El Gueddari, Kamalakar Reddy Dadi, Juan Jesus Torre, Daria La Rocca and Darya Chyzhyk. And more broadly, I want to thank all Parietal and INRIA friends especially: Nicolas Berkouk, Pierre Ablin, Thomas Moreau, Antonia Machlouzarides, Patricio Cerda, Hubert Jacob Banville, David Sabbagh, Maëliss Jallais, Pierre Glaser, Meyer Scetbon, Arthur Mensch, Elvis Dohmatob and Carole Lazarus. I also want to thank Olivier Grisel, Gaël Varoquaux, Alexandre Gramfort and Demian Wasserman for interesting discussions. I also thank Digiteo for funding my thesis and Corinne Petitot and Stéphanie Druetta for helping me navigate administrative processes.

I have a grateful thought for my mother, my sister, and my friends for being there during these years and all others too. Finally, I deeply thank Tania, who inspires me every day.

CONTENTS

ACRONYMS

AP	antero-posterior
AR(1)	Auto-Regressive model of order 1
BOLD	Blood Oxygenation Level Dependent
CCA	Canonical Correlation Analysis
EPI	Echo-planar imaging
fMRI	functional Magnetic Resonance Imaging
fwhm	full-width-at-half-maximum
GLM	General Linear Model
HCP	Human Connectome Project
HRF	Haemodynamic Response Function
i.i.d.	independent identically distributed
IBC	Individual Brain Charting
ICA	Independent Component Analysis
IRMf	Imagerie à résonance magnétique fonctionnelle
MEG	Magnetoencephalography
MNI	Montreal Neurological Institute
MRI	Magnetic Resonance Imaging
MVPA	Multivariate Pattern Analysis
OLS	Ordinary Least Squares
OT	Optimal Transport
PA	posterior-anterior
PCA	Principal Component Analysis
ROI	Region of Interest
RSA	Representational Similarity Analysis
RSVP	Rapid-Serial-Visual-Presentation
SNR	signal-to-noise ratio
SRM	Shared Response Model
SPM	Statistical Parametric Mapping
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SyN	Symmetric Image Normalization

1 OVERVIEW

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive functional brain imaging technique with a millimeter-order spatial resolution. Over the last decades, it has played a pivotal role in *functional brain mapping*: the investigation of brain function organization principles. Thousands of studies using fMRI are published every year, with focuses ranging from understanding general brain organization and how it varies, to identifying patterns of brain activity supporting specialized functional processes.

A key stake for brain mapping is to ensure generalizability and consistency of these results, despite the strong inter-subject variability that is observed in functional images. This *functional variability* stems both from imaging limitations (low signal-to-noise ratio, small sample size) and from natural brain variation across individuals. Although standard fMRI preprocessing includes an anatomical registration step—ensuring the correspondence of individual brain structures to a *template* brain—there remains a high degree of variability in observed functional data, which undermines attempts at precise brain mapping. Variability is a key hurdle for population comparison, which explains why fMRI imaging is seldom used in clinical settings. In general, it remains hard to relate our general knowledge on brain function to a specific patient through imaging.

In this thesis, we focus on a class of methods specifically targeting inter-subject functional variability, *functional alignment*, which take advantage of the rich individual characterization brought by intensive individual scanning experiments to learn data-driven mappings between individual functional responses. We propose a new method, based on *optimal transport*, that enhances alignment performance as well as methodological developments to improve efficiency and usability of this framework. We especially focus on assessing potential benefits of alignment for cognitive research studies and present a thorough methodological benchmark. We release open-source implementations of all the methods introduced. This work opens the door to a broader adoption of functional alignment as a common fMRI analysis step, which will hopefully help progress towards more accurate brain mapping. In this chapter, we outline the structure of this manuscript.

1.1 ORGANIZATION OF THIS MANUSCRIPT

1.1.1 *An introduction to brain mapping and functional alignment*

In the first part, we introduce concepts and techniques underlying functional brain mapping and functional alignment. We give a quick account of functional Magnetic Resonance Imaging principles and standard approach to acquisition and analysis of fMRI data. We present common paradigms of brain function investigation and recent evolutions towards deep characterization of individual functional organization. We proceed with an introduction of seminal functional alignment methods, such as *Hyperalignment* and *Shared Response Modelling*. We frame alignment in the broader machine learning literature and review related approaches for fMRI. Finally, we put these works in context to highlight roadblocks hindering a broad dissemination of alignment methods.

1.1.2 *Alignment methodological developments*

In the second part, we introduce several methodological developments targeted at improving alignment performance and efficiency. We bridge the key idea of alignment—improving correspondence of activation patterns across individuals—with Optimal Transport: an acclaimed mathematical theory which studies matching of probability distributions. We also introduce novel schemes to derive whole-brain matchings from local alignments and to estimate a *functional template*: a well-suited functional representation to which several subjects can be aligned. We evaluate those contributions using several datasets comprising rich individual data.

1.1.3 *Empirical benchmark*

In the third part, we focus on assessing the benefits to be expected from alignment in standard functional imaging studies. To do so, we focus on *inter-subject decoding*, a predictive modelling technique yielding a principled way to measure how much inter-subject variability can be compensated for using functional alignment while preserving signal specificity for later analysis. This evaluation compares five alignment methods on four datasets comprising five decoding tasks. We additionally discuss several common methodological and implementation choices. This study confirms that the best alignment methods are able to recover more than half of decoding accuracy usually lost due to inter-subject variability.

1.1.4 *Conclusion*

In the conclusion, we summarize our contributions and the main outcomes of our empirical benchmark. We also describe the challenges we faced during this thesis, as well as current limitations and future perspectives for this line of work. Additionally, we recapitulate the alignment open-source implementations produced when performing this work, and additional projects conducted during this thesis.

1.2 A NOTE ON CHAPTER ORDERING

To get a broad overview of functional brain mapping concepts and techniques, one should start with chapter 2. A reader especially interested in understanding functional alignment intuitions and its benefits should focus first on chapters 3 and 6. In contrast, chapters 4 and 5 are especially focused on introducing an original mathematical framework for alignment and outlining challenges and directions for building a functional template. Section 6.5 may be skipped at first read, as it is detailing additional analyses which support benchmark conclusions summarized in Section 6.4.

1.3 RELATED PUBLICATIONS

Contributions developed in Chapters 4 & 5 were partly published in a conference paper (Bazeille et al., 2019). Work detailed in Chapter 6 is a journal article (Bazeille et al., under review) currently under review at *Neuroimage*. Additional collaborations detailed in Chapter 7 gave rise to several conference and journal articles (Janati et al., 2019, 2020; Torre et al., in preparation).

Bazeille, Thomas, Elizabeth Dupre, Hugo Richard, Jean-Baptiste Poline, and Bertrand Thirion (under review). "An empirical evaluation of functional alignment using inter-subject decoding." In: URL: https://www.biorxiv.org/content/10.1101/2020.12.07.415000v2.

Bazeille, Thomas, Hugo Richard, Hicham Janati, and Bertrand Thirion (2019). "Local Optimal Transport for Functional Brain Template Estimation." In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 237–248.

Janati, Hicham, Thomas Bazeille, Bertrand Thirion, Marco Cuturi, and Alexandre Gramfort (2019). "Group level MEG/EEG source imaging via optimal transport: minimum Wasserstein estimates." In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 743–754.

Janati, Hicham, Thomas Bazeille, Bertrand Thirion, Marco Cuturi, and Alexandre Gramfort (2020). "Multi-subject MEG/EEG source imaging with sparse multi-task regression." In: *NeuroImage* 220, p. 116847.

Torre, Juan Jesús, Ana Luísa Pinho, Swetha Shankar, Alexis Amadon, Mani Saignavongs, Marcela Perrone-Bertolotti, Thomas Bazeille, Elvis Dohmatob, Isabelle Denghien, Chantal Ginisty, Séverine Becuwe-Desmidt, Séverine Roger, Yann Lecomte, Valérie Berland, Laurence Laurier, Véronique Joly-Testault, Gaelle Mediouni-Cloarec, Christine Doublé, Bernadette Martins, Jean-Philippe Lachaux, Patrick Bisset, Ayse Zeynep Enkavi, Ian Eisenberg, Russel Poldrack, Roberto Santoro, Elia Formisano, Gael Varoquaux, Stanislas Dehaene, Lucie Hertz-Pannier, and Bertrand Thirion (in preparation). "Individual Brain Charting fourth dataset extension, high-resolution fMRI data for cognitive mapping."

Part I

CONTEXT

2 FUNCTIONAL BRAIN MAPPING

This chapter provides an introduction to functional brain mapping. Section 2.1 provides a short technical background on Magnetic Resonance Imaging (MRI) and common methodologies used in human brain function imaging. Section 2.2 introduces central paradigms underlying ***brain mapping*** *studies. Section 2.3 gives an account of* ***functional variability****, one of the main challenges of functional MRI (fMRI) studies today, and explains its consequences on modern brain mapping.*

2.1 BRAIN IMAGING USING MRI

2.1.1 *MRI principle*

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that produces images of organs and their physiological processes. For anatomical purposes, it uses variations in magnetic susceptibility of different molecules to contrast tissues of different composition. In most applications, the signals exploited are the radiofrequency signal emitted by the synchronized spinning of nuclei, most commonly hydrogen nuclei (i.e., protons) which generate transverse magnetization.

These protons exhibit different response to a magnetic field, depending on the environment they are embedded in, e.g., different tissues have different relaxation times. As a first step, nuclei are aligned using a constant and intense static magnetic field. Additional gradient fields (or gradients) are then applied to encode spatial position of voxels. Gradients are small linear spatial variations of magnetic field in x, y, z that are used to produce spatial resolution.

Finally, a radiofrequency pulse is emitted to tilt the nuclei magnetization in the plane perpendicular to the intense static field. When this pulse ends, the spin of hydrogen nuclei relax to their equilibrium states. This relaxation occurs with certain characteristic time constants. The relaxation signal (i.e., spinning of the transverse magnetization) is measured with receptive coils. The response emitted by nuclei contained in each volumic pixel (or voxel) can then be disentangled, knowing precisely the local field that they were subject to. For each voxel, the measure retained is the observed relaxation time. Radiofrequency pulses can be emitted at different timings in so-called sequences. A given sequence produces a particular contrast, for which various type

of tissues will have a dominant influence (e.g., T1-weighted for structural MRI, T2-weighted for functional MRI, diffusion-weighted (Le Bihan et al., 2001) and many more). For structural MRI, this process enables to create images that characterize local magnetic susceptibility and hence tissue composition, at a millimeter scale.

2.1.2 *Functional MRI*

While structural MRI can be used to image anatomical structures of the brain in a static fashion, many applications of MRI actually focus on the dynamical processes ongoing in the brain and how they relate to behavior. It has been known for a long time that an important correlate of brain activity is blood flow (Pauling and Coryell, 1936). Neurons that activate are depleted of oxygen and glucose, which are important primitives of adenosine triphosphate (ATP), a crucial molecule used to produce energy necessary for their biological functions (Vergara et al., 2019). Local brain activation is thus followed by an inflow of oxygenated blood, replenishing these cells and displacing deoxygenated blood away. This phenomenon is called the *neurovascular coupling*.

It was later discovered that oxygenated and deoxygenated blood carry different magnetic susceptibilities, and thus react differently when placed in a magnetic field. Seiji Ogawa and colleagues reckoned in 1990 (Ogawa et al., 1990) that this change in the Blood Oxygenation Level Dependent (BOLD) signal could be used to measure a correlate of brain activity through MRI: the hemodynamic response. This HR shows an increase in BOLD signal, lagging a few seconds behind the neural activation itself, followed by a rapid decrease and a slight undershoot of BOLD signal once the equilibrium blood flow is restored. Although HR can vary with respect to physiological and local effects, it is generally modelled with one common Haemodynamic Response Function (HRF) across brain locations and subjects. The "canonical HRF" (Friston et al., 1998), most commonly used, is depicted in Figure 2.1. When recording BOLD signal evolution during experimental paradigms, local neural activations can thus be deduced by a deconvolution with this HRF.

CAPACITIES AND LIMITATIONS Nowadays, functional Magnetic Resonance Imaging (fMRI) is a ubiquitous tool to investigate brain function in cognitive science research. For human whole-brain imaging, the most common spatial resolution is 2-3 mm, progressing toward submillimeter resolution (Goense, Bohraus, and Logothetis, 2016). A key characteristic of fMRI images is their acquisition cost, which eventually limits data availability. Many challenges faced in this field, including those targeted in this thesis, would be largely mitigated by massive datasets. Unfortunately, such datasets are still rare and will remain so in the foreseeable future.

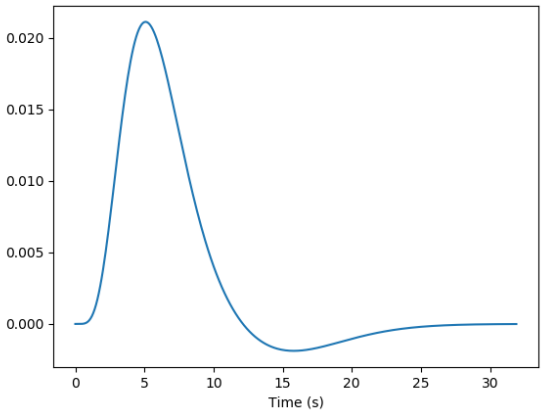

Figure 2.1: Canonical Hemodynamic Response Modelling Function Friston et al., 1998. This linear combination of two Gamma functions was introduced in SPM software and is the default model of HRF.

Another important fMRI limitation is time resolution. Due to the sluggishness of hemodynamic coupling and its timing variability, the time precision that is usually claimed is of the order of one 1 or 2 seconds per image, which limits the span of observable brain processes to rather slow ones (Glover, 2011). Other limitations include the fact that fMRI images have a low signal-to-noise ratio (SNR) (Welvaert and Rosseel, 2013), are subject to artifacts and are not quantitative per se, as the magnetic effects derived from the BOLD signal are hard to relate to a precise quantity of neural activation (Pike, 2012). All these limitations introduce the need for robust processing of images in order to draw statistical conclusions. Since fMRI preprocessing pipelines are often complex and depend on the context (see Figure 2.2), we will especially focus on introducing crucial processing steps.

2.1.3 *Structural preprocessing*

Cognitive studies using fMRI often include an additional structural MRI session to image brain structures and tissue types (gray and white matter). These are first used to restrict studies to brain tissues. Skull-stripping procedures are often used to produce a mask of brain tissues, discarding surrounding regions. Some studies apply an additional tissue segmentation step to distinguish gray matter from white matter and cerebrospinal fluid. It is then possible to refine image masks and restrict studies to gray matter tissues, since BOLD activity is restricted to those regions.

ANATOMICAL REGISTRATION Human brain structures differ in shape in size across individuals. For instance, brain volume or surface can vary almost twofold among typically developing humans of the same age (Giedd et al., 2015). Despite this irreducible anatomical variability,

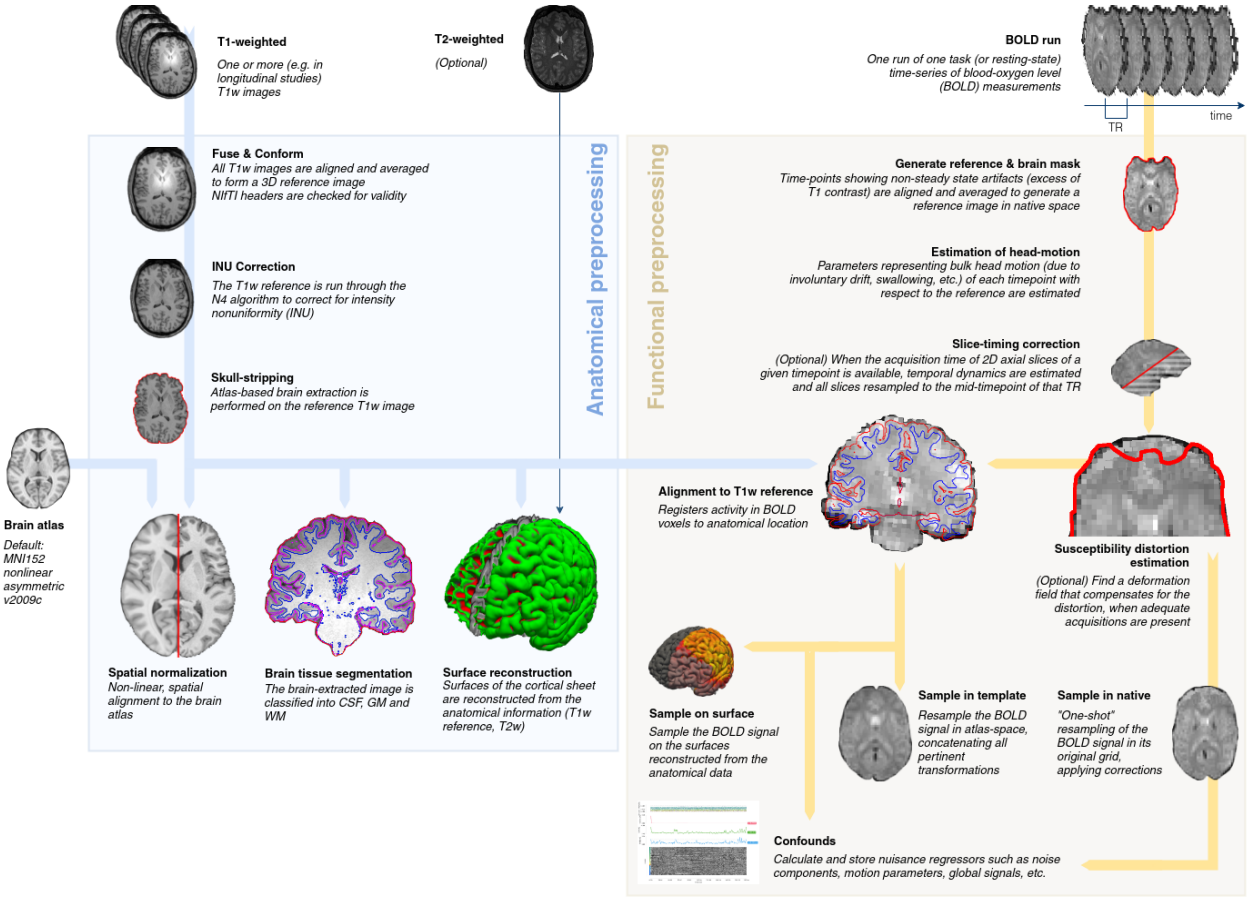

Figure 2.2: Illustrative fMRI preprocessing steps: the fMRIprep normalized pipeline. fMRI studies typically include additional structural MRI acquisition. *Anatomical preprocessing*(2.1.3, left side) is performed on anatomical image(s) and notably include *skull-stripping*, *brain-tissue segmentation* and *spatial normalization to a given template*. *Functional preprocessing*(2.1.4, right side) is performed on BOLD images and can notably include *co-registration*, *mask estimation*, *optional registration to template space reusing normalization learned anatomical images*, *head-motion estimation*, *field inhomogeneities correction*, *signal cleaning*. Figure replicated from *fMRIprep* (Esteban et al., 2019) documentation

cognitive studies are often interested in general features of cognition. Matching brain structures of various individuals to a common space is thus an important preprocessing step, called spatial normalization (Friston et al., 1995). This normalization step consists in registering brain images, maximizing their correspondence to a given template while allowing only smooth deformations. This registration is most often inferred using structural images, and hence often denoted as *anatomical registration*. Estimated registrations can then be applied to other images (e.g., BOLD) acquired for a given subject once those have been rigidly aligned to the structural image used for estimation. Registration techniques usually search for invertible diffeomorphisms between the source and target images, symmetric with respect to these source and target images, and guided by a similarity measure to be chosen among several alternatives. Some datasets used in further experiments, notably Individual Brain Charting dataset, were registered

using SPM software, which relies on a unified step of probabilistic segmentation and diffeomorphic registration using geodesic shooting (Ashburner and Friston, 2005, 2011). Others datasets, used in Chapter 6, were preprocessed using the *fMRIprep* package. Under the hood, this pipeline relies on another popular registration method (used in Chapter 4 experiments): Symmetric Image Normalization (SyN) (Avants et al., 2008). SyN maximizes the cross-correlation within the space of diffeomorphic maps through solving symmetric Euler-Lagrange equations. Other notable registration methods include extensions of Demons algorithm (Thirion, 1998) (e.g., Spherical Demons for surface-based registration from Yeo et al., 2009), and Large Deformation Diffeomorphic Metric Matching (Beg et al., 2005). This registration process makes it possible to project various subject brain structures to a reference image. A common reference space is the Montreal Neurological Institute (MNI) template (Mazziotta et al., 1995), deemed as a good common representation obtained from 152 subjects, and routinely used in group studies. While obviously discarding some subject-specific information, this well-established process enables to map reasonably well information acquired in different subjects and to ease subsequent statistical processing.

2.1.4 *BOLD signal preprocessing*

As a product of a complex acquisition process, the fMRI signal is plagued with multiple sources of noise. These range from experimental and instrumental noise to physiological and movement artifacts. To try to separate those nuisances from the signal induced by neural activity, several denoising techniques are routinely used as preprocessing steps. We don't intend here to be exhaustive about noise sources and available denoising techniques, but instead to give a brief account of commonly used pipelines (Caballero-Gaudes and Reynolds, 2017).

HEAD-MOTION ESTIMATION Motion, and head-motion in particular, create artifactual magnetic effects that confound the signal of interest (Friston et al., 1996). To eliminate this nuisance, preprocessing steps usually include the calculation of a rigid-body transform from BOLD images to a reference frame. These 6 motion parameters estimated for each time step (along three translation directions and three rotation axes) can then be included as confounding regressors in later modelling.

PHASE ENCODING AND FIELD INHOMOGENEITIES MRI images are subject to spatial and intensity distortion due to magnetic field inhomogeneities created by scanner imperfections, as well as spurious fields (including the magnetization of the subject itself). These distortions are most pronounced in the phase-encoding direction (i.e., the direc-

tion chosen to apply one specific gradient called the phase-encoding gradient). A common method to correct for those distortions is thus to acquire protocols twice with opposite phase-encoding polarities, which display opposite distortion patterns, and average those (Holland, Kuperman, and Dale, 2010). Some MRI acquisition sequences also make it possible to either measure directly the magnetic field (and its homogeneities) or deduce them from successive scans and use this model of magnetic field to compensate for induced distortions (Hutton et al., 2002).

SIGNAL CLEANING Ongoing physiological processes, such as cardiac pulsing and respiration, are additional sources of signal fluctuation. Primary effects of these two processes give rise to a relatively low-frequency signal, which can sometimes be cut off using high-pass filtering. When it is not possible or for the sake of removing other artifacts, data-driven denoising methods are used. Principal Component Analysis and Independent Component Analysis (Sui et al., 2009) are two ubiquitous techniques to separate noise artifacts from other signal components (Behzadi et al., 2007).

Final steps of the preprocessing include temporal filtering, confounders regression, detrending and spatial smoothing. Spatial smoothing, using a Gaussian Kernel characterized by its full-width-at-half-maximum (fwhm), is commonly used to remove part of uncorrelated noise and improve SNR. It also helps improve similarity of brain images across subjects, at the cost of spatial resolution (see Section 2.3.2).

2.2 INVESTIGATING BRAIN ACTIVITY WITH FMRI

2.2.1 *Task-based protocols*

In task-based fMRI, a sequence of stimuli is presented to a subject following a carefully monitored design while their brain activity is recorded. This *experimental protocol* entails a controlled priming of various functional processes that can be used to analyze neural activity observed during the experiment. While several types of task-based protocols are used depending on the brain function they study, they all share common principles. Protocols usually rely on sensory stimuli, often visual or auditory, to trigger a behavioral change of interest. A response, not necessary of interest, is asked from the subject at regular intervals in order to ensure their active participation in the protocol. As the stimulus processing can be a factor confounding observed brain activity, trials of interest are intertwined with the presentation of some control stimuli. The effect of interest will thus be the difference of signal acquired in a trial of interest versus a control condition (BOLD

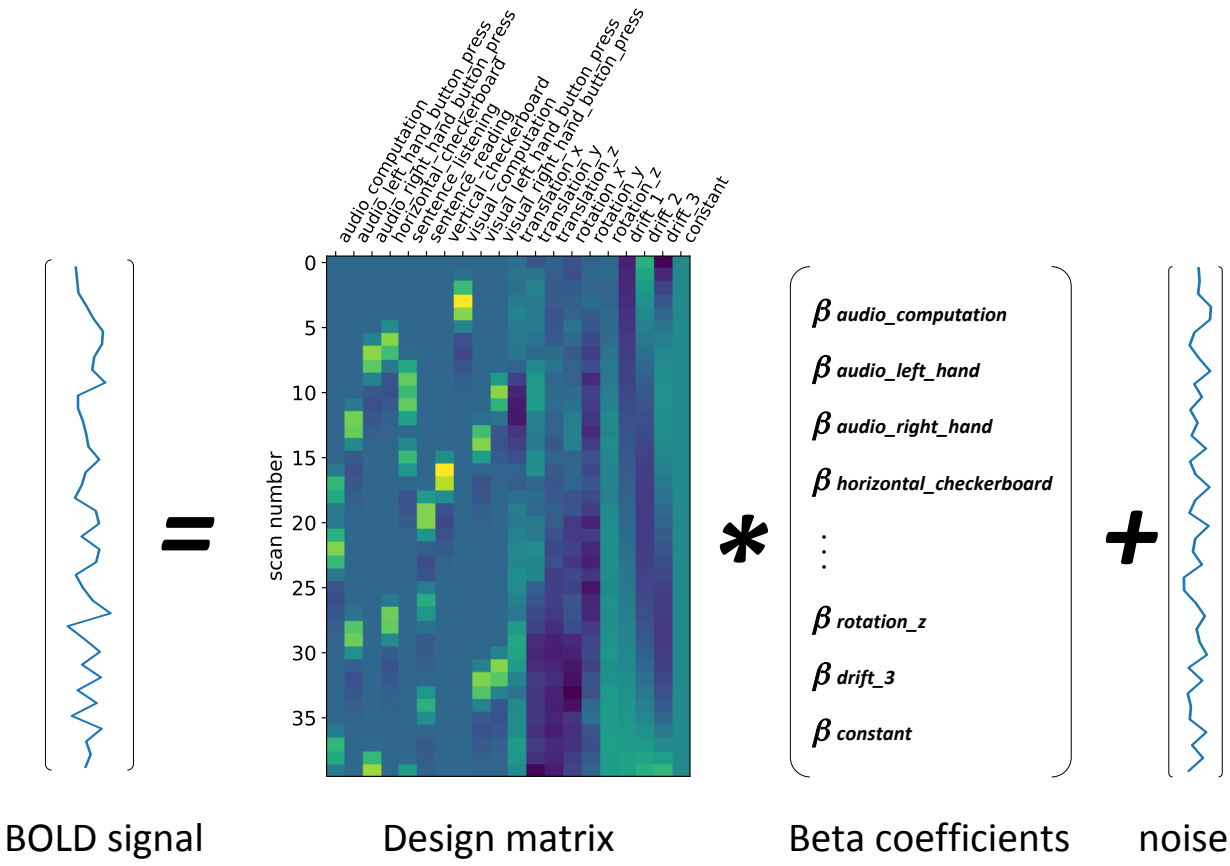

Figure 2.3: General linear modelling of a task-based protocol. A design matrix is derived from regressors of interest, potential confounds (here motion confounds) as well as drift terms. For each voxel separately, a linear regression (Equation 2.1) then estimates β, the coefficients of each regressor. $*$ denotes the matrix-vector product. Figure inspired from Nilearn documentation.

contrast), effectively subtracting a part of the MRI signal unrelated to the cognitive process of interest.

Two main kinds of statistical analysis are performed on task-based fMRI. *Statistical Parametric Mapping (SPM)* (Penny et al., 2011) is a voxel-based approach, employing classical inference to identify functionally specialized brain responses to the presented stimuli. It commonly relies on General Linear models (2.2.2). Conversely, *decoding* (2.2.3) approaches try to *predict stimuli* from brain signals. Models supporting those predictions are multivariate: they search for patterns of activations across several voxels relevant to discriminate between possible stimuli.

2.2.2 *General Linear Models*

As mentioned in Section 2.1.2, images acquired using BOLD mechanism are only a proxy for neural activity through the neurovascular

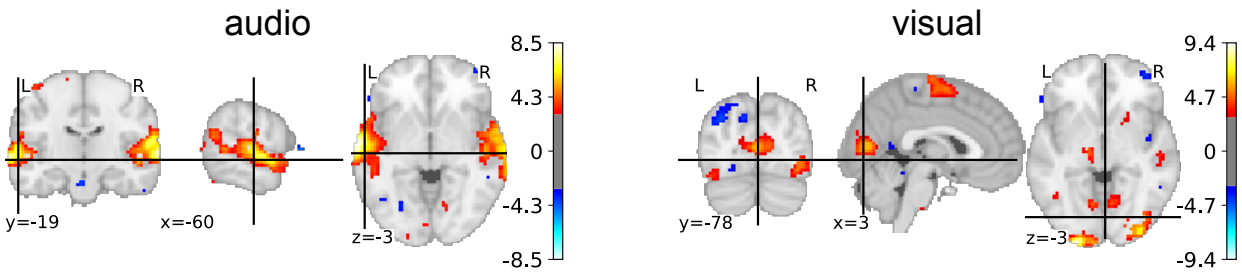

Figure 2.4: Example general modelling output: Base conditions z-maps for a single subject analysis. The design matrix displayed in Figure 2.3 was used to fit a GLM. Thresholded z-maps outline brain regions involved in the sensory process used. Inspired from Nilearn documentation using data from Pinel et al., 2007.

coupling. To deduce brain signals from them, one has to model this coupling using a hemodynamic response function (HRF). As an experimental protocol is a sequence of events, several hemodynamic responses can thus overlap. The routine *linear assumption* states that successive brain responses sum up. To recover the neural signals imputable to a particular event specified by the experimental protocol, one can thus apply a linear convolution between a "block" function representing the event from its onset time for its duration with the chosen HRF.

General Linear Model (GLM) is a generalization of linear regression applied to disentangle effect size imputable to different regressors. They are applied in a *univariate* fashion, i.e., for each voxel independently. The regressors considered can represent single events defined in the protocol, groups of events or a variety of confounding factors (including drifts of various orders accounting for low-frequency noise in the signal). All information available about the p regressors is included in a *design matrix*, as displayed in Figure 2.3. Formally the GLM model solves Equation 2.1 where $y \in \mathbf{p}$ is the signal for one voxel across p scans, $X \in \mathbf{R}^{p,r}$, the design matrix and ϵ, the noise assumed Gaussian, and independent identically distributed (i.i.d.).

$$y = \mathbf{X}\beta + \epsilon \tag{2.1}$$

Ordinary Least Squares (OLS) can then be used to solve Equation 2.1 and find β, the vector of regressors' coefficients as exemplified in Figure 2.3. In practice, the i.i.d noise assumption is not fulfilled since noise display auto-correlation (Monti, 2011). Equation 2.1 is thus routinely solved using an Auto-Regressive model of order 1 (AR(1)).

After fitting these models for all voxels, we can assemble all coefficients estimated for a given regressor in a statistical parametric map, representing the activations linked to this particular regressor. These coefficients maps are often z-scored and the output results are called *Z-maps* as displayed in Figure 2.4. Alternatively, we can make *contrast*

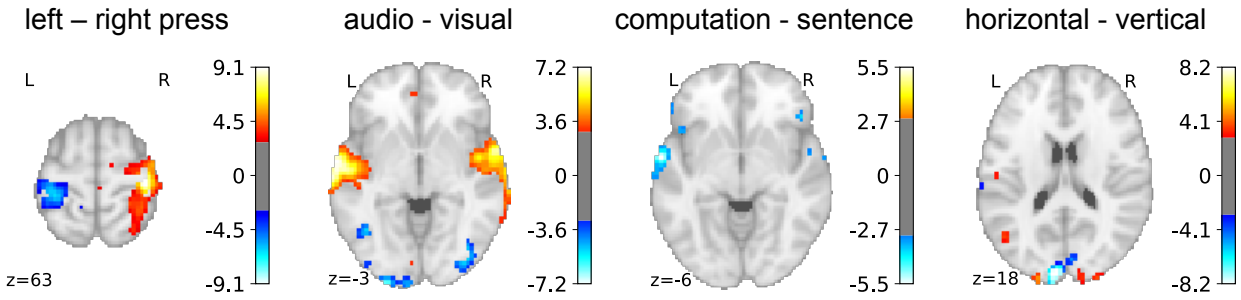

Figure 2.5: Example general modelling output: Contrast z-maps for a single subject analysis. Derived using design matrix displayed in Figure 2.3. Contrasted effects of interest between several conditions outline functional process ascription. "left-right press" contrasts left and right button press events, "audio-visual" contrasts trials using audio versus video instructions. "horizontal-vertical" contrasts horizontal and vertical checkerboards trials. Inspired from Nilearn documentation using data from Pinel et al., 2007.

maps that display activations specific to an experimental condition of interest with respect to a control condition, as displayed in Figure 2.5. Statistical tests are then often conduced on those maps, to derive thresholds that guarantee the statistical significance of observed activations. Several thousands of tests are conducted independently (one per voxel), but the chance that a false positive will arise among all those test is not controlled by their *uncorrected* $p-values$. Thresholds thus often involve compensating for the multiple comparison problem (Alberton et al., 2020). Statistical parametric mapping is a ubiquitous neuroimaging analysis technique. It has been used to produce most of the currently available results linking brain regions to various functional processes.

2.2.3 *Decoding*

Conversely to GLM-based analysis, Multivariate Pattern Analysis (MVPA) or *decoding* (Haxby et al., 2001; Kamitani and Tong, 2005; Haynes and Rees, 2006; Naselaris et al., 2011) is an analysis technique aimed at predicting, from observed brain activity, the stimuli that caused it (or another varying experimental parameter). To make this prediction, it combines information from many voxels (i.e., it is a multivariate method). The *labels* to predict are derived from the experimental protocol (e.g., events, groups of events). To relate those labels to the relevant activation maps, input data are usually preprocessed and deconvolved, possibly through fitting a GLM. A machine learning model is then trained to predict, from these statistical maps, the right label (Haynes, 2015). In a simple classification setting, where we try to predict labels $\mathbf{y} \in \mathbb{R}$ among two possible value encoded as 1 and -1 from n images $\mathbf{x_i} \in \mathbb{R}$, a simple model is the linear Support Vector Machine (SVM). This model searches for a hyperplane $\mathbf{w}$ that

best separate points with label 1 from points with label −1, solving Equation 2.2.

$$\begin{aligned} &\min_{w,b} \quad \|\mathbf{w}\|_F^2 \\ \text{s.t.} \quad &y_i(\mathbf{w}^T\mathbf{x_i} - b) \geqslant 1 \end{aligned} \tag{2.2}$$

The relevance of the model is measured by its *prediction accuracy*. This performance is judged in comparison with *chance level*: the accuracy achieved with random predictions. If model accuracy is significantly higher than chance level, it can be concluded that activation signals contain some information that discriminates across possible labels, i.e., neural processes that support those activations can be meaningfully differentiated based on signal carried by the studied regions. As with any machine learning model, it is important to prevent models from overfitting by using independent data to train the model and test its performance. This practice—most often implemented in cross-validation loops— is especially important when working with neuroimaging data, plagued with a multitude of confounders that can artificially improve model accuracy. It has been shown that cross-validation should be leaving out at least a complete acquisition session, as it would otherwise be prone to optimistic scoring due to session effects (Varoquaux et al., 2017). In neuroscience studies, decoding models performance is often compared across regions to establish which regions hold signal discriminative for given conditions, and is hence most involved in the functional processes probed by those conditions. Additionally, the underlying multi-voxel patterns supporting successful predictions are often reported as candidate neural correlates supporting the functional processes probed. Sparsity inducing models are thus often favored since they provide interpretable "weight-maps" (Gramfort, Thirion, and Varoquaux, 2013; Grosenick et al., 2013; Wu, Koyejo, and Pillow, 2019). On another note, decoding models can both be learned for one subject (intra-subject) or across multiple subjects (inter-subject). We argue later on in this manuscript (6.1) that inter-subject decoding performance can be used as a loose measure of functional variability.

DECODING VS. GENERAL LINEAR MODELS Decoding generates models that are able to predict labels on unseen data. Through cross-validation, they are judged against the variability of the studied samples and can thus be said to be generalizable (Haynes, 2015). Nonetheless, models achieving good decoding performance (e.g Hoyos-Idrobo et al., 2018) come with weak statistical guarantees. It is possible to resort to bootstrapping to ensure that an accuracy difference across models is significantly better than chance. However, assessing the statistical significance of multivariate voxel patterns that support prediction remains an active research area (Chevalier et al., 2021; Schrouff

et al., 2013, 2018b). By contrast, statistical tests as carried out using GLMs enable massively univariate statistical inference but as such, they are prone to multiple comparison bias (Thirion et al., 2007) and do not characterize the global multivariate structure of activations, nor out-of-sample generalizability. Overall, those two techniques provide complementary ways to investigate brain function (Gilron et al., 2017).

2.2.4 *Resting State*

While task-based protocols are often used to study specialized functional processes, resting-state fMRI is a very popular alternative paradigm designed to measure spontaneous brain activity. In this type of protocol, the subject is not instructed to perform any specific task. Low-frequency functional activity is observed nevertheless, which can be related to brain processes identified through task-based paradigms. While the absence of protocol make these data easy to acquire and to compare across studies, it also makes it harder to separate signal from artifacts and only limited inferences can be made on the functional networks involved.

Common analysis techniques of resting state data rely on measuring activity correlations between various brain regions. These regions can be either pre-defined in the case of seed-based connectivity, or defined in a data-driven fashion, e.g., using Independent Component Analysis (Varoquaux et al., 2010). When viewing the brain as a graph, where each brain region is connected, directly or not, with others, these correlations measure how inter-related brain regions are. They can be assembled in a *functional connectome*, a graph encompassing functional connections between various parts of the brain at a given resolution. This family of analysis, called functional connectivity, is often used to segregate brain regions in networks that are tightly related in order to perform some functional processes. Since this "modality" is not selective of specific brain processes, it is often viewed as a "cheap" yet substantial way to characterize individual brain organization.

2.2.5 *Naturalistic*

Naturalistic protocols are those that present rich, multimodal, dynamic stimuli to trigger a perceptual and cognitive experience representative of human ecological state (Sonkusare, Breakspear, and Guo, 2019). They are often opposed to task-based paradigms that involve tightly controlled sequences of sparse, artificial and decontextualized stimuli (Nastase, Goldstein, and Hasson, 2020). They can be employed in order to study complex cognition patterns, obscured by traditional task-based paradigms, such as sustained attentional engagement, context processing, social interaction and dynamical aspects of cognition. More generally, they allow imaging brain activity which better matches

daily state. They have been shown to create intense and highly consistent brain activations (Hasson et al., 2004) across subjects, which is why some studies advocate that they could be a practical way to roughly characterize cognition in a single scanning session. Compared to resting state, these paradigms naturally ensure a greater subject involvement, less head-motion artifacts and overall a more controlled brain activity. Most common naturalistic protocols include movie watching and audio clip listening, but a growing number of studies use varied stimuli such as gambling or video games (Bellec and Boyle, 2019).

In terms of analysis, these paradigms retain a "weak supervision" from task-based protocols which makes it possible, after labelling, to use general linear modelling (e.g., Bartels, Zeki, and Logothetis, 2008 or Huth et al., 2012) as well as decoding of events of interest (Mandelkow, Zwart, and Duyn, 2016). As an alternative to labelling, it is also possible to rely on data-driven event-modelling (Baldassano et al., 2017). It is of course also possible to treat these protocols as unlabeled, and to apply resting state analysis techniques. Finally, inter-subject correlation (Nastase et al., 2019) of brain signals is increasingly used on those datasets to characterize shared responses across subjects, as well as variability in cortical areas.

2.3 FUNCTIONAL VARIABILITY

Most studies using the techniques introduced in Section 2.2 are pursuing two main goals. The first one is concerned with fundamental knowledge: understand brain function organization and its link to cortical anatomy, also called functional brain mapping. The second one is application-driven: find reliable bio-markers that help with diagnosis or monitoring of psychiatric or neurological conditions and brain diseases. These two kinds of links between brain and mind have historically relied on searching for universal brain processes across individuals (Section 2.3.1).

However, even data from large cohorts display a high degree of inter-subject functional variability, undermining attempts to uncover reliable structure-function coupling and derived bio-markers (Section 2.3.2). Originally modelled as noise, this variability actually holds significant signal to relate brain models across individuals. To further study idiosyncratic functional organization and how it can be related across subjects, *deep phenotyping* studies focus on comprehensively mapping brain function for a low number of subjects, and shed a new light on variability (Section 2.3.3).

2.3.1 *Population imaging*

Numerous studies have used the techniques presented previously to link brain regions to various functional processes. They both investigate the neural correlates of specific functional processes (e.g., high-level language functions Fedorenko, Behr, and Kanwisher, 2011, attentional effects on working memory selectivity, Majerus et al., 2018) and models accounting for the underlying computations of complex abilities, such as confidence weighting in an inference process (Meyniel and Dehaene, 2017). These studies often operate under the hypothesis of a tight structure-function coupling, which means that brain functional processes can be ascribed to a specific set of regions responsible for various subprocesses. This hypothesis implicitly guarantees the generalizability of results: if the right regions are found responsible for some functional process in a group of subjects, these regions should display the same functional specificity in the wider population. This type of study has significantly advanced our understanding of core brain function organization principles, and has helped identify population-average effects especially related to large-scale functional networks. Despite this wealth of results, a key stake for the neuroimaging community is to ensure the coherence and generalizability of those results and their extension to even more precise functional characterization.

Small sample size, low signal-to-noise ratio (SNR) and methodological variability are often cited as key challenges (Thirion et al., 2007; Button et al., 2013; Pajula and Tohka, 2016; Turner et al., 2018) that undermine the advance of brain function understanding and its applications. Low SNR also implies that small quantities of per-subject images are not sufficient to yield precise characterization of the studied brain functions. Most traditional studies results have thus limited generalizability on their own and call for meta-analysis in order to build more robust knowledge. To overcome this challenge, many researchers chose to include more subjects in their studies. Among those, several large scale studies have been undertaken to acquire images from a very high number of subjects with fixed protocols, often focused on resting-state MRI (Madan, 2021).

Among those, some studies are primarily designed to analyze the brain correlates of individual characteristics (e.g., age, sex, genetics, sociodemographic factors), or neurological patient condition. As an extension of the UK Biobank medical prospective study, more than forty thousand subjects resting data have already been released, together with genetic data, cognitive measures, psychiatric assessment and a breadth of other health data. The Cam-CAN project uses epidemiological, cognitive, and neuroimaging data to investigate effects of aging on the brain and its impact on cognition. Its imaging data is mostly composed of rest and movie-watching data for more than six hundred subjects.

Alternatively, some studies are especially concerned with improving our understanding of brain function: its general organization and development, its relation to brain network structure or differences across individuals. The most famous one, the Human Connectome Project (HCP) is a large-scale effort to comprehensively chart neuronal connections. Interested in studying both general patterns and variability in the connectome (see 2.2.4) of this large cohort, it contains resting state and task fMRI data for nearly one thousand subjects.

The Adolescent Brain Cognitive Development (ABCD) dataset (Casey et al., 2018) is another large-scale initiative aimed at longitudinally imaging ten thousand adolescents every two years over the course of 10 years. It relies mostly on task-based fMRI to investigate brain development, with a special focus on functional processes previously linked with addiction.

Beyond those exceptional initiatives, data-sharing platform have also made it possible to assemble large collections of raw (Poldrack et al., 2013a) and processed (Gorgolewski et al., 2015) data covering many functional processes and many subjects. The accumulation of available datasets and derived studies significantly advanced our general understanding of brain function; it notably enabled large-scale atlasing (Glasser et al., 2016; Rubin et al., 2016; Schaefer et al., 2018; Varoquaux et al., 2018; Dadi et al., 2020) and general decoding attempts (Koyejo and Poldrack, 2013; Mensch et al., 2017), moving forward with the comprehension of general brain organization.

TRANSLATIONAL RESEARCH For several decades, studies have accumulated evidence on the link between specific brain regions activities and pathologies (e.g., anterior cingulate cortex abnormal activations and depression, Mayberg et al., 1999), or more generally health related outcomes (such as risk aversion and the striatum and amygdala Tom et al., 2007). But the fuzziness of the structure-function coupling makes those measures unreliable for clinical use on individuals. To go beyond those limitations, an increasingly popular strategy is to resort to predictive modelling techniques, such as decoding, introduced previously. These techniques identify activation patterns across several voxels that help predicting accurately a given variable of interest such as a psychiatric condition (e.g., Alzheimer's Disease, Davatzikos et al., 2009), or a brain disorders (e.g., consciousness disorders, Demertzi et al., 2015). In a population context, a key challenge is to find reliable biomarkers that can help in the prognosis, diagnosis and therapeutic monitoring of patients. These biomarkers must additionally be easy to acquire, share and replicate in a variety of contexts and clinical settings (Woo et al., 2017).

2.3.2 *Inter-subject functional variability*

OBSERVED VARIABILITY AND ITS EFFECTS While population imaging approaches help pinpoint central tendencies across large cohorts, they have also outlined a high level of deviation around this average. Some regions' variability makes it hard to draw conclusions at the population level. As an example, the HCP study investigators highlighted a high degree of functional variability across individuals despite applying a tailored areal surface-based registration scheme guided by several modalities, including fMRI (Glasser et al., 2016). They especially described a region termed "area 55b" that they identify as part of the language network in the inferior portion of both the left and right dorsal premotor cortex. This functionally defined area displays a striking degree of variability in location and shape, in fact a substantial minority of subjects show a topological organization that cannot be matched with the most common organization. This high variability is still observed across large cohorts and thus cannot be attributed to a small sample effect.

Beyond spatial variations, another strong source of variability is the variation of activation magnitude recorded for a task across subjects, and sometimes across sessions (Smith et al., 2005). As already mentioned, fMRI activation is only a partial measure of underlying neural activity and doesn't allow quantitative comparison. In its worse case, this activation level variability can lead to some subjects not showing any reliable signal at all for some contrasts. More generally, it undermines activations comparison across subjects.

In traditional studies, a common way to handle variability, which is mostly local, is to apply some Gaussian spatial smoothing to subjects' data. While this technique effectively improves image SNR, it also blurs activation patterns and prevents precise mapping. Even without smoothing, the geometrical effect of averaging many activations patterns in the same neighborhood but not overlapping will create smooth activations "blobs". Overall, brain mapping precision is strongly capped by the spatial scale of functional variability, which is roughly at the cm scale but widely different across cortex locations.

Beyond blurring activation patterns geometry, variability also makes it harder to reconcile functional features across subjects and to build a representative group model of functional organization (Stelzer et al., 2014; Wang et al., 2021). While group average can be coarsely similar to individual data, it has been shown that group models don't account for several topological features that are well-characterized at the individual level (Heun et al., 2000; Seghier et al., 2004; Laumann et al., 2015). Brain mapping of high-level cognitive functions, a very active area of research, commonly displays such problems, and a large portion of the prefrontal lobe is still ill-characterized, largely due to its strong level of functional variability.

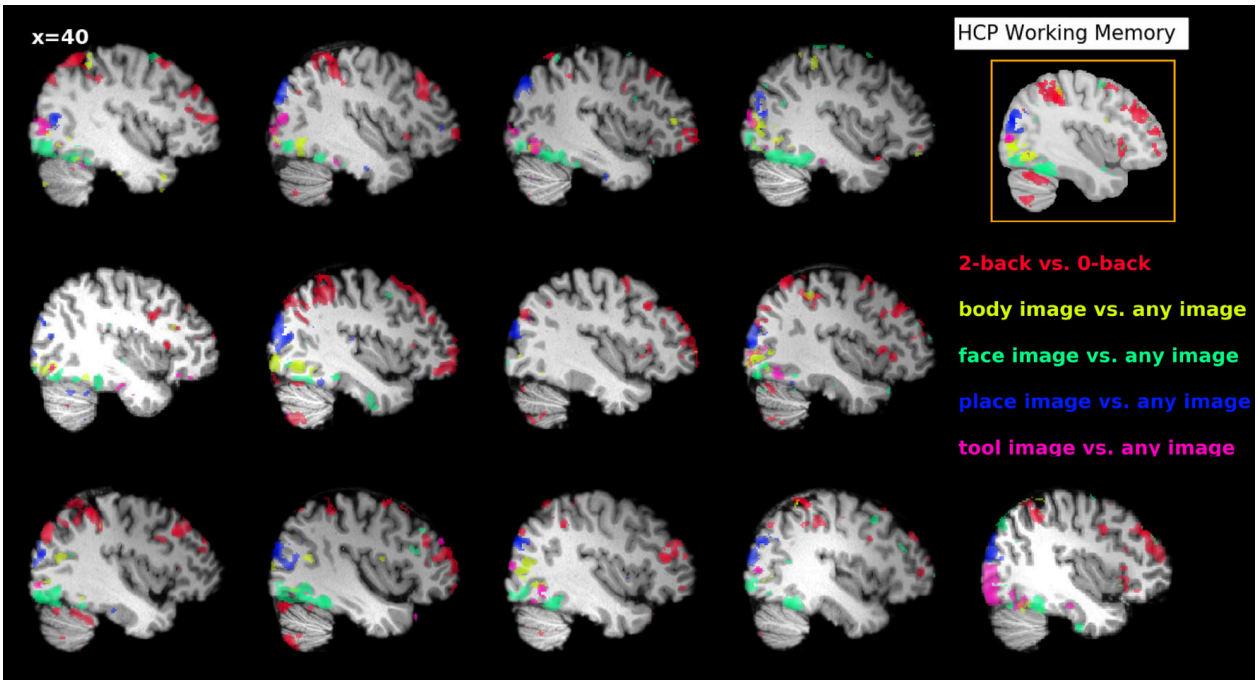

Figure 2.6: Individual contrast maps for 13 IBC subjects and group conjunction map. For each individual, the map is an overlaid view of 5 binarized contrast-maps derived from the HCP Working Memory task. Each contrast displays fixed effects of the corresponding label (see right) using an FDR-corrected threshold q = 0.05. The group-level conjunction map of these individual maps is shown inside the orange frame. All maps correspond to the slice x = 40 mm in the sagittal view. Courtesy of Pinho et al., 2021.

Figure 2.6 illustrates variability commonly observed in fMRI studies, by comparing binarized contrast maps across 13 subjects from the Individual Brain Charting dataset (IBC). All contrasts, derived from a visual n-back task (*HCP Working Memory Protocol*, Barch et al., 2013), display a strong level of variability as well as qualitative difference between individual and group activations. For example, all visual contrasts (i.e., all but *2-back vs 0-back*) probe distinct well-known visual processes (Haxby et al., 2001). However, resulting activations are inconsistent across subjects both in terms of intensity, location, and geometry despite the fact the visual cortex, is one of the *most stable* regions of the brain. We can also observe that the group-level conjunction maps derived from those activation displays a qualitatively different topography, compared with individual images.

Variability can also obscure some fundamental functional organization principles, such as multifunctionality of brain regions. There are ongoing debates among neuroscientists to determine whether several brain regions selectively engage in specific high-level function or if they are fundamentally multifunctional. In Fedorenko, Behr, and Kanwisher, 2011 a detailed account was given of several regions showing language selectivity but often cited as engaging in other processes as well. For most of these regions, the authors conclude that they are, indeed, very selective for language tasks but close to regions involved

in different processes. Even though they are distinct at an individual level, these regions can however be overlapping after group averaging.

With respect to translational biomarkers, inter-subject variability limits the spatial scale at which we can search for patterns and ultimately caps the performance of predictive models on subjects whose activations differ from the average.

In some cases, variability can be extreme, for example due to a pathology with important consequences on structural and functional organization (Blank, Kiran, and Fedorenko, 2017). Our work did not address this aspect, which leads to very specific methodological cases, although the methods that we present could help in those cases to derive matching from functional measurements.

WHAT IS FUNCTIONAL VARIABILITY? From the point of view of traditional fMRI studies, it is hard to find a criterion to distinguish "variability of interest" from sheer noise. Intuitively, it should be defined as functional patterns or signal structure that is stable for an individual (as opposed to intra-subject variability) without generalizing across subjects. However, making such a distinction is possible only with a precise and reliable individual characterization of several subjects. While *deep phenotyping* (2.3.3) fMRI studies have recently emerged, such a characterization remains out of reach. Our goal here is not to give an exhaustive account of these endeavors, that are still exploratory, but rather to give a sense of the complex and mixed nature of the observed "functional variability".

We can first conceptualize it as a remainder of anatomical variability that couldn't be perfectly handled through spatial normalization (Thirion et al., 2006a). As discussed in Section 2.1.3, a typical workflow for group studies includes an anatomical registration step to compensate for brain structures differences across subjects. The transforms estimated on structural images are then used on rigidly coregistered BOLD images to bring them to a common template space. While this standard procedure is a best effort attempt at handling brain structure difference in size and shapes, perfect correspondence cannot be achieved. Functional variability can thus partly be caused by anatomical variability, and be influenced as well by the normalization process (Eickhoff et al., 2009). The main structural and physiological parameters found to be correlated with this variability are: gray matter density, cortical thickness, morphological anatomy, cortical layers, white matter tracts and myelination (Seghier and Price, 2018).

Letting alone anatomical variability, the brain is a plastic object which is thus subject to constant reorganization, and shaped by prior genetic and developmental constraints (Paus, 2013), as well as daily environment. Functional variability in the motor cortex between athletes and general population is indeed not surprising. Even though

principles underlying this kind of reorganization are still an active research area (Sampaio-Baptista et al., 2013), it is self-evident that brain function also depends on cognitive abilities that are not homogeneous in a healthy population. Although many studies attempted to find important physiological (Saygin et al., 2016), and genetic (Thompson et al., 2017) correlates of variability, one would need an extensive individual phenotyping, composed of many interacting factors, to hope model variability.

More formally, inter-subject variability can be described as intrinsic, strategic, or contextual (Seghier and Price, 2018). Intrinsic variability stems from inherent factors fixed on a long time-scale (e.g., language lateralization which is linked to genetic factors). Strategic variability refers to the use of different cognitive strategies to solve the same problem (Miller et al., 2012) and is related to learning. Contextual variability is more spurious and includes subject experimental performance, driven by many factors, including their familiarity with the scanning environment or their mood states.

On top of subject specific factors, it is hard to disentangle idiosyncratic patterns of signal from contextually induced variations. Experimenters, tooling, overall experimental context and data processing account for an additional source of confounding "noise", and it is complex to distinguish a "site-effect" (Brown et al., 2011) from other sources of inter-subject variation. To be fair, a perfect individual characterization is quite unrealistic, since fMRI studies show an important amount of inter-trial (Coste et al., 2011) and inter-session (Bennett and Miller, 2010) variability for any given subject. The low-level of overall fMRI reliability (Elliott et al., 2020) is another major obstacle limiting the translational use of fMRI.

HOW TO HANDLE VARIABILITY? As presented previously, variability designates a complex phenomenon, is hardly distinguishable from noise, encompasses many distinct effects and seems overall impossible to model. The most promising view so far is to focus instead on idiosyncratic—specifically individual—functional organization (Dubois and Adolphs, 2016; Mahowald and Fedorenko, 2016). Finding reliable individual level functional markers could indeed yield insights on how those relate across individuals. Our work builds on this view to learn how to compensate variability, in a model-free fashion. To be realistic, this approach calls for important amounts of per-subject data, that are the only way to differentiate between stable individual functional patterns from noise.

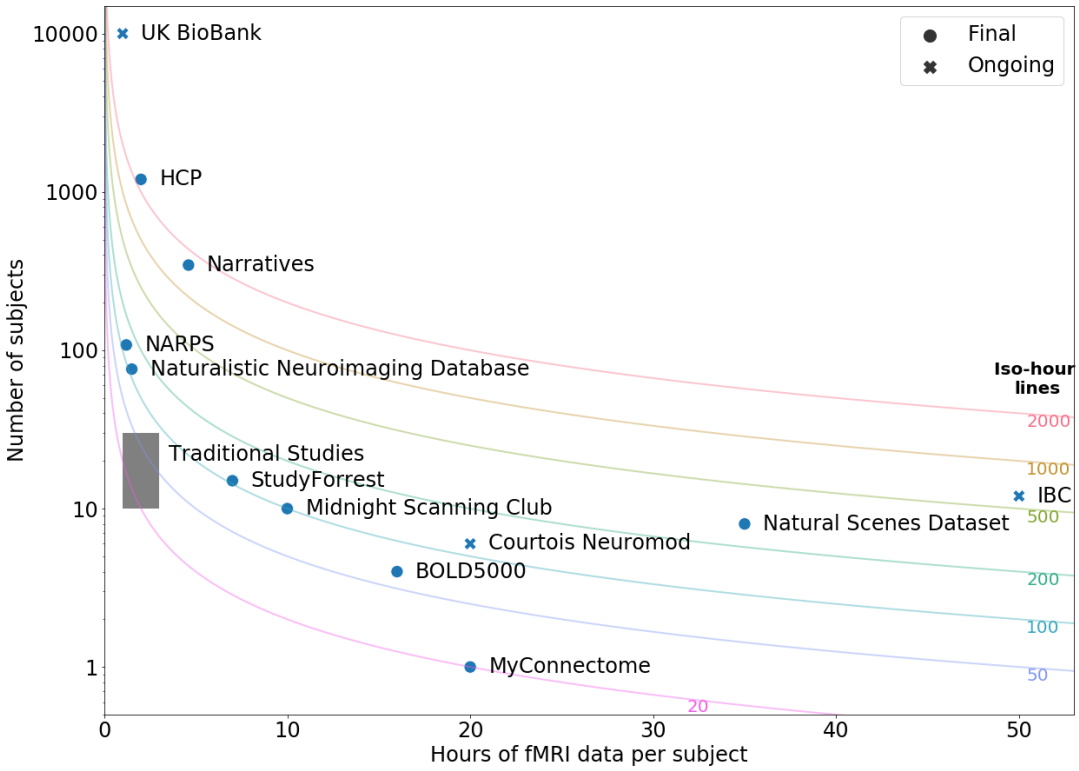

Figure 2.7: Trade-offs between number of participants and amount of data per participant in selected large scale fMRI datasets. Inspired from Naselaris, Allen, and Kay, 2021.

2.3.3 *Deep Phenotyping*

In the last decade, awareness of variability has progressively grown in the neuroimaging community, making it a core characteristic of brain and its derived signals (not only inter-subject variability, but also intra-subject, Salehi et al., 2020). With the influence of precision medicine concepts, *deep phenotyping* (Robinson, 2012) imaging studies have expanded with a common focus on acquiring an important amount of data on a few subjects to extensively describe their idiosyncratic brain function organization (Naselaris, Allen, and Kay, 2021). In sharp contrast with the common wisdom that sampling more individuals is desirable, these studies have stemmed from the conviction that because of the strong inter-subject variability, sampling many individuals introduces an unmanageable amount of variance in observations that would rather obscure precise brain function understanding. They thus chose to extensively sample experimental conditions instead

to precisely characterize brain response to numerous and complex stimuli (see Figure 2.7).

Poldrack et al. investigated long-term brain dynamics of one subject, and their link to physiological processes. To do so, they focused on resting-state fMRI acquisitions repeated for 100 scanning sessions over the course of 18 months (Poldrack et al., 2015). The StudyForrest initiative extensively scanned 10 subjects in a multi-modal fashion. It notably includes both naturalistic and task protocols, primarily focusing on auditory stimuli processing (Hanke et al., 2016, 2014, 2015). The BOLD5000 study focused on characterizing the visual system of four subjects. Each of them was scanned for 20 hours, and overall was presented 5000 different images from classical computer vision datasets: SUN, COCO, and ImageNet (Chang et al., 2019). Courtois Neuromod is an ongoing neural modelling project aimed at training artificial neural networks using extensive individual brain activity data. Six subjects are scanned, with a target of 500 hours of functional data recording per subject. These scans mostly cover movie-watching protocols, as well as a few functional localizers and HCP task-based protocols.

Other datasets have been acquired to give precise individual account of subjects function across a wider range of cognitive tasks. The Midnight scanning club (MSC) dataset includes twelve hours of recording for ten subjects (Gordon et al., 2017). Half the scans images resting state activity while the other half covered three protocols probing motor activity, incidental memory as well as visual and verbal discrimination. The ARCHI database (Pinel et al., 2019) developed four functional localizers acquired for 78 individuals. These localizers total 29 experimental conditions associated with language, arithmetic, social reasoning and visuomotor representations. Finally, the Individual Brain Charting (IBC) dataset (Pinho et al., 2020, 2018) is an ongoing acquisition initiative of 12 participants performing numerous task protocols (already 19 protocols including those of ARCHI, HCP) as well as rest and movie watching data.

These initiatives, accumulating a large set of functional contrasts, look promising to establish a finer-grained functional atlas, and individual signatures for various functional processes. Local individual activation seems quite consistent across contrasts, which might lead to developing reliable brain measures. An interesting illustration of individual variation captured by these characterizations is provided by Pinho et al., 2021, who focus on comparing functional topographies across subjects. To do so, they aggregate individual data over 12 different tasks covering several psychological domains and make a common decomposition of 13 individuals data, using sparse dictionary learning (Abraham et al., 2013). This yields a set of 20 common functional components (labelled according to the contrast mainly contributing to each component) and individual spatial maps for each

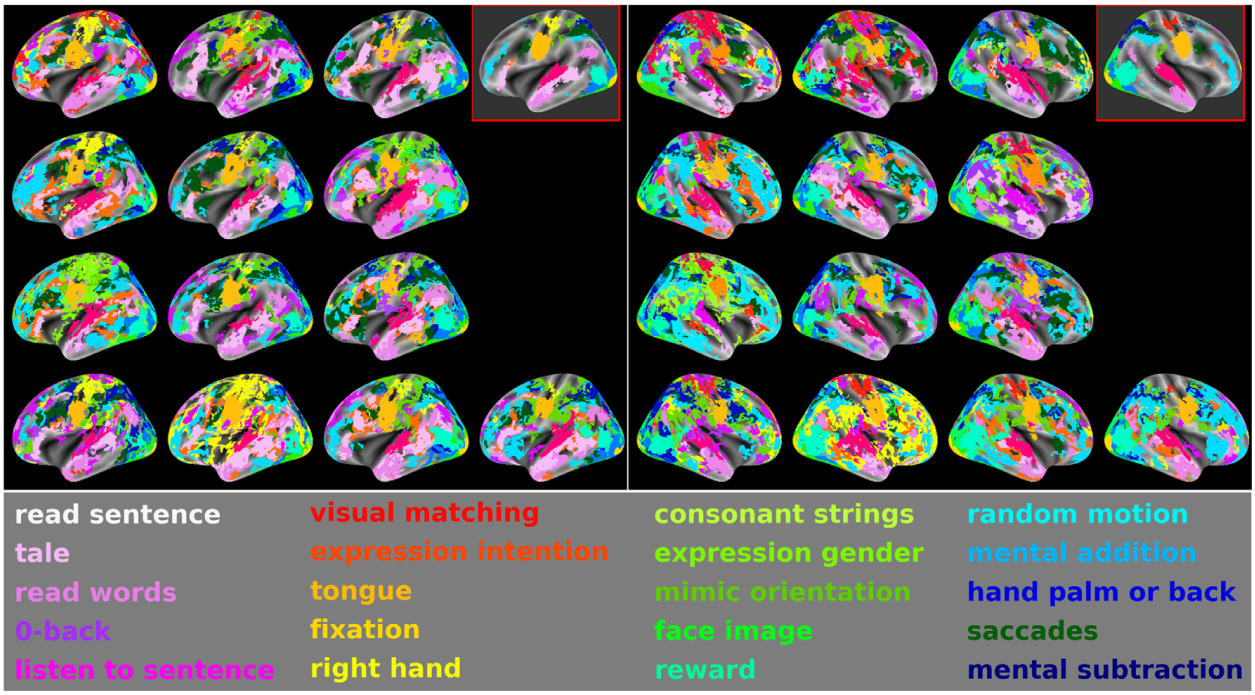

Figure 2.8: Dictionary of 20 cognitive components summarizing 13 individual topographies in fsaverage space. (Top left/right) Labeling of left/right hemispheric cortical regions, according to the strongest dictionary loading in that region. The top-right brain maps outlined in red of each image display a median map obtained at group level, that is, a label is assigned to a component if, at least, half of the participants have that label at that location. (Bottom) The 20 cognitive components are labeled according to the contrast z-map that gets the maximum loading for that component. Overall, this figure illustrates the consistency of some components, while it outlines local and large-scale differences across the IBC participants. Courtesy of Pinho et al., 2021.

component, displayed in Figure 2.8. While coarsely stable, this decomposition display a high level of local organization variability. Beyond some especially stable regions, this variability hampers the ability to derive representative group spatial maps. Compared with individual topographies, those group topographies (in top-right red square) show a qualitatively different geometry and have a hard time representing well some components and characterizing some brain regions. More generally, this points to important differences between individual organization and group estimates reported in traditional studies. On a more positive note, the thoroughness of these individual characterizations also make it possible to derive functional correspondences across subjects to handle variability and progress towards precision brain atlasing.

3 FUNCTIONAL ALIGNMENT

Functional variability, described in Section 2.3.2, is central to the study of brain activity across individuals. Still, when imaging a cohort of healthy patients, this variability is rather considered as an obstacle, hampering the understanding of functional brain organization principles. Section 3.1 introduces ***functional alignment****, a family of methods proposed to compensate for this imperfect voxel-correspondence across subjects by finding matchings that* ***maximize functional similarity*** *across subjects. In this section, we discuss in detail* ***Hyperalignment*** *(Haxby et al., 2011), together with current challenges and extensions. We also describe the closely related* ***Shared Response Model****. Section 3.2 takes a step back so as to frame functional alignment in the broader machine learning literature and discuss commonalities with work performing some kind of* ***feature alignment****. In Section 3.3, we turn a critical eye on this sub-field, given that the proposed methods are still* ***seldom applied*** *in real use-cases to handle variability. This gives context and motivations for our contributions presented in the following chapters.*

3.1 INTRODUCTION TO FUNCTIONAL ALIGNMENT

3.1.1 *Improving functional correspondence across subjects*

A core challenge for cognitive neuroscience is to find similarity across neural diversity (Churchland, 1998); that is, to find shared or similar neural processes supporting the diversity of individual cognitive experience. Anatomical variability and limited structure-function correspondence across cortex (Paquola et al., 2019; Rodriguez-Vazquez et al., 2019) make this goal challenging (Rademacher et al., 1993; Thirion et al., 2006b). Even after state-of-the-art anatomical normalization to a standard space, we still observe differences in individual-level functional activation patterns that hinder cross-subject comparisons (Langs et al., 2010; Sabuncu et al., 2010). With standard processing pipelines, it is therefore difficult to disentangle whether individuals are engaging in idiosyncratic cognitive experience *or* if they are engaging in shared functional states that are differently encoded in the supporting cortical anatomy. Variability additionally hampers attempts at precise functional brain mapping and discovery of reliable translational bio-markers (see Section 2.3).

A note on terminology

Although this class of methods is broadly referred to as both *functional alignment methods* and *hyperalignment methods*, we adopt the term *functional alignment methods* to better distinguish from the specific Procrustes-based hyperalignment implementation in use in the literature.

Functional alignment is an increasingly popular family of methods for fMRI analysis which aims at handling inter-subject variability. Alignment methods derive matchings across subjects' functional data that can be applied in order to improve similarity of their voxels functional activations. Alignment can either be learned after anatomical registration or replace it entirely. This conceptual shift from anatomy-based to functionally-driven mapping has opened new avenues for exploring neural similarity and diversity. In particular, by aligning activation patterns in a high-dimensional functional space (i.e., where each dimension corresponds to a voxel), we can discover shared representations that show similar trajectories in functional space but rely on unique combinations of voxels across subjects.

As a general framework, illustrated in Figure 3.1, functional alignment can be seen as the search for a transform $\mathbf{R}$ that maps a source functional dataset $\mathbf{A^s} \in \mathbb{R}^{p \times n}$ to a target dataset $\mathbf{A^t} \in \mathbb{R}^{p \times n}$ under some given constraints ($\mathbf{R} \in \mathcal{R}$) in order to minimize their dissimilarity (Equation 3.1). When $\mathbf{R}$ is linear, it can be interpreted as a mixing of *source* voxels signals in order to reconstruct *target* signals.

$$\underset{\mathbf{R} \in \mathcal{R}}{\operatorname{argmin}} \|\mathbf{R}(\mathbf{A^s}) - \mathbf{A^t}\|_F^2 \tag{3.1}$$

To learn $\mathbf{R}$, most methods rely on the fact that both $\mathbf{A^s}$ and $\mathbf{A^t}$ were acquired using the same experimental protocol—either task-based or naturalistic—ensuring a *time synchronization* between their n frames. From the initial introduction of hyperalignment in (Haxby et al., 2011, detailed in section 3.1.2), the range of associated methods has grown to include Shared Response Modelling (SRM, Chen et al., 2015 described in section 3.1.5), Optimal Transport alignment (our contribution, introduced in Chapter 4) and many more, reviewed in section 3.2.

SCOPE DEFINITION For simplicity, we focus here on the context of healthy human imaging, and will mostly use deep phenotyping datasets (2.3.3) as their rich individual characterization will help us to estimate mappings and test them. We additionally make the hypothesis that *anatomical* alignment (2.1.3) has been performed across subjects. This means subjects data were brought to a common reference geometry based on anatomical features and their images comprise the same number of voxels p. Although we will not consider it in this thesis, some methods cited, including Optimal Transport alignment, could be applied without any prior anatomical alignment, which makes

them suitable for other use-cases such as functional alignment across different species (Mantini et al., 2012; Xu et al., 2020).

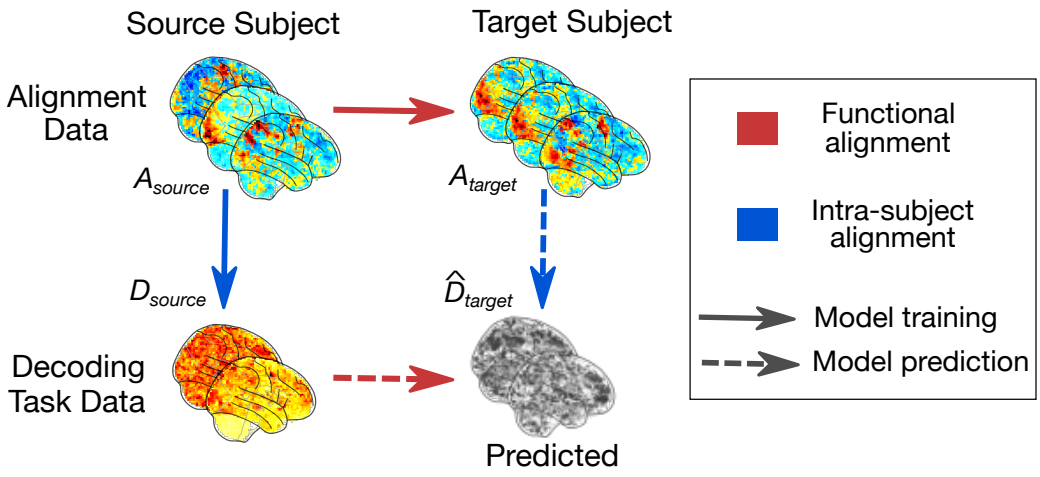

Figure 3.1: Principle of functional alignment The basic goal of *pairwise* functional alignment is to learn correspondence between data drawn from two subjects: from a **source** subject to a **target** subject, using their synchronized **alignment** data $\mathbf{A}$. Each subject often comes with additional data to be aligned, denoted $\mathbf{D}$. Red arrows describe functional alignment methods where correspondence is learned from $\mathbf{A}^{source}$ to $\mathbf{A}^{target}$, while blue arrows describes intra-subject alignment (3.2.3) where we learn correlation structure from $\mathbf{A}^{source}$ to $\mathbf{D}^{source}$. Solid arrows indicate a transformation learned during training. Dashed arrows indicate when the previously learned transformation is applied in prediction to estimate $\hat{\mathbf{D}}^{target}$.

3.1.2 *Hyperalignment*

While seminal work introduced functional alignment as a diffeomorphic registration refinement, guided using functional similarity (Sabuncu et al., 2010, see section 3.2.2), Haxby et al., 2011 proposed to search for matchings exclusively aimed at compensating for functional variability without any anatomical constraint. It is worth noting this is a radical change of perspective. Instead of refining the structure correspondence across subjects—in the hope the structure-function coupling will ensure a functional correspondence—this means turning away from local structural information, to focus only on its functional characterization. Following the intuitions behind multivariate pattern analysis (Haxby et al., 2001), functional signals acquired during a protocol can be seen as a cloud of points in high-dimension, where each dimension is a functional time-point (with a specific stimulation content) and each point is a voxel thus characterized by its activations at each time-point (see Figure 3.2). In this view, alignment can be described as the task of matching those entire point clouds, where each voxel can be mapped to several others that have most similar activation profiles.

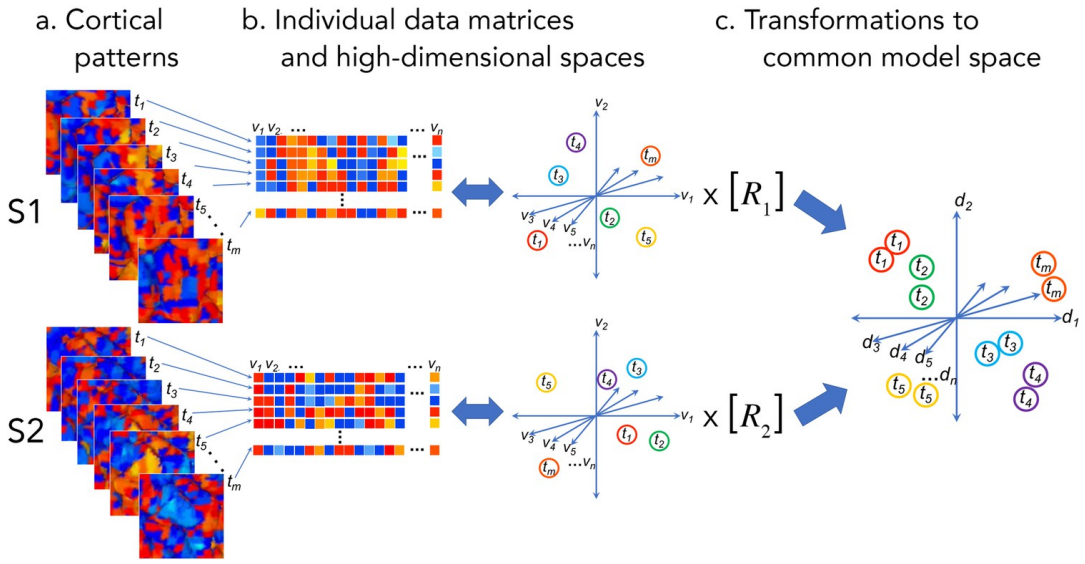

Figure 3.2: Schematic of alignment as a matching task in a high-dimensional space. A voxel is characterized with a series of images, each with a specific functional content. Activation patterns are represented as matrices of activation vectors. The goal of alignment is to improve correspondence between those patterns—the similarity of matrices—through a transform mapping together voxels with similar characterizations. Source: Haxby et al., 2020

To solve this matching problem, Haxby et al. proposed to use Procrustes, an established shape matching method (Goodall, 1991; Gower, Dijksterhuis, et al., 2004). This method yields an orthogonal transform $\mathbf{R} \in \mathbb{R}^{p \times p}$ solving:

$$\underset{\mathbf{R}=\sigma\mathbf{Q},\mathbf{Q}^\intercal\mathbf{Q}=\mathbf{Id}}{\operatorname{argmin}} \|\mathbf{R}\mathbf{A}^{\mathbf{s}} - \mathbf{A}^{\mathbf{t}}\|_F^2, \quad \sigma \in \mathbb{R}^+, \quad \mathbf{Q} \in \mathbb{R}^{p \times p} \tag{3.2}$$

Equation 3.2 is solvable in closed form using the Singular Value Decomposition (SVD) of $\mathbf{A}^\intercal\mathbf{A}^{\mathbf{s}\intercal}$: $(\mathbf{U}, ^\circ, \mathbf{V}) = \text{SVD}(\mathbf{A}^{\mathbf{t}}\mathbf{A}^{\mathbf{s}\intercal})$:

$$\mathbf{R}^* = \sigma\mathbf{Q} \text{ where } (\sigma, \mathbf{Q}) = \left(\frac{\text{tr}(^\circ)}{\|\mathbf{A}^{\mathbf{s}}\|_F^2}, \mathbf{UV}\right)$$

As the best orthogonal matching, this transform can be seen as a rotation matrix mixing signals of voxels in $\mathbf{A}^{\mathbf{s}}$ to reconstruct the signal of voxels in $\mathbf{A}^{\mathbf{t}}$ as well as possible. In their original work, Haxby et al. argued that the orthogonality of the transform aimed at preserving the "representational geometry", meaning the geometry of the point-cloud in this high-dimensional space.

Hyperalignment is a conceptually appealing way to circumvent the ubiquitous variability problem. However, its introduction brought up two main constraints to handle in order to make it applicable to interesting research set-ups: how to handle high-dimensionality of images and the potentially large number of subjects to align.

3.1.3 *High-dimensionality alignment*

Compared to usual registrations, ignoring diffeomorphic constraints gives alignment complete spatial freedom to find matchings that best improve functional similarity. However, it also creates a tension on the spatial scale at which such matchings should be sought for. As seen in Section 2.3.2, inter-subject variability is most striking at a local scale, which is why Hyperalignment was introduced in a well-defined sub-region (Haxby et al., 2011). Moreover, searching for alignment in local regions acts as a form of regularization: considering local inter-subject variability rather than global changes such as large-scale functional reorganization.

Applying the above model directly to full brain would suffer from several issues: *i)* it may create some non-local correspondences (e.g., cross-hemisphere swap) that are not neuroanatomically plausible; *ii)* it is computationally heavy if not intractable: most matching algorithms applied to two images of p voxels, will at least have a complexity of $O(p^2)$; *iii)* it would likely lead to overfitting: to estimate well a $p \times p$ matrix, one needs a high number of samples n, up to p^2, whereas in the typical study $n \ll p$.

SEARCHLIGHT HYPERALIGNMENT A proposed solution was to aggregate many local alignments to compose a full-brain matching, while keeping the benefits of local alignment in terms of regularization. To do so, Guntupalli and colleagues (Guntupalli, Feilong, and Haxby, 2018; Guntupalli et al., 2016), proposed to use the *searchlight* scheme (Kriegeskorte, Goebel, and Bandettini, 2006), popular in brain imaging.

In this procedure, the cortex is divided into a set of small overlapping spheres of radius r covering the brain. A local transform can then be learned in each sphere, and the full alignment is obtained by averaging across overlapping transforms. Although this procedure derives local correspondences, it is computationally costly. One needs at least $\frac{p}{r^3}$ balls to ensure that all voxels are covered (and usually way more are used to ensure balls overlap). Importantly, the aggregated transform produced is no longer guaranteed to bear the type of regularity (e.g., orthogonality, isometry, or diffeomorphicity) enforced during the local neighborhood fit.

LOW-RANK TRANSFORMS As an alternative handling of the complexity entailed by hyperalignment on large regions, Chen et al. proposed instead to do a shared dimensionality reduction across subjects prior to performing alignment (Chen et al., 2014). They suggested joint-SVD to decompose all subjects data in a common basis and in a second step performing Hyperalignment on this lower number of features. This opened the way for the use of lower rank factor models to represent data across subjects, such as the Shared Response Model detailed below.

3.1.4 *Aligning many subjects*

Pairwise matchings between two datasets, such as Procrustes, are usually hard to extend to more than two subjects and thus not suited to group studies (which above all else want to find a common representation for all subjects). In fact, to learn the matchings between all pairs of S subjects, one would need to perform S^2 alignment steps, which is computationally prohibitive in many situations. Moreover, when trying to pull information from all subjects in a common space, it is challenging to find a unique model capturing properly the specific features of each subject. Group averaged data would seem like a fine candidate to use as such a common space, but its topography shows important qualitative difference compared to individual data, which would deter alignment goals. In their original paper, Haxby and colleagues thus proposed an iterative heuristic to build a group "template" suitable for alignment (Haxby et al., 2011).

In this procedure, a subject is arbitrarily drawn as a reference. In a first pass, subjects are taken one by one, aligned to the current reference and this aligned data is averaged with the previous reference to update it. In a second pass, all subjects are aligned to this temporary "common space", and their aligned data are all averaged together to serve as a final "template" space.

This method suffers from several drawbacks. First, it receives non-equivalent contributions from various subjects to the common group model, which is thus strongly dependent on the order in which subjects were aggregated (Al-Wasity et al., 2020). Moreover, it does not follow any clear criterion of optimization, and is thus hard to trust or deem significant. It is merely an intermediate computational step. The complexity of finding better characteristics for potential templates will be detailed in Section 5.2.Another solution to this kind of problem was instead to search for common latent factors across subjects and map their full data in this low-rank basis. This approach was introduced as Shared Response Model.

3.1.5 *Shared Response Model*

The Shared Response Model (SRM), introduced in Chen et al., 2015, is a latent factor model that extracts a common response from different subjects exposed to the same stimuli and subject-specific spatial components. In practice, this shared response corresponds to the activity of a fixed number of components (or "super-voxels") summarizing common activity across subjects. It is jointly estimated with subject-specific "spatial basis" that project the full number of voxels to these super-voxels, in an orthogonal fashion. In its deterministic version, SRM estimates a common shared response $\mathbf{S} \in \mathbb{R}^{k \times n}$ and a per-subject orthogonal basis $\mathbf{W}^i \in \mathbb{R}^{p \times k}$ from subject-level alignment data $\mathbf{A}^i$,

solving Equation 3.3. n is the number of time points, p is the number of voxels, and k is a hyper-parameter indexing the dimensionality.

$$\begin{aligned} &\min_{\mathbf{W}^1,\dots,\mathbf{W}^n,\mathbf{S}} \sum_i \|\mathbf{A}^i - \mathbf{W}^i\mathbf{S}\|_F^2 \\ &\forall i,\ \ \mathbf{W}^{i\intercal}\mathbf{W}^i = \mathbf{I}_k \end{aligned} \tag{3.3}$$

SRM decomposes the signal of many subjects in a common basis, with the same orthogonality constraint as Procrustes. It mostly differs from Procrustes analysis in that it provides a decomposition of all subjects' activity at once, rather than being performed pairwise. This decomposition is low-rank, which is also a way of addressing high-dimensionality. This makes it an effective solution, especially given recent computational improvement (Richard et al., 2019). However, SRM is only effective if the number of components k is large enough to capture all distinct components in the signal.

For full-brain data decomposition, it thus provides limited resolution topographies, since the low number of estimable components is limited by the smaller number n of training samples compared to the number p of voxels. This implicit dimension reduction induces a loss of information compared to using original data. In an ROI, by contrast, we have an increased ratio of n/p which ensures both (1) more stable estimations thanks to increased number of samples (2) that we do not lose relevant information when projecting voxel signal to the *shared response* (n/p must be greater than 1 to ensure full-rank decomposition).

3.2 RELATED APPROACHES

3.2.1 *Machine learning framing: feature alignment*

Trying to minimize differences between different subjects' dataset can be described as a transfer learning problem where each subject is a different domain. More specifically, the primary goal of alignment is to relate datasets of subjects going through the same task, hence it is part of the sub-field of domain-adaptation (Farahani et al., 2020). Domain adaptation algorithms (e.g Schoenauer-Sebag et al., 2019) usually postulate a fixed feature space $\mathcal{X}$, from which each domain data x_s, x_t are drawn, most commonly with a covariate shift, meaning the marginal probabilities of their distributions differ :

$$p_s(x) \neq p_t(x)$$

The *inter-subject decoding* problem, a classification setting used as a validation task in Chapter 6, can typically be considered as a domain adaptation problem.

Functional alignment addresses the same kind of discrepancy, but as a mismatch between domains feature spaces $\mathcal{X}^s$ and $\mathcal{X}^t$. This situation is not common in the transfer learning literature, where it appears for some specific computer vision generalization tasks, natural language translation problems for low-data language (Alaux et al., 2018), to compare information encoded in various neural networks trained in the same way (Lu et al., 2018), or in general machine learning to match embeddings with slightly different dimensions under the names *hyperalignment*, *feature representation transfer* (Kaboli, 2017) or *feature alignment* (Chen et al., 2019). In the most general case, there is neither the same number of features in both spaces nor known common anchor points to relate them. In functional alignment, we rely on common *synchronized* protocols probing those feature spaces in the same way as *supervised* data to learn this mapping. Each synchronized timeframe gives one more dimension which characterizes both features sets in parallel. Each feature is characterized by a vector of p coordinates, and the set of features to match for X and Y can be seen as distributions (or point clouds) in this p-dimensional space.

When working with anatomically-aligned fMRI data, common spatial coordinates are also available to relate both feature sets. While most alignment methods explicitly ignore them, diffeomorphic alignment (3.2.2) use them to impose a strong regularity on the matching. Working on local neighborhoods as proposed in Searchlight alignment (3.1.3) is also a way of using this spatial information as a regularizer.

3.2.2 *A review of functional alignment methods*

LINEAR ALIGNMENT Although various algorithms have been proposed to solve these problems, most linear solutions revolve around a few already evoked ideas. Among the most common are Procrustes and the closely related Canonical Correlation Analysis (CCA) (Hardoon, Szedmak, and Shawe-Taylor, 2004). Equations 3.4 & 3.5 display closely matched formulations of both problem (in their generalized formulation). For each dataset, Procrustes yields an orthogonal basis that forms linear combinations of voxels responses, yielding response features shared across datasets. On the other hand, CCA seeks one set of orthonormal vectors per dataset such that those vectors are maximally correlated across datasets (Xu et al., 2012).

$$\begin{aligned} &\min_{\mathbf{R}_s, \mathbf{R}_t} \|\mathbf{A}^s\mathbf{R}_s - \mathbf{A}^t\mathbf{R}_t\|_F^2 \\ &\text{s.t.} \quad \mathbf{R_k}^\intercal\mathbf{R_k} = \mathbf{Id} \end{aligned} \tag{3.4}$$

$$\min_{\mathbf{R_s},\mathbf{R_t}} \|\mathbf{A^s R_s} - \mathbf{A^t R_t}\|_F^2$$
$$\text{s.t.} \quad \mathbf{R_k}^\intercal \mathbf{A^k}^\intercal \mathbf{A^k R_k} = \mathbf{Id} \tag{3.5}$$

In fMRI, this line of work mostly led to the aforementioned Shared Response Model (which is quite similar to a probabilistic CCA, (Bach and Jordan, 2005; Chen et al., 2015)) with the notable exception of Bilenko and Gallant, 2016 introducing a regularized and kernelized version of CCA. This line of work had recent developments, with the introduction of new multi-view algorithms (Richard et al., 2020; Xu, Yousefnezhad, and Zhang, 2018b), replacing the PCA underlying Shared Response Model with other techniques such as Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000).

NON-LINEAR ALIGNMENT Attempts to find non-linear matchings across feature spaces mostly tried to solve the Procrustes problem after suitable non-linear embeddings of each subject data. Kernel Hyperalignment (Lorbert and Ramadge, 2012) was proposed as a way to improve scalability and open the possibility to use additional features on top of images to learn alignment. The main underlying idea is to search for alignment in a kernel space (see Equation 3.6). This approach remained however limited by the need to choose an arbitrary kernel and the lack of criterion to evaluate such kernels.

$$\underset{\mathbf{R}^\intercal\mathbf{R}=\mathbf{Id}}{\operatorname{argmin}} \|\mathbf{R}\phi(\mathbf{A^s}) - \phi(\mathbf{A^t})\|_F^2 \tag{3.6}$$

In Langs et al., 2014, the authors proposed to go even further in terms of decoupling of functional and anatomical space. This seems particularly meaningful in contexts in which structure/function coupling is not guaranteed across subjects. They state that each subject's high-dimensional fMRI data is lying on its own manifold, that can be well approximated by low-dimensional geometries and matched accordingly. To leverage this assumption, they proposed to learn an embedding from the correlations of fMRI time courses. To find these subject-specific low-dimensional spaces, they use diffusion maps (Coifman and Lafon, 2006) on functional connectivity to derive 300-dimensional embeddings. They also introduce ways to aggregate these embeddings into a common "atlas" in which all subject can be compared. While this technique leverages an interesting view of data, the embedding process used is not invertible, which breaks the overall invertibility of alignment. Moreover, embedding separately the time-dimension of subjects signals destroys time-synchronization and temporal labelling of data. It is thus mainly restricted to resting-state data.

In another original formulation (Wu et al., 2021) proposed to learn a d-dimensional non-linear embedding of anatomical coordinates of voxels based on functional similarity, modelled through Gaussian processes. They argue that the constraints applied to design this embedding space ensure that neural activity changes smoothly as a function of locations in embedding space, and subsequently make a compressed representation matrix of full-brain covariance matrix. However, this non-linear embedding process is non-invertible and performed independently across subjects, which are important limitations to include it in a functional alignment framework.

The small amount of available fMRI data makes it unrealistic to solve this problem using heavily parametrized alignment models without overfitting. For completeness, it should be noted that if it wasn't for this constraint, Auto-Encoders (Liu et al., 2020; Wang et al., 2015), Generative Adversarial Networks, Deep CCA (Benton et al., 2017), and in general networks specifically trained to remove subject-specific information (e.g., including a gradient reversal layer (Ganin et al., 2016)) would be promising directions to explore. However, works that tried to apply those in fMRI are seldom (Chen et al., 2016; Yousefnezhad and Zhang, 2017) and do not translate to typical study context.

DIFFEOMORPHIC ALIGNMENT In a seminal work, Sabuncu et al. (Sabuncu et al., 2010) introduced the idea of treating functional variability as a residual of an imperfect anatomical registration. Building on this view, they proposed to introduce a second fine-tuning step where this diffeomorphic transform would be refined using functional data. Following this idea, several contributions pushed the idea to perform functional alignment as a diffeomorphic registration improving functional similarity. They mostly adapted existing diffeomorphic anatomical registration algorithms to guide them using various functional derivatives. Spatial patterns of functional response (Conroy et al., 2013), functional connectivity patterns (Jiang et al., 2013), functional connectivity low-dimensionnal embeddings from Langs et al., 2014 described previously, multi-modal features (Robinson et al., 2018), (Nenning et al., 2017), or even raw functional images (Dohmatob, Varoquaux, and Thirion, 2018) were proposed to guide registration. This means that the dissimilarity minimized by such algorithms is partly, or even entirely, based on those functional derivatives across subjects instead of being calculated on anatomical structures.

By design, these transforms are far more constrained than the ones described previously. They target precision improvement over preexisting anatomical alignment, or provide more information to guide registration. In fact, enforcing diffeomorphicity in the anatomical space is implicitly imposing a very strong spatial regularization (at least through initialization) on the spectrum of possible "functionally

guided" transforms. By contrast, other alignment approaches posit a strong level of local variability that cannot be compensated adequately with those restrictions.

3.2.3 *Beyond alignment classical set-up*

INTRA-SUBJECT ALIGNMENT Another alternative to functional alignment can be envisioned from Tavor et al., 2016. Alignments are usually learned as source-target mappings on dedicated *alignment* data: $\mathbf{A_s}$, $\mathbf{A_t}$. Commonly, they are later applied on unrelated data, that we denote $\mathbf{D_s}$. Another possibility is to learn a $\mathbf{A}$ to $\mathbf{D}$ mapping within-subjects. In their paper, Tavor and colleagues show that while individual activity patterns in each task may appear idiosyncratic, correspondences learned across different tasks using a general linear model display less inter-subject variability than individual activity maps. For example, this means that it is possible to predict idiosyncratic features of a subject activity performing an object recognition task from its movie-watching data derived features, or to infer its task-based activations from resting-state. This provides an interesting twist on the typical functional alignment workflow: while most methods learn alignments within a single task and across subjects, we can instead learn within-subject correlations across tasks. The structure of learned task-specific correlations should then hold in new, unseen subjects.

Figure 3.3 illustrates how we can learn the local-level correlation structure between two independent tasks $\mathbf{A}^s \in \mathbb{R}^{p \times n}, \mathbf{D}^s \in \mathbb{R}^{p \times d}$ within a single $source$ subject. We denote the mapping between these tasks, as $\mathbf{R}^{intra}$, to distinguish it from mappings that are learned between pairs of subjects. First, we divide both tasks' data into parcels (e.g., using a high-resolution functional atlas). On a local parcel i, each voxel is considered a sample, and we train $\mathbf{R}_i^{intra} \in \mathbb{R}^{p_i \times d}$ through Ridge regression (Hoerl and Kennard, 1970):

$$\mathbf{R}_i^{intra} = \underset{\mathbf{R}_i}{\operatorname{argmin}} \|\mathbf{A}_i^s \mathbf{R}_i - \mathbf{D}_i^s\|_F^2 + \alpha \|\mathbf{R}_i\|_F^2 \tag{3.7}$$

The hyperparameter α can easily be chosen with nested cross-validation. After repeating this procedure for several *source* subjects, we can then use the averaged of those learned correlation structures $\mathbf{R}^{intra}$ to estimate new data for $target$ subject as $\hat{\mathbf{D}}^t = \mathbf{A}^t \mathbf{R}^{intra}$.

CONNECTIVITY-BASED ALIGNMENT As stated previously, sample synchronization is usually necessary to supervise the feature matching process in fMRI. However, when such a constraint is not fulfilled, other methods have been suggested to put datasets in correspondence. In (Guntupalli, Feilong, and Haxby, 2018), the authors propose to use instead coarse features stability across datasets as a way to recreate supervision to refine their matching. To do so, they cluster features into k previously known hierarchical regions and average signal for

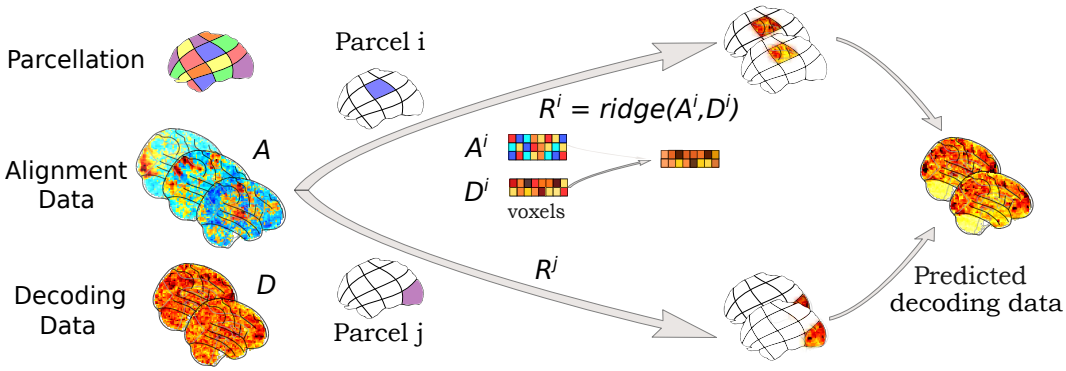

Figure 3.3: Intra-subject alignment. Using intra-subject alignment to learn piecewise correlations between a single subject's alignment and decoding task data. As with other piecewise methods, this mapping is learned separately for all parcels $i \ldots j$ of the chosen parcellation. For each parcel, we use cross-validated Ridge regression to learn the mapping between the two task conditions—alignment data $\mathbf{A}$ and independent decoding task data $\mathbf{D}$—for this source subject. For the i-th parcel, we denote this mapping as the matrix $\mathbf{R}^i$. We then aggregate these piecewise predictions into a single, whole-brain prediction $\hat{\mathbf{D}}$. During training, this prediction can be directly compared to the ground-truth decoding data, $\mathbf{D}$. At test time, we would have access to the target subject's alignment data $\mathbf{A}$ but not their decoding task data, $\mathbf{D}$.

each of these clusters. Then, the correlation of local signals with these stable "averaged signals", which is known as *functional connectivity* in neuroscience, yields a vector of length k that can then be taken as a descriptor of these features. From there, all alignment methods cited previously can be applied (e.g., Shared Response Model Nastase et al., 2020), extending alignment to cases without common synchronized protocol across subjects, and in particular to resting state acquisitions.

We will not cover these techniques further in this thesis, as we will restrain ourselves to the case where we have common time-synchronized protocols across datasets (usually subjects). In these cases, adding supplementary connectivity features do not improve matching compared to regular methods introduced in Section 3.1 (Busch et al., 2021).

REPRESENTATIONAL SIMILARITY ANALYSIS Representational Similarity Analysis (RSA) (Kriegeskorte, Mur, and Bandettini, 2008) is an alternative method which circumvents functional variability problems. Complementary to statistical parametric mapping (2.2.2) and decoding (2.2.3), it is used to study how multivariate activity patterns vary across a set of stimuli in selected regions-of-interest (ROIs). Whereas decoding is looking to discriminate across stimuli through separating the high-dimensional space of activation patterns, RSA measures the distance across patterns as vectors in this space (Haxby, Connolly, and Guntupalli, 2014). In a ROI, the dissimilarity (e.g., correlation, euclidean distance...) between activations patterns elicited by each pair of stimuli are assembled in a Representational Distance Matrix (RDM). Distances between RDMs of various ROIs can also be calcu-

lated, producing second-level RDMs that can be used to investigate how response differences across stimuli are emphasized or weakened in the set of ROI.

To compare first-level RDMs, voxel-level correspondence is unnecessary, which alleviates the effects of functional variability on this kind of analysis when performed across subjects. Put differently, it creates derived representations that are invariant to any orthogonal transformation applied to voxels. However, this is a consequence of the fact that this approach treat functional topography as unimportant, and do not provide any spatial model of common effects nor idiosyncrasies (Haxby et al., 2020). We thus didn't further include RSA in this study, since it does not help to improve functional similarity across subjects.

3.3 CRITICAL LITERATURE REVIEW

3.3.1 *Open questions*

Since its introduction until today, hyperalignment or functional alignment is often cited as a promising avenue to handle variability, improve brain modelling and ease translational research tasks such as biomarkers discovery (Dubois and Adolphs, 2016; Poldrack, 2017; Woo et al., 2017; Gratton et al., 2018; Chiba et al., 2019; Finn et al., 2020; Hamilton and Huth, 2020). However, we have no choice but to note that in this decade, studies devoted to suggest possible methodological variations around alignment outnumbered by far studies actually applying it to investigate other research questions (for a review, see Haxby et al., 2020).

Alignment has often been presented as a conceptually intriguing idea, which it is, but in this thesis, we mostly consider it as a pragmatic way to handle data scarcity and imperfection in order to improve fMRI studies generalizability. In this view, important questions remain weakly addressed in functional alignment literature, which can explain why it has not yet lived up to its potential:

- First and maybe most importantly, alignment use-cases are loosely defined, and it is unclear which method is suited for each application. It is subsequently hard for researchers to understand which method would be beneficial for their own use-cases.
- Following on the previous point, there is a lack of benchmarks comparing methods against one another, which yields a divided literature in which one has a hard time comparing methods.
- Moreover, methodological validation commonly solely relied on similarity metrics, which are hard to relate to broader use-cases.

This also prevents potential users to understand which gains can be expected from alignment.

- Crucial practical questions on how to reliably handle high-dimensional data, and datasets containing many subjects, are not completely answered in the literature.
- Alignment is a complex technique which could lead to double-dipping if used improperly, hence the lack of well-documented and computationally efficient implementations will deter any users that would have wanted to use it.

3.3.2 *Our work*

Our early work focused on introducing *optimal transport* as a reliable alignment method, which had the potential to improve distribution matching calculation. In doing so, we also proposed a new aggregation scheme as well as a template design method introduced in Chapters 4 & 5. While doing this, we realized many of the aforementioned shortcomings. Our subsequent work aimed at addressing those. We identified two major research scenarios to which alignment could be beneficial: (i) improving group studies inference power and generalizability, (ii) improving signal transfer across subjects and datasets for more powerful decoding. As our priority was to evaluate functional alignment potential in applied settings, we mostly focused on the latter direction, which provided us with a solid paradigm for an empirical benchmark (see Chapter 6). We also tried to clarify stakes and metrics of functional template creation and although no formal contribution has stemmed from this work yet, we give a quick account of our endeavors in Section 5.3.

Part II

ALIGNMENT METHODOLOGICAL DEVELOPMENTS

4 LOCAL OPTIMAL TRANSPORT ALIGNMENT

In this part, we bridge the key idea of functional alignment—finding the best matching to improve correspondence of activation patterns across individuals—with ***Optimal Transport (OT)****: a mathematical theory that had a very rapid development in the last two decades and is especially well-suited to estimating such matchings (Section 4.1). Without any regularization, optimal transport yields a sparse* ***orthogonal mapping with a strong regularity*** *resulting from mass conservation constraints. The key idea of this contribution, developed in Section 4.2, is that this regularity is well-suited to find matchings* ***bridging signals geometry*** *across subjects. The derived transport plans are invertible and preserve the structure of individual topography after deformation (especially compared to smoother Procrustes solutions). Additionally, the* ***entropic smoothing*** *procedure used to regularize optimal transport also contributes to the efficient estimation of brain mappings. As we implement this method as a local alignment procedure, Section 4.3 introduces an additional* ***piecewise aggregation scheme*** *to combine non-overlapping matching derived on many parcels. Section 4.4 describes an experimental evaluation of this method on rich multi-subject, multi-contrasts datasets. This chapter is based on work from Bazeille et al., 2019.*

4.1 A PRIMER ON OPTIMAL TRANSPORT

Searching for a mapping between two groups of voxels without taking their anatomical coordinates into account can be described as the search for a matching between two independent and unordered sets. In such case, the proper metric to consider is functional similarity between features belonging to each set. We can then consider various mathematical frameworks to derive optimal matching.

4.1.1 *Assignment problem*

The most classical related problem, known as the assignment problem, was introduced by Monge (Monge, 1781). In this formulation, the two sets contain the same number of elements, and each element from one set can be related to any element from the other set through a cost function $C : \mathcal{X}^s \times \mathcal{X}^t \to \mathbb{R}$. In our case, a natural cost function could be a measure of dissimilarity between features. The assignment problem is then to link elements from both sets in a one-to-one fashion in order

to minimize the overall cost of this assignment (defined as the sum of chosen pairs costs). Formally, we search for a bijection $f : \mathcal{X}^s \to \mathcal{X}^t$ which minimizes $\sum_{x_s \in \mathcal{X}^s} C(x_s, f(x_s))$. A well-known solution to this problem is the Hungarian algorithm (Munkres, 1957), which iteratively refine the overall matching by focusing at each step on minimally dissimilar pairs of elements.

To put this in context, searching for alignment as an optimal assignment means looking for the permutation of voxels maximizing functional similarity to another subject. Of note, a permutation is an orthogonal matrix, hence it is part of the space over which Procrustes minimizes its cost function. However, mappings obtained in such a restricted class meet strong constraints, in particular in terms of sparsity. While enforcing a one-to-one voxel correspondence seems somewhat unrealistic when dealing with inter-subject variability, it is an interesting baseline.

4.1.2 *Transportation problem*

The assignment problem is in fact a special case of the broader transportation problem. As presented in the previous section, assignment aim is to find a bijective pairing between two sets, i.e., each source element can be matched to only one target element. On the other hand, transport is concerned by matching two measures (i.e., histograms of unit norm) supported by those sets in a n-to-n fashion. In this more general formulation (Kantorovitch, 1958), admissible couplings are probabilistic, i.e., they can split the mass of a source location towards several target locations.

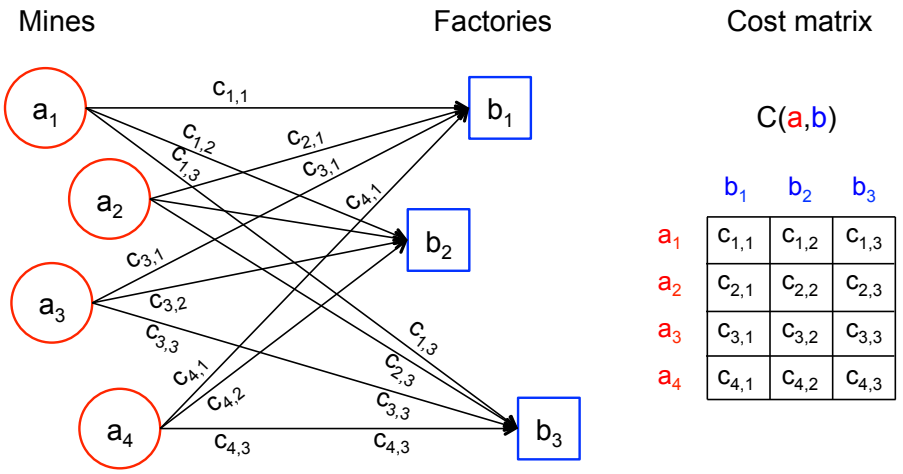

Figure 4.1: Mines and Factories problem. Resources must be transported from mines to factories at a minimal overall cost. The cost to transport a quantity of resource between histograms **a** and **b** is defined by a cost matrix **C**.

OF MINES AND FACTORIES A classical example of a transportation problem is the one said of "mines and factories", depicted in Figure 4.1, in which p mines (the source support) produce each a quantity a_i for $i \in [1, p]$ of a given resource. This resource is consumed by q factories (the target support) where each factory consumes b_j resources for $j \in [1, q]$. Both histograms to match are normalized: the mass of resource produced and consumed are equal and thus considered summing to one, such that $\sum_i a_i = \sum_j b_j = 1$.

To transport a quantity x of resource from a mine i to a factory j takes a cost $c_{ij}(x)$ depending on the distance between the mine and the factory. The transportation problem itself is to find the coupling $\mathbf{R}$ that transports the entire resource's mass from mines to factories at a minimal overall cost (which is the sum of all costs of effective transportation). This coupling is admissible only if it enforces the *mass conservation* constraint, which means that the quantity of resource displaced from a mine i must sum to exactly a_i and the amount of resource displaced to a factory j must sum to b_j.

OPTIMAL TRANSPORT More formally, let us denote two unit-norm histograms $\mathbf{a} \in \mathcal{X}^s$, $\mathbf{b} \in \mathcal{X}^t$ as the source and target histograms to match supported by both sets. As opposed to the assignment problem, the disjunction between histograms to match and their supports mean that both supports need not have the same cardinality. We'll denote $\mathbf{card}(\mathcal{X}^s) = p$ and $\mathbf{card}(\mathcal{X}^t) = q$. We also define $\mathbf{C} \in \mathbb{R}_+^{p \times q}$ a cost matrix relating both supports through the cost function $C : \mathcal{X}^s \times \mathcal{X}^t \to \mathbb{R}_+$. The objective is to a find the coupling $\mathbf{R} \in \mathbb{R}_+^{p \times q}$ which transports $\mathbf{a}$ to $\mathbf{b}$ at a minimal cost. This cost can be written as the element-wise product between $\mathbf{R}$ and $\mathbf{C}$ or simply $\sum_{i,j} \mathbf{R}_{i,j} \mathbf{C}_{i,j}$.

To be admissible, a coupling must respect mass conservation problems. The transportation problem is thus a constrained minimization problem that search for the *optimal transport plan* $\mathbf{R}^*$ solving Equation 4.1:

$$\begin{aligned} &\underset{\mathbf{R}}{\operatorname{argmin}} \sum_{i,j} \mathbf{R}_{i,j} \mathbf{C}_{i,j} \\ &\text{s.t.} \left(\sum_j \mathbf{R}_{i,j} \right)_i = \frac{1}{p}, \left(\sum_i \mathbf{R}_{i,j} \right)_j = \frac{1}{q} \end{aligned} \tag{4.1}$$

Notably, this transport plan can match a source element to several targets by dividing its mass, and can also be interpreted as a probabilistic matching between source and target histograms. In this basic formulation, the transportation problem is convex and thus yields a unique solution. The overall transportation cost enticed by this plan is called the *Wasserstein distance*: $W_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = \sum_{i,j} \mathbf{R}^*_{i,j} \mathbf{C}_{i,j}$.

Given two sets related by a cost function, the Wasserstein distance can be seen as a natural geometrical distance to compare probability distributions. In our work, we are mostly interested in finding the *optimal transport plan* which can be interpreted as an optimal matching between subjects data (our detailed methodology, including additional regularization of this minimization, is described in Section 4.2).

WASSERSTEIN BARYCENTERS Let us define a set of S histograms living on the same support: $(\mathbf{x}^s)_{s=1}^S \in \mathcal{X}$. When $\mathcal{X}$ is equipped with a distance d, a common problem is to derive their barycenter $\mathbf{t}$ (also called Fréchet-mean for distance d) solving Equation 4.2.

$$\underset{\mathbf{t}\in\mathcal{X}}{\operatorname{argmin}} \sum_{s=1}^{S} d(\mathbf{x}^s, \mathbf{t}) \tag{4.2}$$

When d is the ℓ_2-norm, the barycenter comes back to the euclidean mean, and for the ℓ_1-norm the barycenter becomes the median. While Equation 4.2 can be difficult to solve for a generic distance, it is a convex problem when using the Wasserstein distance (Peyré and Cuturi, 2018). It is possible to derive a barycenter where each $\mathbf{x}^s$ comes with a different cost matrix $\mathbf{C}_s$. However, in the typical *eulerian* set-up, all cost matrices are the same and define a single Wasserstein distance W such that the Wasserstein barycenter $\mathbf{t}$ minimizes Equation 4.3. Using entropic regularization, this minimization can also be solved efficiently using Sinkhorn algorithm.

$$\underset{\mathbf{t}\in\mathcal{X}}{\operatorname{argmin}} \sum_{s=1}^{S} W(x^s, t) \tag{4.3}$$

UNBALANCED OPTIMAL TRANSPORT When trying to apply optimal transport to real world noisy data, the mass conservation hypothesis is rarely fulfilled. Thus, a common extension of the aforementioned problem is the one of "unbalanced" optimal transport, which relaxes the mass conservation constraints using a divergence $\mathbf{D}$. The transport plan marginals are allowed to be only approximations of $\mathbf{a}$ and $\mathbf{b}$, as penalized through $\mathbf{D}$. A parameter τ is introduced to control how much mass variations are penalized compared to transportation cost (Peyré and Cuturi, 2018). Equation 4.1 can then be generalized as Equation 4.4:

$$\underset{\mathbf{R}}{\operatorname{argmin}} \sum_{i,j} \mathbf{R}_{i,j}\mathbf{C}_{i,j} + \tau\mathbf{D}(\mathbf{R}\mathbb{1}, \mathbf{a}) + \tau\mathbf{D}(\mathbf{R}^\intercal\mathbb{1}, \mathbf{b}) \tag{4.4}$$

Where $\mathbf{R}\mathbb{1}$ denotes $\mathbf{R}^\intercal\mathbb{1}$ the right and left marginals of the coupling, respectively denoted in Equation 4.1 as $\left(\sum_j \mathbf{R}_{i,j}\right)_i$ and $\left(\sum_i \mathbf{R}_{i,j}\right)_j$

A usual divergence to use in this context is the Kullback-Leibler divergence. Efficient algorithms to estimate unbalanced Wasserstein barycenters can also be derived (Benamou et al., 2015; Chizat et al., 2018; Janati, Cuturi, and Gramfort, 2020).

4.2 OPTIMAL TRANSPORT ALIGNMENT

PREVIOUS WORK In Gramfort, Peyré, and Cuturi, 2015, the authors proposed to apply optimal transport to match brain signals across subjects in order to build group averages better accounting for variability. In this model, the authors define signals as the histograms to match and the spatial distance between each signal location as the cost of displacement. The key idea here is, for one contrast, to match similar *quantity of activations* even when they display a slight spatial shift, and calculate a Wasserstein barycenter of subjects' activations. They illustrate this method both on fMRI data and in the context of the MEG inverse problem. This nicely fits the intuition that activation foci should be moved conservatively to match individuals, which makes it quite alike diffeomorphic alignment.

When considered as a potential tool for functional alignment, this approach suffer from several limitations. First, it is calculated for each timeframe separately, which means that, when applied to several images or contrasts, the underlying voxel displacement induced on each frame will be unrelated. Hence, it lacks consistency and generalizability for practical applications. Pragmatically, we must also note that fMRI signals *are not measures*: they are not entirely positive nor normalized across subjects. This introduces the need to resort to unbalanced optimal transport and additional computational heuristics and introduce additional hyperparameters which are hard to set. More fundamentally, this kind of matching does not convey functional similarity across subjects, but merely tries to match sparse blobs of brain activity.

OUR APPROACH We chose to apply optimal transport in another way, building on functional alignment intuitions. Instead of searching for an explicit matching between observed signals, we searched for the underlying matching between voxels as characterized by their functional similarity. We thus defined the sets of voxels to match directly as our measures and defined a cost matrix conveying functional dissimilarity across voxels.

NOTATIONS Let $p \in \mathcal{N}$, we denote $[p]$ the set of integers from 1 to p. Let $\mathbf{A_i} \in \mathbb{R}^p$, $\delta_{\mathbf{A_i}}$ will denote the Dirac mass at location i. Given a brain region comprising p voxels, we consider the d-dimensional signals observed in these voxels $\{\mathbf{A}_1, ..., \mathbf{A}_p\}$. Here, these d-dimensional

signals correspond to d activation maps observed in a given subject. We denote by $\mathbf{A}$ the $p \times d$ matrix obtained by concatenating those vectors. $\|\|_F$ denotes Frobenius norm, and $\mathrm{tr}(.)$ the trace operator.

CORRESPONDENCES FROM AN OPTIMAL TRANSPORT GEOMETRY PERSPECTIVE Let us consider the set $\mathbf{A}^s$ of functional signals in a given source subject s. Following the intuitions of Langs et al., 2011, the p vectors $\{\mathbf{A}^s_1, ..., \mathbf{A}^s_p\}$ together make up a measure μ^s lying on a latent manifold $\mathcal{F}_s$ embedded in $\mathbb{R}^p$. The discrete measure μ^s with positive weights $\mathbf{w}^s > 0$ and support $\{\mathbf{A}^s_1, ..., \mathbf{A}^s_p\} \in \mathcal{F}_s$ is defined as $\mu^s = \sum_{i=1}^p \mathbf{w}^s_i \delta_{\mathbf{A}^s_i}$. Note that in the present framework $\mathbf{w}^s = [\frac{1}{p}...\frac{1}{p}]$.

The difference between two measures $\mu^s = \sum_{i=1}^p \mathbf{w}^s_i \delta_{\mathbf{A}^s_i}$ and $\mu^t = \sum_{i=1}^q \mathbf{w}^t_i \delta_{\mathbf{A}^t_i}$, reflects the differences between individuals s and t. However, for $s \neq t$, the support of the $\mathcal{F}_s$ and $\mathcal{F}_t$ manifolds are distinct in general: two subjects do not exhibit the same set of responses, due to intrinsically different brain organization. Directly computing Kullback-Leibler divergence between μ^s and μ^t is useless, as non-coincident support leads to infinite values; fixing this mismatch by smoothing induces a loss of information. By contrast, the Wasserstein distance between μ^s and μ^t is well-defined (Peyré and Cuturi, 2018).

In this framework, distance evaluation is tightly linked to functional alignment, as it is formulated as the task of finding an optimal coupling $\mathbf{R}^* \{\mathbf{A}^s_1, ..., \mathbf{A}^s_p\} \rightarrow \{\mathbf{A}^t_1, ..., \mathbf{A}^t_q\}, \mathbf{R} \in \mathbb{R}_+^{p \times q}$. Enforcing signal conservation and optimality of the alignment cost $\mathbf{C}(\mathbf{A}^s, \mathbf{A}^t)$ yields:

$$\begin{aligned} &\mathbf{R}^* = \min_{\mathbf{R}} \sum_{i,j} \mathbf{R}_{i,j} \mathbf{C}(\mathbf{A}^s, \mathbf{A}^t)_{i,j} \\ &\text{s.t.} \left(\sum_j \mathbf{R}_{i,j} \right)_i = w^s_i \text{ and } \left(\sum_i \mathbf{R}_{i,j} \right)_j = w^t_j \end{aligned} \tag{4.5}$$

To use the discrepancy of functional features as a cost function, we define:

$$\forall i, j \in [p] \times [q], \mathbf{C}(\mathbf{A}^s, \mathbf{A}^t)_{i,j} = \|\mathbf{A}^s_i - \mathbf{A}^t_j\|_2^2$$

and couple the input measures μ^s and μ^t, where all voxels have a constant weight, respectively $1/p$ and $1/q$. If both subjects functional data share a common number of voxels (i.e., $p = q$), and we search for a deterministic coupling, this falls back to the assignment problem(4.1.1).

ENTROPIC REGULARIZATION. We define the entropy of a coupling as

$$h(\mathbf{R}) = -\sum_{i,j} \mathbf{R}_{i,j} (\log(\mathbf{R}_{i,j}) - 1)$$

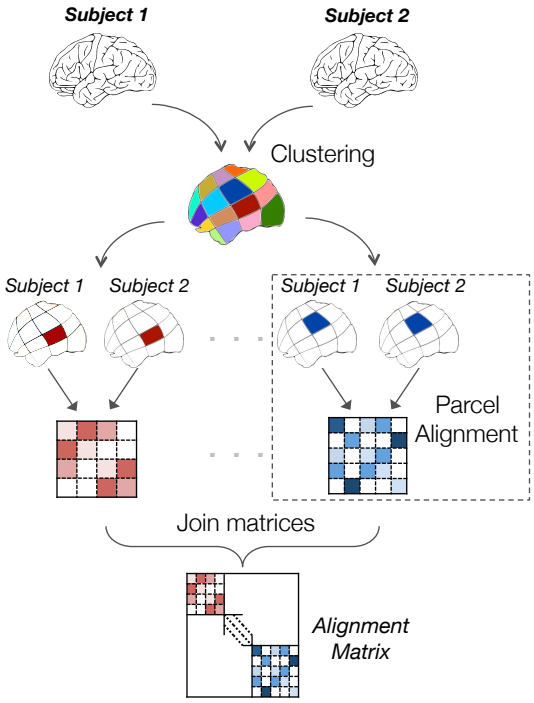

Figure 4.2: Piecewise alignment methodology. In a first step, a parcellation is derived using functional clustering or a functional atlas. Local alignments are learned between subjects on each parcel. The whole-brain alignment is recomposed as a block-diagonal aggregation of local alignments.

and use it as a regularization function in Equation 4.5, which becomes :

$$\begin{aligned} \mathbf{R}^* &= \min_{\mathbf{R}} \sum_{i,j} \mathbf{R}_{i,j} \mathbf{C}(\mathbf{A}^s, \mathbf{A}^t)_{i,j} - \epsilon h(\mathbf{R}) \\ \text{s.t.} \left(\sum_j \mathbf{R}_{i,j} \right)_i &= \frac{1}{p}, \left(\sum_i \mathbf{R}_{i,j} \right)_j = \frac{1}{q} \end{aligned} \tag{4.6}$$

The entropic term makes the objective function ϵ-strongly convex, hence leading to a unique optimal solution for a given ϵ. Besides making computation of transport faster using Sinkhorn algorithm (Cuturi, 2013), this entropic regularization also acts as a smoothing of the solution.

4.3 PIECEWISE ALIGNMENT

Most alignment methods perform well in a local setting, and optimal transport is no exception. Computational cost becomes prohibitive for tens of thousands of voxels, and searching for such high-dimensional matching introduces the need for more samples than available and deters the idea of picking a strongly regularized transformation. As introduced in Section 3.1.3, *Searchlight* was proposed as an aggregation

procedure to circumvent this problem. It consists in searching for local alignments on overlapping local neighborhoods and in summing these transforms to recompose a full-brain alignment. However, it is computationally very costly and the aggregate transform lacks the regularity enforced locally. Enforcing sparsity of local transforms using optimal transport to later average them into a dense global one would make little sense.

To overcome this limitation, we propose an alternative aggregation scheme: *piecewise alignment*. In a first step, we do a functional clustering of data to find local non-overlapping clusters $c_1, \cdots, c_K$ of voxels with common activity patterns (those parcels could alternatively be derived from a functional atlas). In each of these clusters, we find the optimal alignment transform and concatenate these local transforms to recover a full-brain transform with the desired regularities. Formally, the optimal alignment transform to align two subjects $\mathbf{x}^s_{train}$ and subject $\mathbf{x}^t_{train}$ on the training session is obtained by solving the problem in each cluster $c \in \{c_1, \cdots, c_K\}$:

$$\mathbf{R}^*[c] = \mathbf{R}^*(\mathbf{A^s}_{train}[c], \mathbf{A^t}_{train}[c]) \tag{4.7}$$

On the test session $\mathbf{A^t}_{test}$ is predicted using $\mathbf{A^s}_{test}$ by:

$$\forall c \in \{c_1, \cdots, c_K\} \mathbf{A^t}_{test}[c] = \mathbf{R}^\star[c]\mathbf{A^s}_{test}[c] \tag{4.8}$$

This returns a full-brain transformation matrix with the desired regularities, but could theoretically induce boundary effects in the alignment recovered. In Chapter 6 we compare Piecewise Alignment with Searchlight Alignment and show its efficiency. In terms of parcellation, several algorithms perform well for functional clustering. Moreover, our later work show that the parcellation method chosen and its resolution (or working with a functional atlas instead) doesn't have much impact on the quality of the alignment learned (see Section 6.5.3).

4.4 EXPERIMENTAL VALIDATION

DATASETS To benchmark these methods and assess their prediction accuracy, we ran experiments on two datasets, where individual data were previously registered in MNI-space following standard procedure (SPM12 software called though Nipype for IBC, HCP pipeline for HCP). The "Individual Brain Charting" (IBC) (Pinho et al., 2018) contains scans of the same 13 participants for a wide variety of cognitive tasks. The data were acquired using a 3T scanner (acquisition resolution of 1.5mm resampled at 3mm after spatial normalization). We worked directly on activation maps: for a given functional contrast, they yield an activation statistic at each voxel. We learn alignment between subjects between $d = 53$ contrasts derived from data acquired

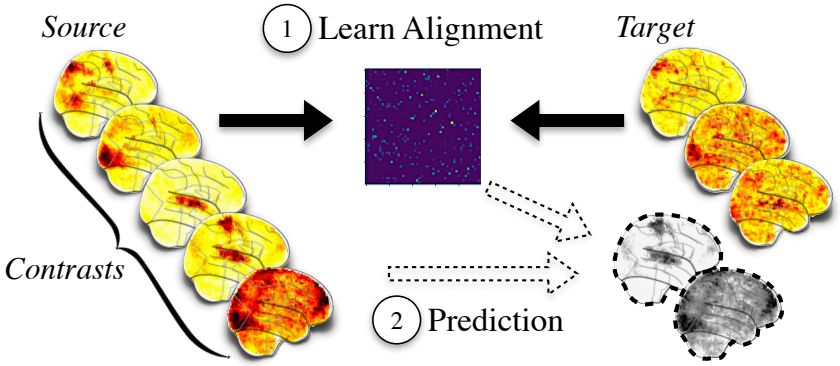

Figure 4.3: Pairwise prediction of target subject from aligned source subject. In step (1), alignment between *source* and *target* subjects is learned on *alignment data*. During step (2), it is applied on *additional* source subject data to predict corresponding maps of target subject. These predicted maps are compared against corresponding *left-out* target data.

with antero-posterior (AP) EPI phase encoding; to assess the quality of our predictions, we also use 53 contrasts acquired in separate sessions using the same experimental paradigms with posterior-anterior (PA) EPI phase encoding. Note that the resulting AP/PA distortions were estimated and corrected with FSL's topup software prior to image pre-processing.

Human Connectome Project (HCP, 2.3.1, Van Essen et al., 2013) is a collection of neuroimaging and behavioral data on 1,200 normal young adults, aged 22-35. For our experiment, we focused on 20 randomly chosen subjects. For each, we used the $d = 25$ statistical maps available in both left-to-right (LR) and right-to-left (RL) phase encoding, resampled at 3mm after spatial normalization. We learn alignment between subjects using LR images and assess prediction on RL acquisitions.

PAIRWISE ALIGNMENT BENCHMARKS In pairwise prediction, we first learn the optimal alignment operator between a source and a target subject on training data. We then use this alignment and supplementary images of the source subject to predict additional data of the target subject and score this prediction against the true target subject images using a prediction metric (see Figure 4.3).

We use this straightforward set-up on IBC and HCP datasets to compare prediction performance of several alignment methods (detailed below) and applied on a parcellation of the brain: (i) Scaled-Orthogonal Transform, (ii) Ridge Regression, (iii) Optimal Permutation, (iv) Optimal Transport with entropic smoothing. We compare these to two baselines, the identity transform (that predicts the target subject as the source subject data), and a multi-purpose state-of-the-art diffeomorphic medical image registration algorithm: symmetric image normalization (SyN)(Avants et al., 2008).

In a second experiment, we study the influence of the amount of entropic regularization used in OT loss on both datasets. Obviously, some pairs of subjects have more similar functional data than others. To make our evaluation process robust to this variability, we tested every method on the same set of 20 pairs of subjects chosen randomly in each dataset.

IMPLEMENTATION DETAILS Here we give a few details about our implementation of methods present in this benchmark. (i) Scaled-Orthogonal Transform (reimplementation of hyperalignment, 3.1.2) solves the scaled Procrustes problem and yields an orthogonal transform. (ii) Ridge Regression (Hoerl and Kennard, 1970), further relaxes this orthogonality constraint, and simply look for a linear coupling $\mathbf{R}$ with a small norm. For computational efficiency, we consider here ℓ_2 norm penalization, yielding $\min_{\mathbf{R}} \|\mathbf{A^t} - \mathbf{R}\mathbf{A^s}\|_F^2 + \frac{\lambda}{2}\|\mathbf{R}\|_2^2$. Using such a model, the alignment problem boils down to a ridge regression also solvable in closed form:

$$\mathbf{R}^* = \mathbf{A^t}\mathbf{A^{s\intercal}}(\mathbf{A^s}\mathbf{A^{s\intercal}} + \lambda \mathbf{I_d})^{-1}$$

(iii) Optimal Permutation search for a permutation of voxel improving functional similarity (4.1.1) (iv) Optimal Transport with entropic regularization (4.2). Notably, in our experimental case where $p = q$, permutations and every coupling acceptable in strict optimal transport sense (Equation 4.5) are part of the broader class of orthogonal transforms, however their stronger regularization could help to preserve signal structure better than Procrustes solution. By contrast, entropic regularization (Equation 4.6) yields a non-orthogonal solution in general.

In this study, we applied all alignment on whole-brain data, using piecewise alignment with *Hierarchical K-means* parcellation algorithm. Hierarchical K-means (HK-means) is a recursive K-means method aimed at efficiently deriving well-balanced clusterings. For a target number of k regions, HK-means performs a first clustering into $\sqrt{k}$ pieces, and each of them is clustered in turn into $\sqrt{k}$ parts.

We include SyN, as an additional diffeomorphic baseline in this study. SyN yields a diffeomorphic mapping maximizing Mattes mutual information between local regions. Since it works only for scalar images, it was applied only on the principal components of the training set of images. Our implementation relies on Nilearn for data handling, open-source solvers (Scikit-learn, POT for Sinkhorn, antspy for SyN).

PREDICTION METRICS To measure the quality of our prediction $\mathbf{R}\mathbf{A}^{\mathbf{s}}$, against the ground truth $\mathbf{A}^{\mathbf{t}}$ at the voxel level, we defined η^2, the normalized reconstruction error, as:

$$\eta^2_\star(\mathbf{A}^{\mathbf{t}}, \mathbf{R}, \mathbf{A}^{\mathbf{s}}) = 1 - \frac{\sum_{i=1}^{n} (\mathbf{A}^{\mathbf{t}}{}_i - \mathbf{R}_i \mathbf{A}^{\mathbf{s}}{}_i)^2}{\sum_{i=1}^{n} \mathbf{A}^{\mathbf{t}2}_i},$$

A η^2 value of 0 means that the quality of the prediction is equivalent to predicting 0 along all the dimensions. A perfect prediction yields a value of 1. To focus on the prediction improvement that can be made through alignment—independently of the preexisting distance between data to align—we assess performance quantitatively using a reconstruction ratio R_{η^2}. This ratio is also defined at voxel level and is superior to 0 if the voxel is predicted better by aligned data than by raw data.

$$R_{\eta^2_\star}(\mathbf{A}^{\mathbf{t}}, \mathbf{R}, \mathbf{A}^{\mathbf{s}}) = 1 - \frac{\sum_{i=1}^{n} (\mathbf{A}^{\mathbf{t}}{}_i - \mathbf{R}_i \mathbf{A}^{\mathbf{s}}{}_i)^2}{\sum_{i=1}^{n} (\mathbf{A}^{\mathbf{t}}{}_i - \mathbf{A}^{\mathbf{s}}{}_i)^2} = 1 - \frac{1 - \eta^2_\star(\mathbf{A}^{\mathbf{t}}, \mathbf{R}, \mathbf{A}^{\mathbf{s}})}{1 - \eta^2_{id}(\mathbf{A}^{\mathbf{t}}, \mathrm{Id}, \mathbf{A}^{\mathbf{s}})}$$

RESULTS Figure 4.4 shows that functional alignment methods generally improve prediction quality from one subject to another with respect to the identity, though not uniformly over the cortex. Sensory and motor regions typically obtain high scores, showing the stability of the signals across subjects in these areas; by contrast, other regions obtain low scores overall. Syn offers no improvement of prediction scores, nor does the optimal permutation of voxels. This means that a strict one-to-one mapping of voxels is not suitable for functional alignment. For the three other methods, we clearly see different behavior between regions with high signal-to-noise ratio (SNR) and regions with lower SNR. Figure 4.5 (a-b) report the compared distributions of predictions ratios R_{η^2} on IBC and HCP datasets and are consistent with previous observations. Ridge and OT outperform all other methods on IBC dataset, whereas Scaled Orthogonal and Ridge perform slightly better on HCP dataset. Figure 4.5 (c) shows that entropic regularization strongly improves prediction scores up to an inflexion point.

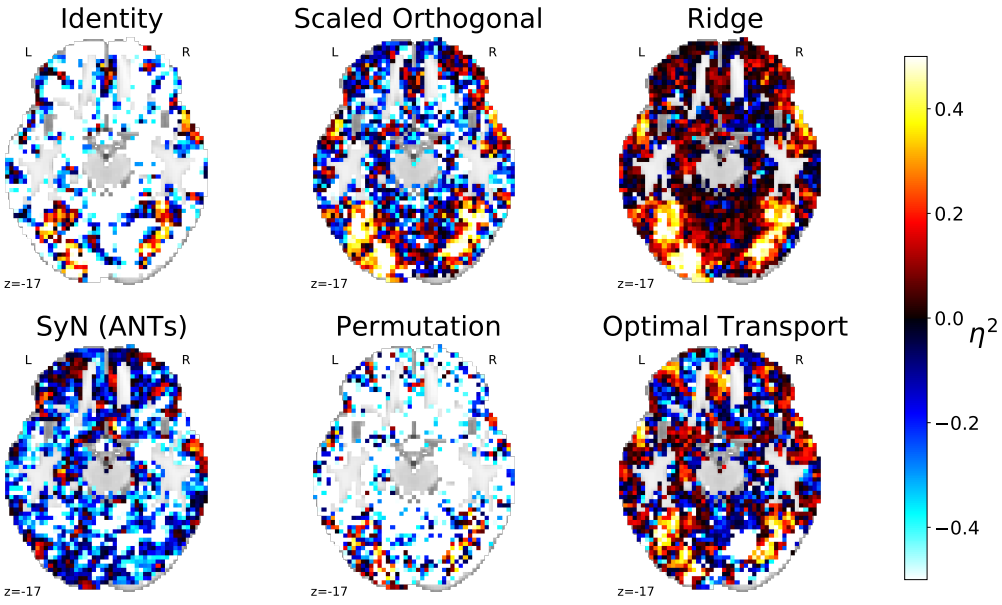

Figure 4.4: Qualitative display of pairwise alignment methods prediction accuracy. Prediction accuracy is measured through *normalized reconstruction error* $\eta^2(\mathbf{A^t}, \mathbf{R}, \mathbf{A^s})$ of target subject 9 using alignment with subject 15 (IBC dataset, z=-17 mm). A η^2 value of 0 means that the quality of the prediction is equivalent to predicting 0 along all the dimensions.

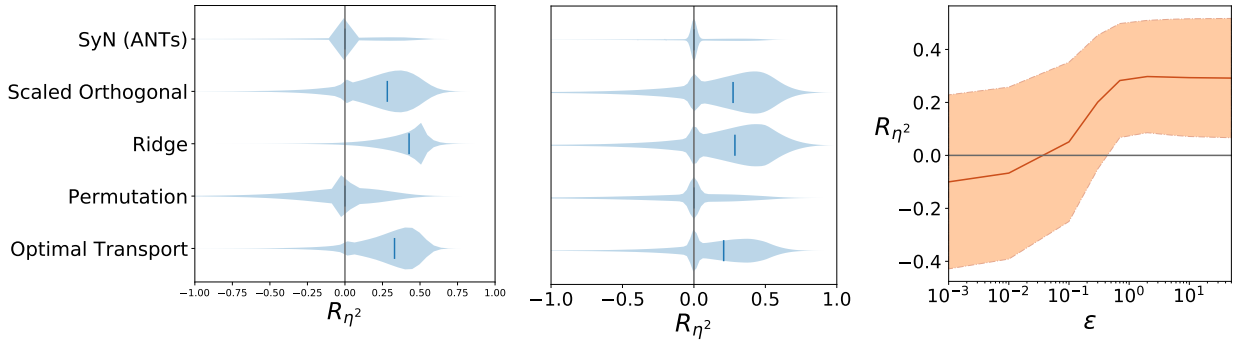

Figure 4.5: Quantitative assessment of pairwise alignment methods prediction accuracy.: Accuracy is measured using *reconstruction ratio* R_{η^2}, which is superior to 0 if the voxel is predicted better by aligned data than by raw data. The two left panels display R_{η^2} distribution across voxels when aligning 20 pairs of subject from IBC (left) and HCP (middle). (right) panel display effect of Optimal Transport entropic regularization parameter on averaged R_{η^2} for IBC dataset.

DISCUSSION

In this chapter, we introduced a new alignment method based on Optimal Transport, as published in Bazeille et al., 2019. Defining functional dissimilarity as a cost function, we propose to search for optimal transport plan of voxels. We derive those matchings on local

non-overlapping parcels and use them to recompose a whole-brain transformation. We validate its performance of this method on two datasets. Although potentially more conservative than some less regularized solvers, it is also a mean to better preserve signal specificity for later analysis. A limitation of this method is that it is a *pairwise* matching of subjects. In Chapter 5, we develop potential ways to estimate good common representations to which all subjects could be aligned. In Chapter 6, we assess the performance of both Optimal Transport alignment and the Piecewise alignment scheme to improve *inter-subject decoding*.

5 ALIGNMENT TO A TEMPLATE

In Chapter 4, we have introduced a new technique to derive a pairwise matching: a transform mapping one subject to another. In section 3.1.4, we have also motivated the need for a ***common representation space*** *to build an efficient framework to* ***transfer signal*** *across many subjects and introduced state-of-the-art methods. Beyond this computational aim, an overarching goal of alignment is also to help uncover more reliable group—and even population—brain function organization principles. In this view, our ability to draw improved* ***group statistical inference*** *hinges on the quality of the common representation chosen for individual data. In Section 5.1, we outline the challenges of finding a suitable template to do so and the limits of currently available approaches. In Section 5.2, we introduce an alternative iterative templating scheme and assess its relevance (Bazeille et al., 2019). Finally, Section 5.3 details our further attempts at improving template-related metrics and using optimal transport to build an adequate template.*

5.1 CHALLENGES OF FUNCTIONAL TEMPLATE DESIGN

For a long time, brain imaging studies have been infused with the concept of a stereotaxic space, a standardized brain representation (in terms of orientation and spatial coordinates) to ease combination and comparison of data and results from several subjects and studies (Evans et al., 2012). Among several options, the standard space has turned out to be a combination of (i) a well-defined 3D stereotaxic coordinate space, usually a representative 3D structural MRI called a "template" (or tissue probability maps for segmentation-based approaches) (ii) a class of mappings to put brains "native" space in correspondence to that template space. While both components of this mapping can vary depending on the precise use-case, the most common spatial normalization procedures map diffeomorphically each subject structural MRI to the *MNI152* template (see Section 2.1.3).

Functional alignment can be thought to be part of an additional normalization process guided by functional data. Pairwise alignment methods (as reviewed in Section 3.2 or introduced in Chapter 4), with their pairwise formulation, can be seen as the mapping class. However, this procedure commonly lacks an appropriate standard functional space (or functional "template") to which subjects brain activations

could be well-represented after functional alignment. When designing such a template, one faces two related challenges:

- enable good functional matching without including idiosyncratic features
- create a high-SNR representation of group effects, with a sharp functional topographic organization.

Since these characteristics are hard to measure, a common solution is instead to use similarity after functional alignment to the template space as a sole measure of its optimization and validation. Although this is a necessary constraint, it is not enough to ensure meaningfulness of the derived representation. In Section 5.3, we discuss alternative metrics that could be used for this purpose.

RELATED WORK Beyond the Euclidean mean of subjects images, which does not create a suitable topographic organization to represent individual data (Laumann et al., 2015), two main template estimation procedures have been proposed in the literature: *greedy bagging* of hyperaligned data (3.1.4, Haxby et al., 2011) and the *latent factor* approach to template from the Shared Response Model (3.1.5, Chen et al., 2015). While these are pragmatic solutions, they reduce the template to an intermediate variable of the alignment procedure, almost like a nuisance parameter. The templates proposed by these methods are actually never shown nor validated in the studies that introduce them.

The hyperalignment templating method (Haxby et al., 2011) yields a full-brain template that could be a candidate representation summarizing group activations at each frame after alignment. However, as an unbalanced algorithm where subjects contribute unequally to the final group representation, the derived template (and its performance to transfer signal across subjects) depends heavily on the order in which subjects were considered (Al-Wasity et al., 2020). Put differently, it does not follow any clear definition or criterion of optimization. Moreover, it is implemented on whole-brain using the searchlight scheme, which will cause its spatial organization to be smoother than individual data.

On the other hand, shared response modelling is a procedure jointly decomposing subjects data on a common basis in a balanced way. This low-dimensional basis can be thought of as a template. However, drawing statistical group inferences in this basis would make little sense and has not been carried out so far.

Importantly, these templates display a strong spatial indeterminacy, since applying any additional orthogonal transform to templates and their inverse to mappings would yield equally valid alignment schemes. Put differently, there are an infinity of templates solving the basic optimization problem. Since functional alignment does not

enforce any anatomical regularity, spatial activations observed on most of those valid templates would be meaningless.

5.2 ITERATIVE TEMPLATE GENERATION

In order to try to overcome those constraints, our article (Bazeille et al., 2019) introduced a new general-purpose iterative procedure to derive a principled template and alignments using any pairwise method. This method progressively refines a template through alternate minimization for a given class of functional alignment. It can be applied together with an aggregation scheme to derive whole-brain templates, here piecewise alignment (4.3). Reusing image-based statistics introduced in Section 4.4, we show that this new template-based alignment method helps increase similarity across subjects while being better principled than methods previously available in the literature.

5.2.1 *Template generation procedure*

Pairwise correspondences do not scale well with many individuals. A template measure $\mathbf{T}$ is needed, which can be obtained by solving

$$\min_{\mathbf{T},\mathbf{R}_1 \dots \mathbf{R}_n} \sum_{s=1}^{n} \|\mathbf{R}_s(\mathbf{T}) - \mathbf{X}^s\|_\star^2 \tag{5.1}$$

for the chosen loss $\|.\|_\star$ (Wasserstein or Frobenius). We solve it through alternate minimization iterating over :

- a $\mathbf{R}$-step of independent alignment of the current template to every sample, thus estimating $\mathbf{R}_i, i = 1..n$
- a T-step where $\mathbf{T}$ is regressed to minimize jointly its distance to the samples.

$$\min_{\mathbf{R}_s} \|\mathbf{R}_s(\mathbf{T}) - \mathbf{X}^s\|_\star^2, \forall s \in [n] \qquad \mathbf{R}\text{-step}$$

$$\min_{\mathbf{T}} \sum_{s=1}^{n} \|\mathbf{R}_s(\mathbf{T}) - \mathbf{X}^s\|_\star^2 \qquad \text{T-step}$$

Note that, for all norms considered here, the T-step results in a quadratic problem solved by conjugate gradient. Crucially, Equation 5.1 is only element-wise convex. Starting from a fixed initialization, alternate minimization will thus recover a unique minimum, but without any guarantee that it is a global minimum. Moreover, template and alignments still suffer from a strong spatial indeterminacy, since their defining criterion is invariant to orthogonal transforms. As initialization, we set all alignment operators to the identity, thus the first

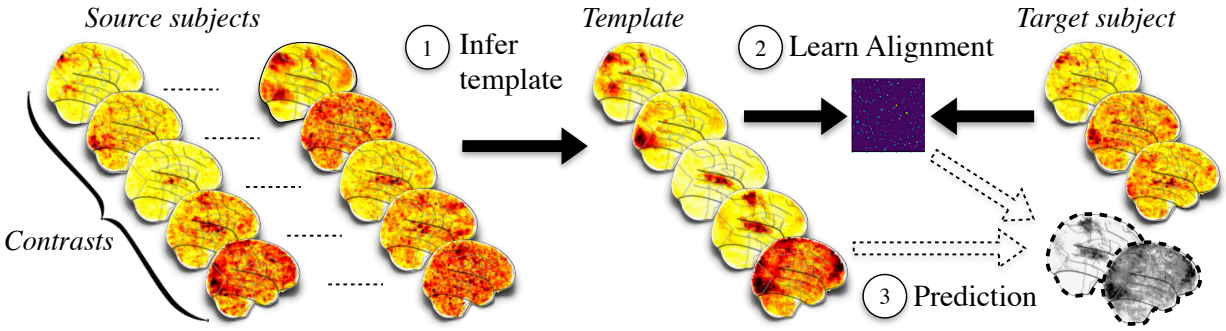

Figure 5.1: Template-based prediction of left-out subject data. In step (1), a template representing all available contrasts of *source subjects* is learned during 4 iterations of alternate minimization. In step (2), pairwise matchings are derived from the template towards each *target* subject using common *alignment data*. In step (3), those alignments are applied to *additional* template data to predict corresponding maps of target subject. Predicted maps are compared against corresponding *left-out* target data.

T-step is $\min_{\mathbf{T}} \sum_{s=1}^{n} \|\mathbf{T} - \mathbf{X}^s\|_2^2$ and the first template is the sample mean. In practice, we run 4 iterations of the alternate minimization, as we found that was sufficient for convergence.

5.2.2 *Experimental validation*

In Section 4.4, we introduced a benchmarking experiment to compare Optimal Transport alignment to other algorithms. In this section we reused the same set-up and metrics to assess template-based alignment performance using (i) Scaled-Orthogonal Transform, (ii) Ridge Regression, (iii) Optimal Permutation, (iv) Optimal Transport with entropic smoothing, all applied in a piecewise fashion.

To evaluate template-based prediction accuracy, we split the IBC dataset randomly into two folds of 7 and 6 subjects. We estimated a template from the train subjects using all AP and PA contrasts (AP and PA contrasts are the same contrasts derived from two independent acquisition repeating the same protocol while varying the phase-encoding direction, as detailed in 4.4). We then learned an alignment operator between each of the test subjects and the template using AP contrasts and try to predict their PA contrasts that we scored using predictions metrics, as illustrated in Figure 5.1. We compare the results of the same methods quantitatively in terms of prediction loss. In a second experiment, we infer a template on which we learn alignment for all subjects using AP data. We then apply these alignments to left out PA data to bring all our subject in a common space. In this common space, we run a one sample-test and compare the group effects detected by each method, for specific conditions.

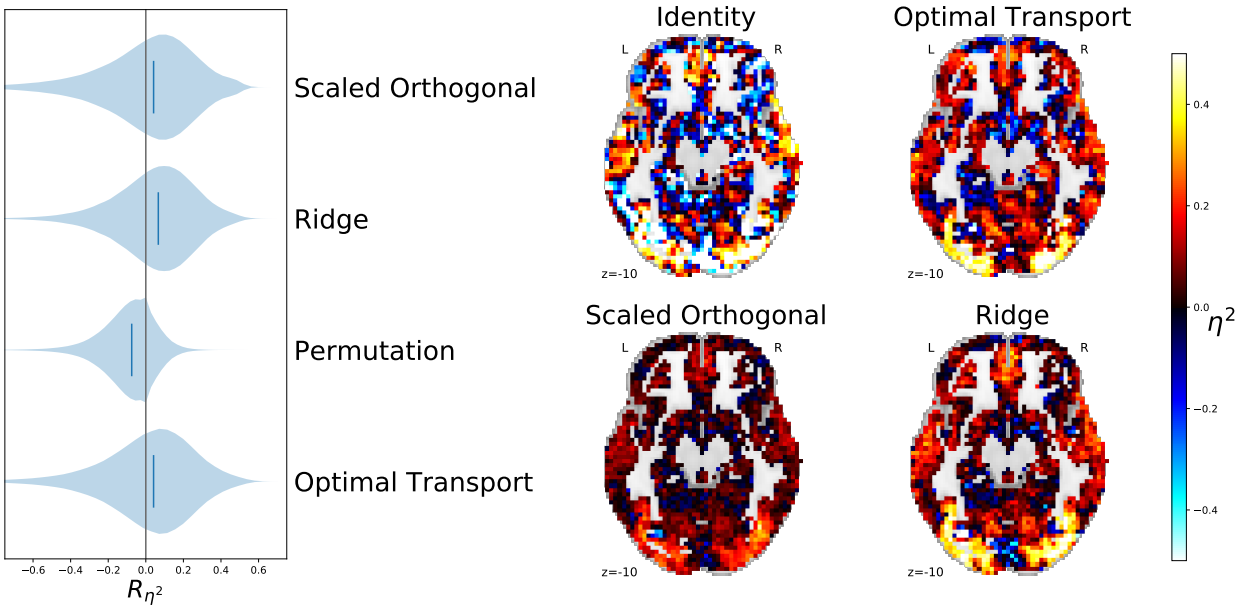

Figure 5.2: Template-based prediction performance. (left) Quantitative assessment of prediction accuracy for 6 left-out subjects of IBC dataset, measured using *reconstruction ratio* R_{η^2}, i.e. gain from alignment over prediction from group average. Performance is weaker than pairwise alignment accuracy (Figure 4.5) (right) Qualitative display of missing data prediction accuracy (*IBC-subject 11*, z=-17mm). Accuracy is measured through *normalized reconstruction error* $\eta^2(\mathbf{A^t}, \mathbf{R}, \mathbf{A^s})$. A η^2 value of 0 means that the quality of the prediction is equivalent to predicting 0 along all the dimensions. Overall, prediction performance seems lower but more consistent than observed in pairwise prediction (Figure 4.4).

5.2.3 *Validation results*

In Figure 5.2 left panel, we can observe that prediction accuracy is strongly improved by the use of a learned template in most brain regions, which establishes that functional alignment performs well to estimate cross-subject correspondences and identify a latent brain activity template. It also shows that our template estimation procedure manages to capture some inter-subject variability when mapping individual data. However, it must be noted that, compared to the performance achieved by pairwise alignment (reported in Figure 4.5), template-alignment gains are lower. This is possibly a consequence of the performance lost to the sub-optimality of this group representation to represent each subject data.

Figure 5.2 right panel shows that, despite the fact that templates derived using Scaled Orthogonal, Ridge and Optimal transport are equally accurate at predicting new subject data overall, they do so in dissimilar ways. Ridge regression tends to predict 0 in regions with low-SNR. This strong smoothing effect comes at the expense of providing predictions that are not very precise for high-SNR regions. On the contrary, Optimal Transport makes large mistakes in low-SNR regions but predicts high-SNR regions more accurately. This behavior

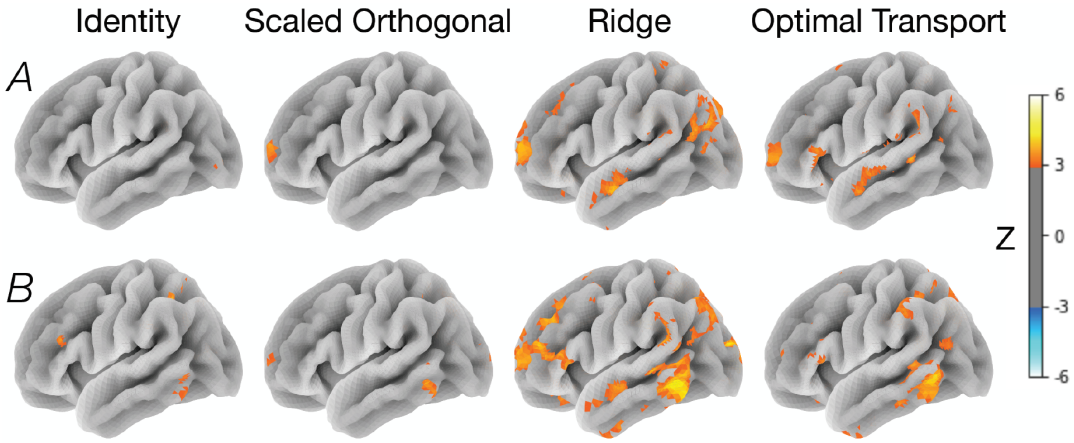

Figure 5.3: Z-scored group maps after alignment : 1 sample test of subjects data brought in common template space. All 13 IBC subjects maps were used to design a template and derive pairwise alignments of subjects toward this representation. Template and pairwise alignments were learned through various alignment methods. *Identity* corresponds to the euclidean mean. Displayed contrasts are (A) *Match (hcp-relational protocol)* (B) 0 *back place(hcp-wm protocol)*

is desirable in a template-building procedure since it better preserves specificity and structure of functional signals.

This difference between both methods is also visible in Figure 5.3. The two sets of maps show the efficiency of alignment methods to bring all subjects in a common space, especially for Optimal Transport and Ridge. The former identifies localized zones of interest across aligned subjects, that suit well previous neuroanatomic findings for the given conditions. On the contrary, Ridge tends to predict strong group effects in numerous large regions that do not seem specific to the studied contrast. This means that Ridge rely on a very smooth common representation where each template voxel activations are a mix of numerous voxels. This smoothing behavior —not preserving spatial specificity of activations in order to maximize overall similarity—can be well-suited for some kinds of applications, such as aligning several datasets in a common space to learn better decoding, for example. However, when aiming at the creation of an interpretable template, Optimal Transport seems to better preserve valuable information on the specificity of activation patterns for a given condition.

We finally consider a realistic use case, where test data would be functionally aligned in view of a group study (one-sample t-test). We present the ensuing group-level brain maps, projected on the cortical surface, in Figure 5.3, for 2 contrasts. The *Match* contrast is derived from the HCP Relational protocol (Barch et al., 2013), focusing on a visual comparison task on artificial textures. For this contrast, we show that both Ridge and optimal transport recover regions in the anterior and posterior segments of the superior temporal sulcus that would not

be detected using standard approaches, or scale-orthogonal functional alignment. The *0-back place* contrast is derived from an acquisition using HCP Working Memory protocol (Barch et al., 2013). Inspired from the visual n-back task, this contrast is a matching task on place images. For these contrasts, OT- and Ridge-based alignment recover regions in the temporo-occipital junction and inferior parietal sulcus that would not be detected using standard approaches, nor by scaled-orthogonal functional alignment. In particular, place sensitive regions of the ventral and dorsal visual cortex seem to have been successfully recovered by the functional alignment approach.

5.3 FURTHER WORK ON TEMPLATE DESIGN

5.3.1 *Better metrics to assess template significance*

The problem of evaluating a template is hard, since it is a problem of creating an abstract object accounting for several datasets. Obviously, there exists no such thing as a perfect template or a ground truth to which we could compare the object that we create. A reasonable idea, explored before, is thus to resort to a well-principled design process, i.e., minimizing a certain objective criterion. However, here we want to focus on giving an interpretable model of group activations and validate how well it accounts for individual models after alignment. In order to do so, we need to define some external validation criteria.

A first solution to try to judge the external validity of such an object is to take some stable functional characteristics, (e.g., those derived using a very large cohort of subjects) and take those as a ground truth, hoping that the large sample size used will yield a reasonable representation. However, this "ground-truth" is also plagued with inter-subject variability and its spatial topography will probably bear some artifactual patterns because of this variability, (e.g., smoothness, functional overlap...)

Another solution is to reflect on interpretable criterions of generalizability of this template to represent unseen subjects. Building on the idea of judging voxels by their functional similarity (e.g., used in Section 4.2), it seems reasonable to compare voxel signals in the sense of the l2-distance. A well-defined template, the empirical mean of training subjects images, can be derived when using this distance at the image level. However, as extensively described in Section 2.3.2, inter-subject variability undermines anatomical correspondences across subjects and l2-distance between anatomically aligned voxels is thus a very biased metric. Instead, we propose to define a comparable metric $\mathcal{L}$ which includes a spatial tolerance for misalignment. For a given voxel i of an unseen subject data $\mathbf{X}$, we would thus measure template

$\mathbf{T}$ dissimilarity to this voxel as the minimal dissimilarity for any voxels contained in the ball of radius r, $\mathcal{B}_i(r)$ around i. In order to make this metric more robust to extreme use-cases where the template is especially designed to overfit this constraint, we add a symmetrical term where each voxel in the template must also be reasonably similar to data. $\mathcal{L}$ is a distance, is equal to the euclidean distance when $r = 0$ and grows more tolerant spatially when r increases. It could also be used to measure the amount of inter-subject spatial variability.

$$\mathcal{L}(\mathbf{X}, \mathbf{T}, r) = \frac{1}{\sqrt{2}}\Big(\sum_{i\in[0,\dots,n]} \min_{j\in\mathcal{B}_i(r)} \|\mathbf{X_i} - \mathbf{T_j}\|^2 + \sum_{i\in[0,\dots,n]} \min_{j\in\mathcal{B}_i(r)} \|\mathbf{T_i} - \mathbf{X_j}\|^2\Big)^{1/2}$$

5.3.2 *Deriving better templates using Wasserstein barycenters.*

The template procedure introduced in Section 3.1.4 relies heavily on a good initialization and is still subject to spatial indeterminacy. As noticed in Section 4.1, Wasserstein Barycenters are a promising avenue to craft geometrically-informed distances to define the template. However, several problems limit the out-of-the-box use of Wasserstein Barycenters to generalize Optimal Transport alignment presented in Section 4.2:

- To properly model pairwise functional alignment as a transport problem, we used a cost matrix depending on both histograms to match. However, this breaks the convexity of the Wasserstein Barycenter formulation, since the cost matrix (and hence the Wasserstein distance) used to compare histograms depends on the template. Put differently, the cost to relate a template and histograms would change at each template iteration, which would hamper convergence.
- Keeping only a function-based cost matrix will always entice a spatial indeterminacy on the template.

To overcome those limits, we tried to use spatial Wasserstein barycenters (in Gramfort, Peyré, and Cuturi, 2015 fashion) to find a better initialization for our iterative algorithm, which would allow starting from a spatially regularized template and refine its functional similarity with subjects. As discussed already, this comes with its own load of practical problems.

NON-POSITIVITY OF ACTIVATIONS As pointed out in Chapter 4, activations obtained in fMRI images are not measures, hence optimal transport cannot be applied directly on the signals in its classical formulation. For one thing, activation histograms contain both a positive and negative part (for deactivations) whereas optimal transport is well-defined only on positive histograms. As deactivations and

activations have different meaning, a sensible heuristic to solve the positivity issue is to separate the positive and negative part of the histograms, and calculate separately transport plans. We proceeded with this heuristic, although it could induce other problems, such as the overlap of activations and deactivations barycenters.

ACTIVATION VALUES ARE NOT NORMALIZED Another hurdle is that histograms are non-normalized, i.e., subjects display different overall quantity of activations (both positive and negative). As already cited, it is possible to use unbalanced optimal transport (Gramfort, Peyré, and Cuturi, 2015) to relax this constraint, as described in Section 4.1. However, the main problem of this approach is the introduction of an additional hyperparameter τ to penalize mass discrepancy against transportation accuracy. This parameter, related to variability, is very hard to set in practice.

We thus opted for a straight forward alternative to restore normalization across histograms, namely *histogram equalization*, sometimes applied in structural MRI preprocessing (Sun et al., 2015). In this step, denoted $\mathrm{Norm}(\mathbf{X})$, we focus on a contrast and treat images as activation histograms (i.e., spatial coordinates of voxels are ignored). After sorting every subjects' histograms, we replace activation value of voxels at a given rank by their average across subjects. Derived normalized histograms respect the individual signal patterns, while activation and deactivation quantity are matched.

GEOMETRICAL BIAS OF SPATIAL BARYCENTERS IN VOLUMIC REPRESENTATION When trying to derive Wasserstein barycenters on volumic images of activation using a cost matrix based on voxels pairwise spatial distance, we quickly observed geometrical artifacts. In the 3D grid geometry implied by the volumic representation, this spatial cost matrix is biased towards the center of the considered region. Voxels in the middle of this region are closer on average to the rest of voxels, especially than those on the edges of the region. By design, transportation to central voxels will thus entice a lower cost and be favored by any algorithm searching to minimize spatial displacement of activations. Implementing this process using an aggregation approach wouldn't solve this problem (e.g., barycenters after parcellation would be biased towards the centers of parcels containing activations).

To overcome this artifactual behavior, it is possible to work instead on surfacic representations of data in each hemisphere. This is known to be a relevant spatial representation of the cortex, where the cortical surface is roughly homeomorphic to a sphere, and well-suited for inter-subject registration (Robinson et al., 2018). Using this representation coupled with its natural distance between vertices (distance on the

mesh), the overall distances better any two vertices from the mesh become way better balanced.

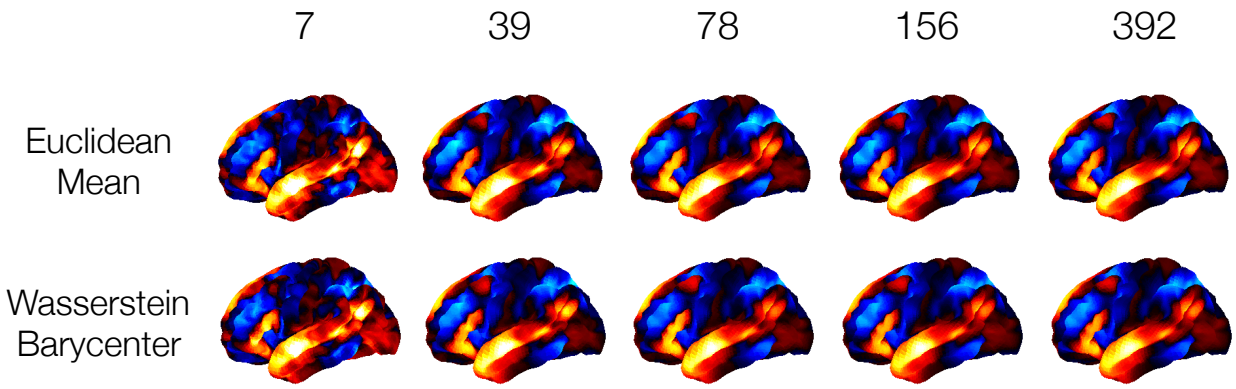

Figure 5.4: Wasserstein barycenter compared to euclidean mean of individual contrasts. Barycenters of "faces-shapes" contrast from HCP Emotional protocol (Barch et al., 2013) for a variable number of HCP subjects (x-axis). Individual contrasts were resampled in *fsaverage5* surface representation (10000 vertices per hemisphere). Balanced Wasserstein barycenter was calculated on the full-hemisphere histograms after histogram normalization, using a spatial cost matrix pairwise linear distance on the mesh. Activation and deactivation barycenters were calculated separately and summed.

EXPERIMENTAL TEST To test this procedure, we used the HCP dataset, focusing on the "faces-shapes" contrast from HCP Emotional protocol (Barch et al., 2013), which is a visual matching task of either faces displaying emotions or geometrical shapes. With the wealth of subjects available in this cohort (more than 700), we took a variable number $n \in \{7, 39, 78, 156, 392\}$ to build a template and kept a reasonable 300-subjects independent validation set. To avoid geometrical artifacts that would be induced by discontinuities introduced through a parcellation, we had to match entire hemisphere histograms in one transport plan. To make computation tractable despite quadratic complexity of optimal transport, we had to resort to use low-resolution surfacic representations (*fsaverage5*, 10000 vertices per hemisphere).

We define $\mathbf{X}$, the concatenation of subjects surfacic activation maps $\mathbf{X_i}$ where $\forall i \in [1, n], \mathbf{X_i} \in \mathbb{R}^p$ with p the number of vertices. We performed histogram normalization (described previously) to derive $\bar{\mathbf{X}} = \mathrm{Norm}(\mathbf{X})$. We separate $\bar{\mathbf{X}}$ in a negative part $\bar{\mathbf{X}}^-$ and a positive part $\bar{\mathbf{X}}^+$ such that $\bar{\mathbf{X}}^+ + \bar{\mathbf{X}}^- = \bar{\mathbf{X}}$. After Norm operation, all columns have equal sum, both positive and negative, so we can normalize them. We denote these sums $s^+ = \sum_{j=1}^{p} \bar{\mathbf{X}}^+_{ij}$ and $s^- = \sum_{j=1}^{p} \bar{\mathbf{X}}^-_{ij}$. Additionally, we define a spatial cost matrix $\mathbf{C} \in \mathbb{R}^{p \times p}$, relating all vertices through their pairwise linear distance on the $\mathrm{fsaverage5}$ mesh used. We then calculate independently the templates $\mathbf{t}^+$ and $\mathbf{t}^- \in \mathbb{R}^p$ as balanced Wasserstein Barycenters that solve Equation

4.3 for cost $\mathbf{C}$ and histogram sets $\frac{\bar{\mathbf{X}}^-}{s^-}$ and $\frac{\bar{\mathbf{X}}^+}{s^+}$. We recompose our final barycenter after rescaling as $\mathbf{t} = s^-\mathbf{t}^- + s^+\mathbf{t}^+$. We assessed the quality of template as a representation of unseen subjects using loss $\mathcal{L}(\mathbf{t}, \mathbf{X}_{\mathtt{left-out}})$, introduced in the previous section.

We implemented balanced Wasserstein barycenters, with an entropic regularization parameter as small as numerically possible, $\epsilon = 0.006$. For implementation, we rely on the efficient, GPU-enabled, Sinkhorn solver for barycenters (Janati, Cuturi, and Gramfort, 2020) from `https://ott-jax.readthedocs.io/en/latest/`

RESULTS Figure 5.4 displays the Euclidean and Wasserstein barycenters obtained using a various number of subjects for the HCP "faces-shapes" contrast. As we can see, the functional topographies are very similar and varying similarly with the growing number of subjects used. We confirmed this by an evaluation of those barycenters compared to left-out subjects using the $\mathcal{L}$ loss. This showed no meaningful difference between both barycenters in their ability to represent well topography of left-out subjects.

This lack of improvement is related to the dimensionality of the barycenter to estimate, which is way higher than the number of samples available to do so. Moreover, the very low-resolution of representations deters the purpose of alignment, which aims at handling local variability in functional topographies. Overall, the main obstacles hampering the geometrical matching of signals are the geometrical effects induced when working on local parcels, in tension with the estimation difficulties on whole-brain signals.

DISCUSSION

In this chapter, we have discussed challenges and methods to derive a functional template, a common functional representation to which individual data can be aligned. Besides being a computational object enabling signal transfer, templates should ideally display a sharp and representative functional topography to enable improved group inference. However, it is hard to control their topographical characteristics, and they are subject to a strong spatial indeterminacy. We introduced a well-principled templating scheme, published in Bazeille et al., 2019, relying on an alternate template minimization and alignment refinement and validated its performance in two experiments. Although this procedure is not sufficient to ensure all desirable properties previously cited, we discussed new metrics to assess those characteristics and potential refinements based on Wasserstein barycenters.

Part III

APPLICATIVE BENCHMARK

6 EMPIRICAL BENCHMARK OF FUNCTIONAL ALIGNMENT

A collection of methods have been proposed for functional alignment (3). However, these methods are seldom used in brain mapping studies. For a given application, it remains unclear how to choose among them, which implementations to use and how much gain can be expected. In this thorough empirical validation, we assess whether alignment can help handle variability in ***typical research set-ups*** *(Bazeille et al., under review).*

As developed in Section 6.1, we consider performance to include both (1) improving inter-subject similarity while retaining individual signal structure and (2) computational efficiency, as the latter is an important consideration for scientists who may not have access to specialized hardware. Section 6.2 develops the ***inter-subject decoding framework*** *we use to measure alignment performance through a downstream predictive task. Using this* ***pairwise*** *framework, we assess five alignment methods on several publicly accessible fMRI datasets. Beyond Procrustes and Optimal transport, we included an original extension of Shared Response Model:* ***Piecewise SRM****, as well as a new Intra-subject alignment method.*

Section 6.3 develop the results of this benchmark. Notably, we: (1) establish that functional alignment improves decoding accuracy above anatomical-only alignment, (2) investigate the impact of common methodological choices such as whether alignment is learned in subregions across the whole brain or in a pre-defined region-of-interest (ROI), and (3) compare the impact of specific alignment methods in whole-brain and ROI-based settings. We then provide a qualitative comparison of the transformations learned by each method to "open the black box" and provide insights into how potential accuracy gains are achieved. Finally, we discuss the availability, usability and scalability of current implementations for each of the methods considered.

Section 6.4 summarizes our findings and put them in context with the literature. Section 6.5 details supplementary analyses assessing the impact of additional methodological choices on alignment performance. It can be skipped at first read, as its main results are summarized in the previous section. For all alignment methods considered here, technically up-to-date and efficient implementations to reproduce these results are provided at `https://github.com/neurodatascience/fmralign-benchmark`.

6.1 QUANTIFYING THE ACCURACY OF ALIGNMENT

6.1.1 *Image-based statistics*

A key question is how to objectively measure the performance of functional alignment. One approach, used in Chapters 4 & 5, is to consider alignment as a reconstruction problem, where we aim to learn a functional alignment transformation which allow imputing missing images in a target subject using data from source subjects. These functionally aligned maps can then be compared with held-out ground-truth maps from the target subject. We can quantify this comparison using image-based statistics such as the correlation of voxel activity profiles across tasks (Guntupalli et al., 2016; Jiahui et al., 2020), spatial correlation or Dice coefficient between estimated and held-out brain maps (Langs et al., 2014) or other metrics such as *reconstruction ratio* introduced in Section 4.4.

However, these image-based statistics are sensitive to low-level image characteristics (e.g., smoothness, scaling), and their values can therefore reflect trivial image processing effects (such as the smoothness introduced by resampling routines) rather than meaningful activity patterns.

6.1.2 *Quantifying alignment accuracy in a predictive framework*

Rather than using image-based statistics, an alternative approach is to test functional alignment accuracy in a predictive framework. Prior works adopting this framework has used tests such as time-segment matching from held-out naturalistic data (e.g., Chen et al., 2015; Guntupalli et al., 2016). However, because time-segment matching relies on the same stimulus class to train and test the alignment, it is unclear whether the learned functional transformations extend to other, unrelated tasks—particularly tasks with low inter-subject correlation (Nastase et al., 2019). We are therefore specifically interested in predictive frameworks that probe model validity by measuring accuracy on held-out data from a different stimulus class, with or without functional alignment.

6.1.3 *Inter-subject decoding*

Inter-subject decoding (2.2.3) is a well-known problem in the literature aimed at uncovering generalizable neural coding principles. Specifically, in inter-subject decoding we learn a predictive model on a set of subjects and then test that model on held-out subjects, measuring the extent to which learned representations generalize across individuals. In an information-mapping framework (Kriegeskorte and Diedrichsen, 2019), decoding allows us to assess the mutual

information between task conditions. Functional alignment, therefore, should facilitate information-mapping by increasing the similarity of condition-specific representations across subjects, thus improving their decoding.

Although the link between mutual information and decoding accuracy is non-trivial (Olivetti, Veeramachaneni, and Avesani, 2011), we consider that measuring alignment with decoding accuracy on unseen subjects better fulfills neuroscientists' expectations of inter-subject alignment in two main ways. First, decoding accuracy provides a more interpretable assessment of performance than other measures such as mutual information estimates. Second, decoding accuracy on a held-out sample provides insight into the external validity and therefore generalizability of derived neural coding principles. Compared to image-based measures, decoding accuracy is thus a more rigorous measure of whether functional alignment improves the similarity of brain signals across subjects while also preserving their structure and usability for broader research use cases. In this work, we therefore quantify functional alignment accuracy by assessing improvements in inter-subject decoding when using functional alignment over and above anatomical alignment. That is, the field-standard approach of normalizing subjects to a standardized anatomical template using diffeomorphic registrations, as implemented in, e.g., fMRIprep (Esteban et al., 2019).

6.2 INTER-SUBJECT DECODING BENCHMARK

6.2.1 *Experimental procedure*

For each dataset considered (6.2.5), we calculated the inter-subject decoding accuracy for anatomical-only alignment and for each of the five considered functional alignment methods. To calculate inter-subject decoding accuracy, we took the trial- or condition-specific beta maps generated for each dataset (see Section 6.2.5 for full details on beta-map generation) and fit a linear SVM (2.2.3). In order to ensure fair comparisons of decoding accuracy across experiments, we chose a classifier with no feature selection and default model regularization ($C = 1.0$). Classifiers were implemented in Scikit-learn (Pedregosa et al., 2011), and decoding accuracy was assessed using a leave-one-subject-out cross-validation scheme. That is, the linear SVM was trained to classify condition labels on all-but-one subject and the resulting trained classifier was used without retraining on the held-out subject, providing an accuracy score for that cross-validation fold.

For each dataset, we first calculated the inter-subject decoding accuracy using anatomical alignment. This served as a baseline accuracy against which we could compare each functional alignment method.

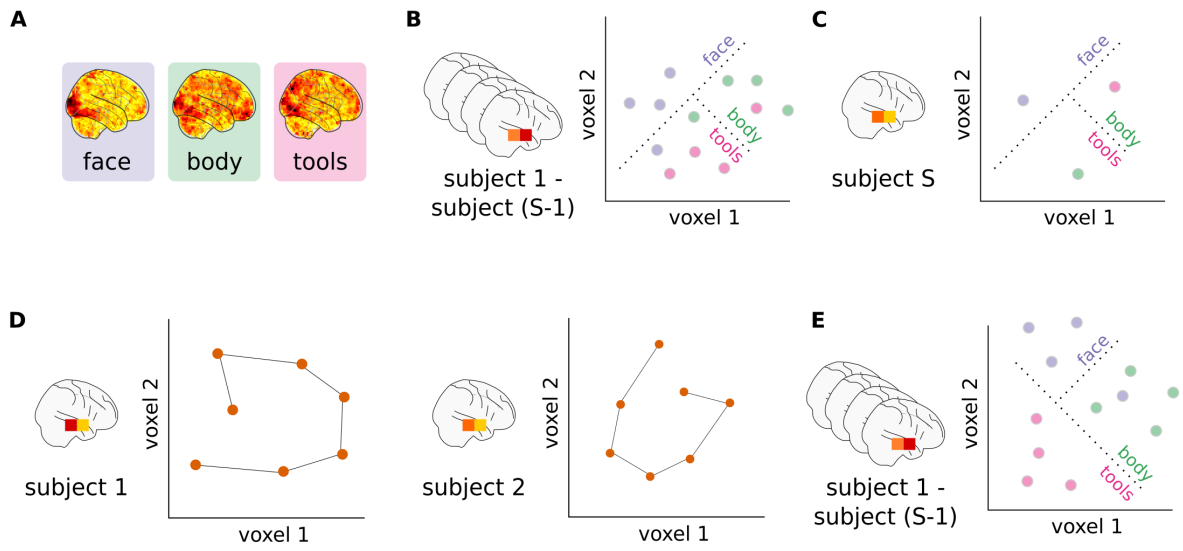

Figure 6.1: Analysis pipeline. (A) First-level general linear models are fit for each subject to derive trial- or condition-specific beta-maps for each session. (B) These beta maps and their matching condition labels are used to train a linear SVM on the training set of subjects. (C) The trained classifier is applied to a held-out test subject, and accuracy is assessed by comparing the predicted and actual condition labels. (D) On a separate task, we compare subject-level activation patterns as trajectories in the high-dimensional voxel space. This allows us to learn functional alignment transformations that maximize the similarity of these high-dimensional spaces. (E) These voxel-wise transformations are applied on the decoding beta maps, and a new linear SVM is trained to predict condition labels. This trained classifier can then be applied to the held-out test subject and decoding accuracy assessed as in (C).

Using alignment data, functional alignment transformations were then learned for each pairwise method, where the left-out subject for that cross-validation fold was the target subject for functional alignment. Inter-subject decoding accuracy was then re-calculated after applying functional alignment transformations to the decoding beta maps.

In the special case of SRM—which calculates an alignment from all provided subjects in a single decomposition—we withheld the left-out subject from the shared response estimation step to avoid data leakage. The projection of the left-out subject is learned from previously estimated shared space. Then, the learned projections are applied to the decoding data and decoding is performed on the projected data.

For each cross-validation fold, we report the inter-subject decoding accuracy of a given functional alignment method after subtracting the baseline, anatomical-only accuracy for that same fold. An overview of the experimental procedures is provided in Figure 6.1.

6.2.2 *Benchmarked methods*

AGGREGATION SCHEMES USED IN THIS BENCHMARK To align the entire cortex across subjects, two main frameworks have been proposed: searchlight (3.1.3) and piecewise (4.3). Each of these frameworks

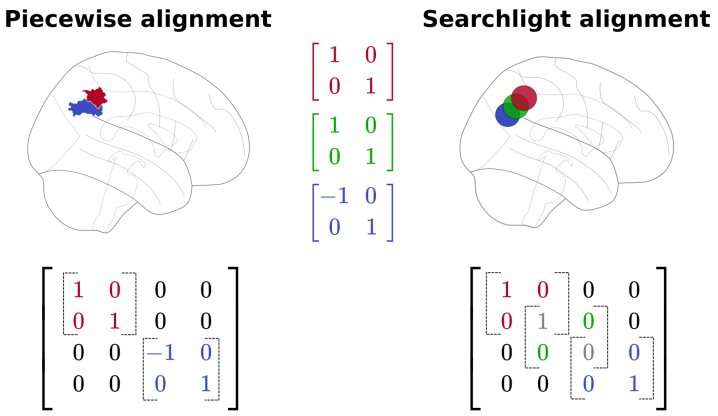

Figure 6.2: Comparing piecewise and searchlight alignment. In this illustration, transformations are derived for the blue, green, and red areas separately. Note that the piecewise alignment does not include a green area, as this corresponds to a searchlight overlapping both the red and blue areas. For non-overlapping parcels, these transformations are stacked into a larger orthogonal matrix. For the overlapping searchlight, these transformations are aggregated, with overlapping values averaged. Note that the final transformation for the searchlight alignment is no longer orthogonal in this example.

use functional alignment methods to learn local transformations and aggregate them into a single large-scale alignment; however, searchlight aggregates overlapping transforms where piecewise acts on segregated transforms, as illustrated in Figure 6.2.

In the literature to date, searchlight and piecewise aggregation schemes have both been used in conjunction with Generalized Procrustes Analysis, under the names Searchlight hyperalignment (3.1.3, Guntupalli et al., 2016) and scaled orthogonal alignment (4.4, Bazeille et al., 2019), respectively. We therefore include both searchlight Procrustes and piecewise Procrustes in our benchmark. Every other method is regularized at the whole-brain level of analysis through piecewise alignment.

As piecewise alignment is learned within a parcellation, an important question is: which brain atlas should be used for piecewise alignment? In Section 6.5.3, we compare results from the Schaefer et al., 2018 atlases to those from parcellations derived directly on the alignment data. By default, the results presented below are derived with the 300 ROI parcellation of the Schaefer atlas, unless noted otherwise. In the case of searchlight Procrustes, we selected searchlight parameters to match those used in Guntupalli et al., 2016 as implemented in PyMVPA (Hanke et al., 2009).

METHODS INCLUDED IN THIS BENCHMARK In this study, we specifically focus on pairwise alignments wherein subjects are directly aligned to a target subject's functional activations. Although template-based approaches are an important area of research (as developed in Chapter 5), the question of how best to generate the reference template

is distinct from determining the best alignment method and beyond the scope of the current work.

As we use inter-subject decoding to compare functional alignment methods, we only consider methods that meet the following two criteria. First, the alignment transformations should be learned on activations evoked during temporally synchronized (i.e., co-occuring) task data, or on contrasts matched across individuals. Second, the learned transformations must be invertible or quasi-invertible linear mappings and applicable as-is on unseen data with a different task structure. These two criteria exclude several methods introduced in Section 3.2, such as regularized canonical correlation analysis (rCCA; Bilenko and Gallant, 2016), gradient hyperalignment (Xu, Yousefnezhad, and Zhang, 2018a), connectivity hyperalignment (Guntupalli, Feilong, and Haxby, 2018), and methods based on Laplacian embeddings (Langs et al., 2014).

In our whole-brain benchmark, we consider five different alignment methods: searchlight Procrustes (3.1.2; Guntupalli et al., 2016; Haxby et al., 2011), piecewise Procrustes, piecewise Optimal Transport (4.2; Bazeille et al., 2019), piecewise Shared Response Modelling (SRM; 3.1.5; Chen et al., 2015), and intra-subject correlations across tasks (Tavor et al., 2016), here referred to as "intra-subject alignment" (3.2.3).

IMPLEMENTATION DETAILS For Optimal Transport, the level of sparsity is controlled by ϵ, a user-supplied hyper-parameter, which we set to 0.1 throughout our experiments. For our implementation, we rely on the `fmralign` package. Optimal transport transformations are calculated in a piecewise fashion, following Bazeille et al., 2019.

For Intra-subject alignment (3.2.3) we observed in preliminaries observations that—unlike other piecewise techniques (see Section 6.5.3)—the decoding accuracy strictly improved with the number of parcels used, so we used the highest resolution atlas available. Thus, divide alignment and decoding data into 1000 parcels using the highest-resolution Schaefer atlas (Schaefer et al., 2018) and train models to predict decoding data from alignment data on each parcel. After repeating this procedure for all *source* subjects, we then use $\mathbf{R}^{intra}$ to estimate decoding data for $target$ subject as $\hat{\mathbf{D}}^{target} = \mathbf{A}^{target}\mathbf{R}^{intra}$. As with other functional alignment methods, we can evaluate the quality of our estimation using an inter-subject decoding framework

PIECEWISE SHARED RESPONSE MODEL For Shared Response modelling, we specifically use the FastSRM implementation proposed by Richard et al., 2019 and available in the BrainIAK library (RRID: SCR_01 4824), that approximates this calculation with an emphasis on

improved computational performance. For full details on the computational advantages of FastSRM, we direct the reader to their work.

In order to align our SRM implementation with the other considered alignment algorithms, we introduce a new piecewise SRM method to aggregate SRM transformations across the whole brain. Thus, within each parcel or across an *a priori* ROI, SRM decomposes the signal of many subjects in a common basis, with the same orthogonality constraint as Procrustes.Given the strong dependency of SRM performance on the selected hyper-parameter k, this parameter requires additional experimenter consideration. For piecewise SRM, we perform a grid search to select the relevant Schaefer parcellation resolution and number of components k (see Section 6.5.4). From these results, we chose to use Schaefer atlas 700 and run one SRM on each parcel searching for 50 components—or equal to the number of voxels if less than 50 voxels are in a given parcel. For ROI-based analyses, we set k to 50 components as in our piecewise analyses and matching the original SRM benchmarks provided in Chen et al., 2015.

6.2.3 *Main experiments*

Experiment 1 uses the experimental procedure described previously to assess accuracy gains provided by alignment methods with respect to anatomical alignment when applied to whole-brain images. We benchmarked the five methods described in Section 6.2.2: piecewise Procrustes, searchlight Procrustes, piecewise Optimal Transport, piecewise SRM, and intra-subject alignment, with relevant hyperparameters selected as described previously. Results of this benchmark, comprising five tasks from four datasets, are presented in Section 6.3.1. For each method, we also assessed its computation time relative to piecewise Procrustes alignment. Piecewise Procrustes provides a reasonable computational baseline as it is the only considered alignment method that does not include a hyperparameter and therefore shows a stable computation time across experiments.

We estimate the noise ceiling for this task as within-subject decoding accuracy. Within-subject decoding was calculated separately for each subject as the average leave-one-session-out decoding accuracy. We can then directly compare this accuracy value to the inter-subject decoding accuracy when that subject is the target—that is, the left-out—subject. The difference between within- and anatomical inter-subject decoding accuracies, then, is a good approximation of the decoding accuracy lost due to inter-subject variability; therefore, it provides a range of possible accuracy gains that can be expected from functional realignment.

We then conducted *Experiment 2* to understand how whole-brain results compare to ROI-based analyses. Specifically, we replicated *Experiment 1* within selected ROIs, such that local alignment methods were applied directly without any aggregation scheme. ROIs were

chosen based on *a priori* expectations of each decoding task (see Section 6.2.5 for details for each dataset). Results from *Experiment 2* are shown in Section 6.3.2.

Experiment 3 tackles the notoriously hard problem of understanding how each of the considered methods align subjects by examining qualitatively their impact on activity patterns across individuals. To "open the black-box," we reused IBC dataset full-brain alignments learned in *Experiment 1*. Specifically, we consider the transformation to sub-04's activity pattern from all other subjects' functional data. With these transformations, we align two contrasts from the IBC dataset: Rapid Serial Visual Presentation of words (RSVP language task) and sound listening. Finally, we run a group conjunction analysis (Heller et al., 2007) on these aligned contrasts and visualize the results. This statistical analysis, more sensitive than its random effect equivalent on small samples, allows one to infer that every subject activated in the region with a proportion γ showing the effect considered. Here we use $\gamma = 0.25$ to recover all regions selectively activated by at least a few subjects, and we show in Section 6.3.3 how this group functional topography is modified by alignment.

6.2.4 *Control analyses*

In addition to our three main experiments, we ran three additional control analyses on the IBC dataset. First, we assess the impact of the brain parcellation and its resolution on piecewise alignment by comparing whole-brain decoding accuracy for IBC tasks using piecewise Procrustes across both data-driven and pre-defined parcellations (Section 6.5.3). As piecewise SRM displays an interaction between parcellation resolution and SRM's hyperparameter k, we ran a grid search for this algorithm to determine its optimal parameters (Section 6.5.4).

Second, we calculated inter-subject decoding performance after applying Gaussian smoothing kernels of several widths on both IBC dataset decoding tasks (Section 6.5.5). Gaussian smoothing is an interesting baseline, as it is commonly used to facilitate inter-subject comparisons by smoothing over residual variance in functional mappings. In a third control experiment, we assessed the impact of whether data is represented on the surface or the volume and resolution on decoding accuracy in the IBC Rapid-Serial-Visual-Presentation (RSVP) language task (Section 6.5.6).

6.2.5 *Datasets and preprocessing*

In order to assess the performance of each functional alignment method in a range of applications, we searched for publicly accessible datasets that included both a task suitable to learn the alignment

Dataset	S	Alignment data	p	Decoding task	Decoding categories	d
Individual Brain Charting	10	Contrast maps from HCP and ARCHI task batteries	53	RSVP Language	Words, Non-Words, Consonants, Sentences, Jabberwocky	360
				Sounds dataset	Voice, Nature, Animal, Music, Speech, Tools	72
BOLD5000	4	COCO, ImageNet, and Scenes images	300	Imagenet images content	Plant, Animal, Food, Artifact	350
Forrest	10	*Forrest Gump* audio-movie listening	1600	Music genre	Country, Metal, Ambient, Symphonic, Rock	200
Courtois Neuromod	6	*Life* movie watching	2008	Visual *n*-back condition	Body 0-back, Body 2-back, Face 0-back, Face 2-back, Place 0-back, Place 2-back, Tools 0-back, Tools 2-back	72

Table 6.1: Datasets used to benchmark alignment methods. The four datasets used in this benchmark, where each dataset consists of S subjects. We note the alignment data used for each dataset and p the number of timeframes it comprises. These datasets show the range of possible task structures which work for alignment—from static images for BOLD5000, to statistical contrast maps for IBC, to complex audio or audio-visual movies for Forrest and Courtois Neuromod. A full listing of included 53 contrast maps for IBC is included in Section 6.5.7. We also include the decoding task(s) used for each dataset. Each subject's decoding task data comprises d images evenly divided across the listed stimulus categories (except for BOLD5000 categories that are unbalanced). Of note, IBC dataset has two independent decoding tasks, bringing the total number of decoding tasks to five.

(e.g., naturalistic or localizer protocols) and an independent decoding task on which we could evaluate functional alignment performance. After discarding datasets where we could not obtain above-chance accuracy levels for within-subject decoding, we retained four datasets: BOLD5000 (Chang et al., 2019), Courtois NeuroMod (Boyle et al., 2020), Individual Brain Charting (IBC; Pinho et al., 2018), and Study Forrest (Hanke et al., 2016). For the IBC dataset, we included both a language (RSVP language) and auditory (Sounds dataset) decoding task, yielding a total of five decoding tasks that probe visual, auditory, and language systems. For a complete description of the alignment and decoding data included in each experiment, please see Table 6.1.

BOLD5000, StudyForrest and Courtois NeuroMod were preprocessed with fMRIprep (Esteban et al., 2019), while IBC data were preprocessed using an SPM-based pipeline as described in Pinho et al., 2018. A complete description of the fMRIprep preprocessing procedures is available in the appendix (Section 6.5.8). Preprocessed data were then masked using a gray matter mask, detrended, and standardized using Nilearn (Abraham et al., 2014a). To reduce the computational cost of functional alignment, we downsampled all in-

cluded datasets to 3mm resolution. Both alignment and decoding task data were then additionally smoothed with a 5mm Gaussian kernel. A general linear model (GLM) was fit to each decoding task run to derive trial-specific beta maps (or condition-specific beta maps for the Courtois Neuromod and IBC Sounds tasks), which were carried forward for inter-subject decoding.

As described in Section 6.2.3, *Experiment 2* uses pre-defined regions of interest (ROIs). We selected large, task-relevant ROIs to ensure that sufficient signal was available when decoding. A large visual region, extracted from the Yeo7 (Buckner et al., 2011) atlas, was used for the visual tasks in BOLD5000 and Courtois NeuroMod. For Forrest and IBC Sounds—which are auditory tasks—we took the Neuroquery (Dockès et al., 2020) predicted response to the term "auditory". We then compared this predicted response with the BASC (Bootstrap Analysis of Stable Clusters) atlas (at scale 36; Bellec et al., 2010) and took the parcel most overlapping with the predicted response; namely, parcel 25. For IBC RSVP, which is a reading task, we extracted the BASC (at scale 20) atlas components most overlapping with MSDL (Multi-Subject Dictionary Learning; Varoquaux et al., 2011) atlas parcels labeled as left superior temporal sulcus, Broca and left temporo-parietal junction: namely, the 8 and 18 BASC components. We then kept only the largest connected component. All included ROIs are displayed in Figure 6.5.

6.2.6 *Software used and released*

With the exception of Courtois Neuromod, all other included datasets are available on OpenNeuro (Poldrack et al., 2013b) under the following identifiers: *ds000113* (Study Forrest), *ds001499* (BOLD5000), and *ds002685* (IBC). Courtois Neuromod `2020-alpha2` release will be available under a data usage agreement as outlined on `https://docs.cneuromod.ca`.

Our pipeline entirely relies on open-source Python software, particularly the SciPy stack (Virtanen et al., 2020). All included methods are implemented in `fmralign` or accessed through their original, open source implementations, as described in Section 6.2.2. To ease replication and extension of the presented results, we have created the `fmralign-benchmark` repository under `https://github.com/neurodatascience/fmralign-benchmark`. This repository provides an implementation of the procedures adopted in these experiments, building on `fmralign` and previously cited tools.

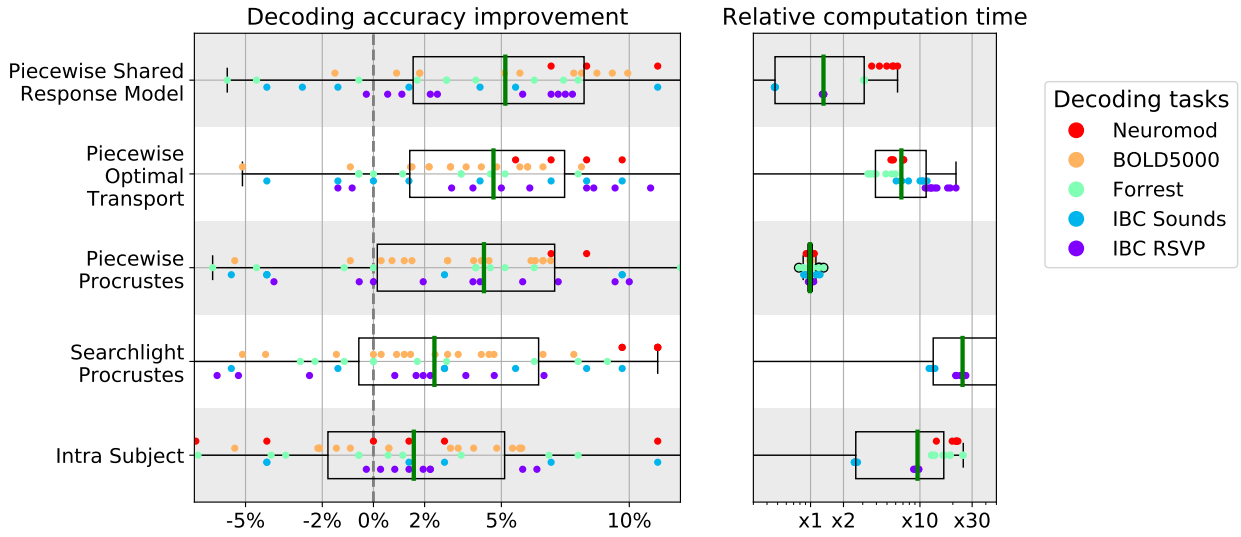

Figure 6.3: Decoding accuracy improvement and computation time after whole-brain functional alignment. In the *left panel*, we show decoding accuracy improvement for each of the considered functional alignment methods at the whole-brain level of analysis. Each dot represents a single subject, and subjects are colored according to their decoding task. To aggregate results across datasets, we show accuracy scores after subtracting inter-subject decoding accuracy for the same leave-one-subject-out cross-validation fold with anatomical-only alignment. In the *right panel*, we show the computational time for each of the considered methods. All computation times are depicted as relative to piecewise Procrustes. For both panels, each box plot describes the distribution of values across datasets, where the green line indicates the median. We see that piecewise Procrustes, Optimal Transport, and intra-subject alignment consistently improve decoding accuracy across datasets. We also see that piecewise Optimal Transport is 10 times slower, and searchlight Procrustes is more than 30 times slower than piecewise Procrustes.

6.3 EMPIRICAL VALIDATION OF ALIGNMENT

6.3.1 *Functional alignment improves inter-subject decoding*

The *left panel* of Figure 6.3 displays absolute decoding accuracy change brought by each functional alignment method relative to anatomical alignment on whole-brain images. As every method is trained and tested on the same cross-validation folds, we report the fold-by-fold performance change. The *right panel* displays each method's relative computation time compared to piecewise Procrustes alignment. For each panel, each point displayed is the result for one leave-one-subject-out cross validation fold, and each color corresponds to one of the five decoding tasks. Note that these timings are based on available implementations — `fmralign` for piecewise alignment methods, `pymvpa2` for searchlight, and `BrainIAK` for SRM— and are therefore subject to change as implementations improve. Nonetheless, these estimates provide insight into the current state-of-the-art.

ALIGNMENT SUBSTANTIALLY IMPROVES INTER-SUBJECT DECODING ACCURACY Overall, we can conclude that most functional alignment

methods consistently improve decoding accuracy, with gains from 2-5% over baseline. This trend is relatively consistent across datasets and target subjects. Thus, alignment methods manage to reliably reduce individual signal variability while preserving task-relevant information in a variety of conditions. Although we see that there is noticeable variance in performance across data sets, these methods on average show significant effects on inter-subject decoding accuracies. As reported in Table 6.2, baseline accuracy is around 20% above chance on average. In this setting, the observed 5% average improvement across datasets is a substantial increase in performance.

In order to provide further context on these results, we also estimated the noise ceiling for inter-subject decoding. Figure 6.4 reports that across datasets, the leave-one-session-out (i.e., within-subject) decoding accuracy for the target subject is on average 8.5% higher than the corresponding leave-one-subject-out (i.e., inter-subject) decoding accuracy after anatomical alignment for the same target subject. Thus, we expect that functional alignment methods will achieve at most an 8.5% increase in inter-subject decoding accuracy over anatomical alignment. In this light, we can see that the best functional alignment method recovers more than half of the decoding accuracy lost to inter-subject variability. Additional control analyses suggest that this effect cannot be explained by smoothing (Section 6.5.5). We further find that the presented results are largely insensitive both to whether the data is represented on the cortical surface or in volumetric space, and to the parcellation resolution used (see Section 6.5.6).

PIECEWISE METHODS SHOW COMPUTATIONAL AND ACCURACY ADVANTAGES Procrustes alignment results in better inter-subject decoding accuracies when performed in a piecewise as compared to a searchlight approach. Specifically, searchlight shows lower decoding accuracies on average, suggesting that its internal averaging destroys part of the signal structure recovered by Procrustes. With respect to computational cost, we can see that searchlight Procrustes is 25 times slower on average than piecewise Procrustes. These results suggest that piecewise alignment is a better choice when calculating functional alignment transformations on full-brain data. Moreover, Section 6.5.3 shows that gains to expect from piecewise alignment are largely insensitive to the resolution and type of parcellation used; i.e., taken from an atlas or learned directly from subject data.

The two best performing methods also use a piecewise aggregation scheme. Specifically, Piecewise SRM and Optimal Transport yield the highest decoding scores with a slightly lower standard deviation in accuracy scores than Procrustes. Piecewise SRM is faster to train than piecewise Procrustes for a fixed set of hyperparameters; however, identifying the ideal hyperparameters for a new dataset requires a computationally costly grid-search. Our results (see Section 6.5.4)

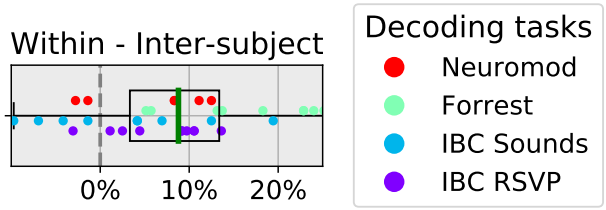

Figure 6.4: Within-subject minus inter-subject decoding accuracy. We show the difference between the average leave-one-session-out within-subject decoding accuracy and anatomically-aligned leave-one-subject-out inter-subject decoding accuracy, when that target subject is left-out. Thus, each dot corresponds to a single subject, and the dot's color indicates the decoding task. Of note, BOLD5000 was dropped as it did not have independent folds, and therefore could not be used for within-subject cross-validation. The box plot describes the distribution of differences, where the green line represents the median value. We argue this difference approximates the effects of inter-individual variability, and so the best average accuracy improvement we can hope for using functional alignment is around 9%.

suggest that, in general, a large number of components k and a high-resolution parcellation are likely to give reasonable performance across datasets.

The second best performing method, Optimal Transport, gives non-trivial accuracy gains in most configurations and only rarely decreases decoding accuracy, likely because of the stronger constraints that it imposes. However, this extra-performance comes at a computational cost: it is on average 7 times slower than Procrustes. For data sets without sufficient data or computational power to perform a hyper-parameter grid search for piecewise SRM, we suggest that Optimal Transport offers robust decoding performance with little hyper-parameter tuning. It remains, however, more computationally costly than the reference implementation of piecewise Procrustes.

TASK-SPECIFIC MAPPINGS CAN BE LEARNED WITHIN SUBJECTS

The intra-subject alignment approach differs from other considered functional alignment methods in that it learns mappings between the alignment data and decoding task data, with the assumption that these mappings can be generalized across subjects. Our results support this assumption, although this method yields gains half as large as the best performing alignment method and comes with a significant computational cost. Part of this cost can be accounted for by the increase in the number of parcels that are used to preserve signal specificity. Nonetheless, using task-specific mappings as a functional alignment method suggests that future work on refining related methods may be a promising direction of research.

6.3.2 *Whole-brain alignment outperforms ROI-based alignment*

The *left panel* of Figure 6.5 displays the performance of each functional alignment method relative to anatomical alignment within

task-relevant ROIs. The *right panel* displays each method's relative computation time compared to piecewise Procrustes alignment. When visually compared to Figure 6.3, ROI-based decoding accuracies appear to be slightly lower than whole-brain decoding accuracies for most of the considered methods. We directly compare ROI-based and whole-brain alignment in a supplementary analysis, depicted in Figure 6.7, confirming that ROI-based decoding accuracies are in fact lower on average for the datasets considered in this work. Our results support previous work from the inter-subject decoding literature (Chang et al., 2015; Schrouff et al., 2018a) and suggest that full-brain piecewise alignment yields the best overall decoding pipeline, though we note that this conclusion may change depending on the exact research context.

OPTIMAL TRANSPORT AND SRM SHOW HIGH ROI PERFORMANCE

Overall, we find that the best performing methods bring a 3-5% improvement in decoding accuracy at the ROI level of analysis. Specifically, Optimal Transport is on average the best performing method, with a median accuracy increase of 5% within task-relevant ROIs. Here, we see that baseline decoding accuracy is less than 10 % above chance in all datasets (except for Courtois Neuromod; see Table 6.3 for exact accuracy values). Thus, the 5% accuracy increase brought by Optimal Transport represents a strong effect.

SRM yields the second-best performance within ROIs, showing reasonable decoding accuracy gains on most datasets. It shows more variance across datasets, however, than the other considered methods. In particular, SRM decreases inter-subject decoding accuracy on the visual ROI for Courtois Neuromod, with accuracy values dropping by approximately -20% compared to anatomical alignment (see Table 6.3). Performance was not significantly improved by using a higher number (up to 600) of components, highlighting the unique difficulty in identifying well-suited hyper-parameters for SRM. Interestingly, Procrustes shows substantially lower performance on average in the ROI compared to the whole-brain level of analysis, largely due to its sensitivity to high-variance voxels.

Computationally, we see that SRM is the fastest method and runs roughly 3 times faster than Procrustes, while Optimal Transport remains 10 times slower than Procrustes. We also note that—on average—intra-subject alignment does not show increased inter-subject decoding accuracy within task-relevant ROIs. We suspect that this is likely because when restricting the learned relationship between data types (e.g., movie-watching to classification task data) to a single ROI, the low number of predicted features precludes the identification of stable multivariate patterns that can transfer across subjects.

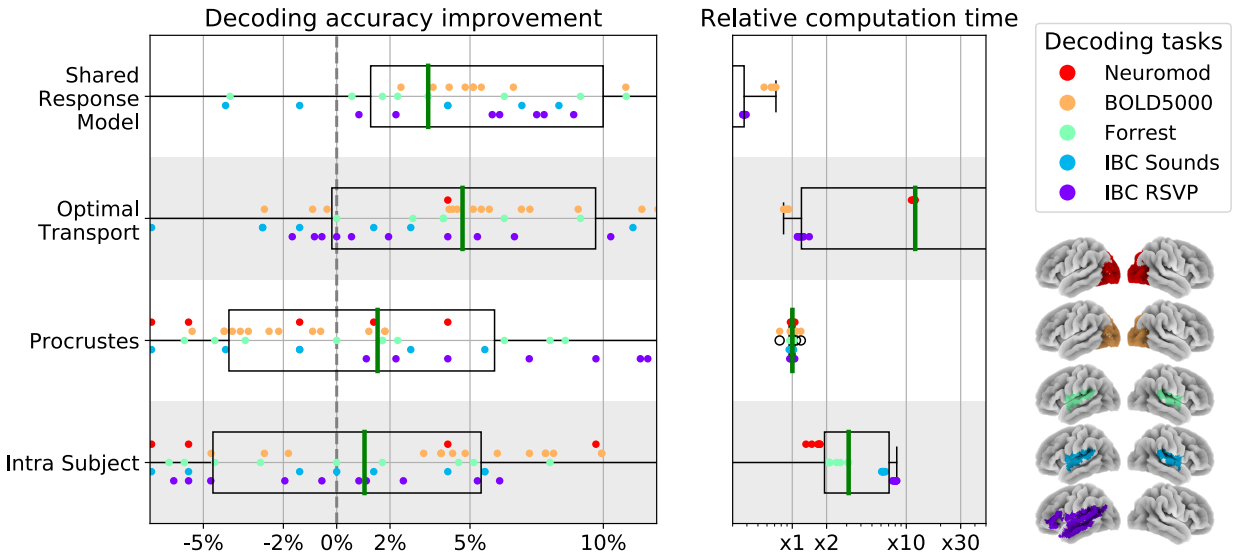

Figure 6.5: Decoding accuracy improvement and computation time after ROI-based functional alignment. In the *left panel*, we show decoding accuracy for each of the considered local functional alignment methods at the ROI level of analysis. The ROIs used for each dataset are displayed on the *far right*. Each dot represents a single subject, and subjects are colored according to their decoding task. Rather than raw values, we show accuracy scores after subtracting inter-subject decoding accuracy for the same leave-one-subject-out cross-validation fold with anatomical-only alignment. Note that all methods are applied without aggregation, so only the method name is given. In the *right panel*, we show the computational time for each of the considered methods. All computation times are depicted as relative to piecewise Procrustes. For both panels, each box plot describes the distribution of values, where the green line indicates the median.

6.3.3 *Qualitative display of learned alignments*

Understanding the effects of high-dimensional transformations—such as those used in functional alignment—is non-trivial. To aid in this process, we "open the black box" by functionally aligning a group of subjects to an individual target subject's functional space and depict the resulting maps in Figure 6.6. Here, we reuse whole-brain alignments learned in *Experiment 1*. We also display the ground-truth individual activation maps in *panel A*, in order to better highlight how each method affects the signal distribution. As a reminder, the contrast data displayed here was not used to learn alignments, so it means that alignment learned on various task data, not specifically related to language nor audition, carried enough information for fine-grain registration of these networks.

We can see that overall, functional alignment methods enhance group-level contrasts compared to anatomical-only alignment; i.e., activation maps are more similar across functionally-aligned subjects. This result is not at the expense of signal specificity, since the aligned group topographies are still sharp. From the comparison between panels *A* and *B*, we can also conclude that alignment methods bring group topography much closer to the targeted subject topography across many contrasts. Nonetheless, we can still observe that there seems

to be a trade-off between sharpness of activation (low smoothness of image, due to low variance across aligned subjects) with Optimal Transport, and accuracy of their location compared to the target ones (low bias introduced by the matching) with searchlight Procrustes.

6.4 DISCUSSION

In this work, we have proposed a new procedure to measure the information recovered through functional alignment using inter-subject decoding, and we subsequently used this framework to benchmark five functional alignment methods on five distinct decoding tasks across four publicly available datasets. In general, we find that functional alignment improves inter-subject decoding accuracy in both whole-brain and ROI settings. These results, combined with our qualitative visualization of the effects of functional alignment on signal structure, suggest that functional alignment improves inter-subject correspondence while matching signal to realistic functional topographies. This finding extends and supports conclusions from earlier work (Güçlü and Gerven, 2015; Guntupalli et al., 2016).

At a whole-brain scale, the best performing methods are piecewise SRM, piecewise Optimal Transport, and piecewise Procrustes which each bring 5% improvement over this baseline on average. As the baseline inter-subject decoding accuracy is roughly 20% above chance across datasets (Table 6.2), this 5% increase represents a substantial improvement. We also note that this represents recovering more than half of the accuracy lost to inter-subject variability. The considered functional alignment methods also improve decoding performance when applied *without* an aggregation scheme (i.e., piecewise or searchlight aggregation) within task-relevant ROIs. Here, Optimal Transport and SRM bring 5% and 3% improvement in inter-subject decoding accuracy, respectively, over a baseline accuracy which is on average 10-15% above chance across datasets (Table 6.3). From our control analyses, we observe that these increases in decoding accuracy were reliably greater than the effect of Gaussian smoothing (see section 6.5.5). In a minimalistic replication, this effect seems to hold for both volumetric and surface data and at different parcellation resolutions (see section 6.5.6; cf. Oosterhof et al., 2011).

Our benchmark also brings new evidence that the latent correspondences that can be learned between different tasks display less inter-individual variability than the task-specific activation maps (Tavor et al., 2016). *Experiment 1* indeed showed that such correspondences could even be used at a whole-brain scale to transfer signals across subjects to solve an inter-subject decoding problem, which is—to the best of our knowledge—an original experimental result. By releasing efficient and accessible implementations of these methods in the

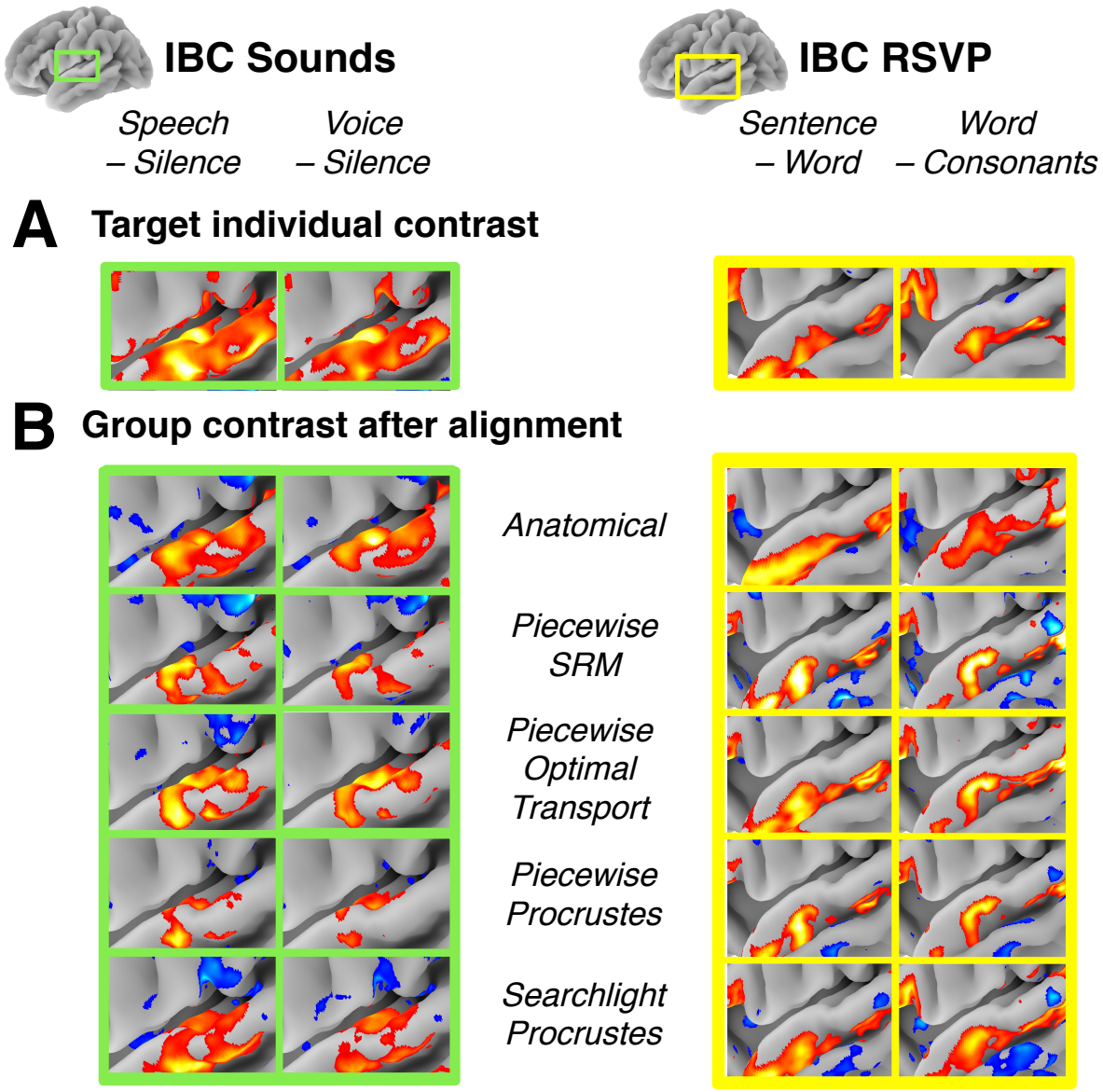

Figure 6.6: Comparison of alignment methods' geometrical effects. (A) Activation patterns for the Target subject (IBC sub-04) for two contrasts from the IBC Sounds task (*Speech* > *Silence*, *Voice* > *Silence*) and IBC RSVP task (*Sentence* > *Word*, *Word* > *Consonants*). Here, we only show contrast maps from a sub-region of the temporal lobe containing contrast-relevant information. Note that this sub-region differs slightly between the Sounds and RSVP task. (B) Visualization of a group conjunction analysis of all IBC subjects after alignment to the target subject for each of the considered methods. We used a γ value of 0.25 in the conjunction analysis, which corresponds to at least 25% of the IBC sample showing activation in this temporal region after alignment. For ease of comparison, the color bar for each contrast and method was scaled to show the full range of values (i.e., the color bar spans different intervals across methods and contrasts) and so is not included here. All displayed maps were thresholded at 1/3 of their maximum value. We see that functional alignment yields stronger contrasts overall when compared to anatomical alignment. Piecewise Procrustes and piecewise Optimal Transport yield less smooth representations, better preserving signal specificity.

`fmralign` package, we hope to facilitate future cognitive neuroscience research using functional alignment methods.

6.4.1 *Combining local alignment models*

Across datasets, we find that the spatial framework of alignment and decoding significantly affects subsequent performance. Notably, piecewise Procrustes outperforms searchlight Procrustes, both in terms of accuracy and computational performance. The methodological difference between these aggregation schemes is whether alignment transformations are learned within overlapping neighborhoods (as in searchlight Procrustes) or not (as in piecewise Procrustes). Searchlight alignment suffers in that the overlap between searchlights requires multiple computations for a given neighborhood, and the aggregated transformation is no longer guaranteed to reflect properties of the original transforms, e.g., orthogonality. Although piecewise alignment may theoretically introduce discontinuities at parcel boundaries, in our results we do not find evidence of this effect and indeed find that piecewise aggregation overall benefits decoding performance. Importantly, we found that the improved performance of piecewise Procrustes was largely insensitive to parcel size and definition (see Figure 6.8).

6.4.2 *Evaluating alignment performance with decoding*

We use inter-subject decoding to quantify the amount of mutual information recovered by functional alignment methods. In general, identifying publicly available datasets with tasks appropriate for both inter-subject decoding and functional alignment remains a challenge. Beyond the four datasets included in these results, we investigated several other publicly available datasets such as the Neuroimaging Analysis Replication and Prediction Study (NARPS; Botvinik-Nezer et al., 2020),the Healthy Brain Network Serial Scanning Initiative (HBN-SSI; O'Connor et al., 2017), the interTVA dataset (Aglieri et al., 2019, available as OpenNeuro *ds001771*) and the Dual Mechanisms of Cognitive Control Project (DMCC, Braver et al., 2020).

We had difficulties in achieving sufficient baseline accuracy levels in these and other datasets, and we therefore chose not to include them in the present study. This suggests that the amount of signal discriminating complex experimental conditions is not strong enough to find inter-subject patterns robust to variability in many publicly available datasets, likely due to limited sample sizes and unoptimized experimental designs. We hope that broader recognition of the benefits of using inter-subject decoding to uncover neural coding principles across subjects—using functional alignment if necessary—will encourage investigators to collect and share more datasets supporting this type of analysis. Greater data availability will encourage robust, principled comparisons of alignment methods and foster progress in the field.

6.4.3 *Study limitations and future directions*

Although our study provides a broad evaluation of the performance of several functional alignment methods, there are several dimensions which we hope future work will better address. Notably, we did not thoroughly investigate how alignment performance is impacted by image resolution and whether data are represented on the surface or the volume. Using volumetric images downsampled to a standard resolution of 3mm isotropic enabled us to make fair comparisons across datasets at a reasonable computational cost. We also show in Section 6.5.6 that results from piecewise Procrustes alignment on the IBC dataset hold in a higher resolution, surface-based setting. Nonetheless, other functional alignment methods might show different patterns of performance in this setting or at different resolution levels. Moreover, applying these methods on high-resolution images is an exciting perspective to better understand how brain function details vary across subjects. To progress in this direction, a stronger focus on developing computationally efficient methods will be needed. The use of high-resolution parcellations—combined with more efficient implementations of piecewise Optimal Transport or a piecewise Shared Response Model—seem to be particularly promising directions.

We have not examined either the impact of alignment data on the learned transformations and whether this impact varies across cortex. That is, we could further ask whether certain kinds of stimuli may produce more accurate functional alignments for specialized functional regions. In general, the surveyed functional alignment methods view each subject alignment image as a sample, and the resulting transformation is trained to match corresponding samples across subjects. If some training images lack stable signal in a given ROI, functional alignment methods are unlikely to learn meaningful transformations in this region. Finally, this benchmark largely focussed on pairwise alignment models. Template-based models—beyond latent factor models as SRM—are an important area of research to further improve the usability of functional alignment methods, particularly in research settings comprising many subjects. In future work, we intend to address the above questions to learn more about when functional alignment methods are most appropriate.

6.5 ADDITIONAL ANALYSIS

6.5.1 *Absolute decoding accuracy of various methods*

Tables 6.2 and 6.3 report absolute decoding accuracies for *Experiment 1* and *Experiment 2*, to bring a different view of results presented in Figures 6.3 and 6.5, as relative improvements brought over anatom-

Methods/Dataset	IBC RSVP	IBC Sounds	Forrest	BOLD5000	Neuromod
Chance	16.7	16.7	20	25.5	12.5
Anatomical	38.2 ± 4.1	32.7 ±4.8	31.4 ± 4.7	33.3 ± 2.9	54.1 ± 6.7
Intra-subject	39.6 ± 3.1	36.4 ± 8.6	31.9 ± 5.4	34.9 ± 2.8	55.6 ± 7.7
Searchlight Procrustes	39.0 ± 5.1	32.6 ± 6.5	**33.6** ± 6.1	35.2 ± 2.0	65.5 ± 6.6
Piecewise Procrustes	42.0 ± 4.7	36.6 ± 5.5	**33.8** ± 6.4	36.4 ± 1.8	**67.4** ± 10.6
Piecewise Optimal Transport	**43.5** ± 5.5	**38.0** ± 9.5	**33.8** ± 5.6	36.6 ± 2.1	65.3 ± 9.1
Piecewise Shared Response Model	42.4 ± 4.0	37.0 ± 6.8	**33.7** ± 7.2	**39.6** ± 2.9	66.2 ± 7.0

Table 6.2: **Full-brain benchmark absolute decoding accuracy (%)**. Another view of *Experiment 1* results. Complementary display to Figure 6.3.

Methods/Dataset	IBC RSVP	IBC Sounds	Forrest	BOLD5000	Neuromod
Chance	16.7	16.7	20	25.5	12.5
Anatomical	22.3 ± 3.3	26.9 ± 6.9	28.6 ± 6.0	33.5 ± 3.3	50.2 ± 5.3
Intra-subject	22.0 ± 1.7	25.6 ± 2.7	27.5 ± 3.2	38.0 ± 2.6	51.4 ± 10.2
Procrustes	**31.0** ± 4.3	**32.7** ± 8.3	30.1 ± 7.0	30.1 ± 2.2	46.5 ± 7.8
Optimal Transport	24.9 ± 2.5	29.9 ± 5.4	28.9 ± 2.2	39.7 ± 2.7	**64.4** ± 8.9
Shared Response Model	30.4 ± 4.4	30.9 ± 6.3	**34.3** ± 5.2	**40.9** ± 3.6	32.6 ± 5.5

Table 6.3: **ROI benchmark absolute decoding accuracy (%)** Another view of *Experiment 2* results. Complementary display to Figure 6.5.

ical registration by various alignment methods. This "per dataset view" highlight that gains brought by best methods are substantial improvement over baseline, especially when compared to chance.

6.5.2 *Whole-brain decoding provides better accuracy than ROI-based decoding*

In Figure 6.7, we compare ROI-based and whole-brain inter-subject decoding accuracy improvements for piecewise Procrustes alignment above anatomical-only alignment. We see that whole-brain alignment generally shows higher inter-subject decoding improvements compared to ROI-based alignment. As mentioned in the main text, this result supports previous work from the inter-subject decoding literature (Chang et al., 2015; Schrouff et al., 2018a), and it suggests that full-brain piecewise alignment yields the best overall decoding pipeline.

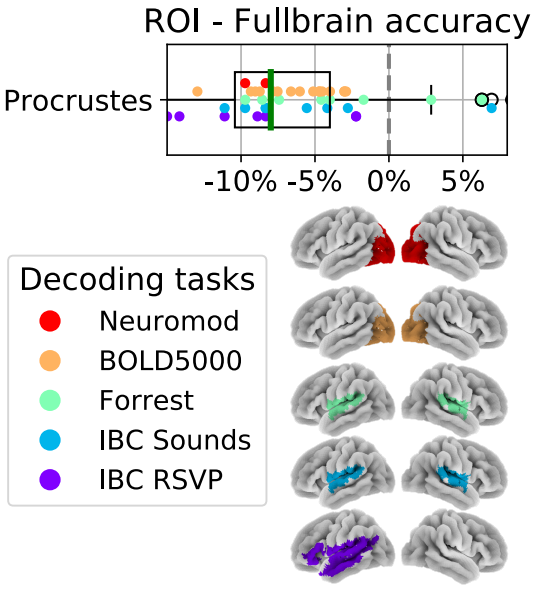

Figure 6.7: Comparing ROI and whole-brain decoding accuracy after piecewise Procrustes alignment. The ROIs used for each dataset are displayed on the *lower panel*. In the *upper panel*, we show the distribution of differences in decoding accuracy scores between ROI-based and whole-brain piecewise Procrustes alignment. Each dot represents a single subject, and subjects are colored according to their decoding task. Each difference score is calculated by subtracting the inter-subject decoding accuracy for whole-brain piecewise Procrustes alignment from the ROI-based piecewise Procrustes alignment accuracy score—for the same leave-one-subject-out cross-validation fold. The box plot thus describes the distribution of differences, where the green line represents the median value. We see that decoding accuracy is lower when performed within ROIs than when performed on the whole-brain data.

6.5.3 *Parcellation has limited impact on decoding accuracy*

To assess the impact of the parcellation used on piecewise alignment results, we compared decoding accuracy gains while varying the parcellation kind and resolution. First, we consider the multi-resolution (Schaefer et al., 2018) atlas, which was learned through a gradient weighted Markov random field method on resting state data from 1489 subjects. We compare this *a priori* parcellation to two parcellations learned directly on the subject's alignment data after 5mm fwhm Gaussian smoothing: K-means or Hierarchical K-means. All these parcellations were taken at ten resolutions from 100 to 1000 parcels.

As hierarchical K-means may be less familiar to readers, we briefly describe it in more detail here. This method is a variant of K-means, aimed specifically at obtaining more balanced parcels. To identify k parcels, we first apply K-means to cluster the voxels in $\sqrt{k}$ big clusters. Each of these "big clusters" is then clustered again in $\sqrt{k}$ to obtain a total of k smaller, well-balanced parcels. In this experiment, K-means and Hierarchical K-means implementations used are respectively from `scikit-learn` and `fmralign`, and fitted as part of `fmralign` alignment functions on the source subject data.

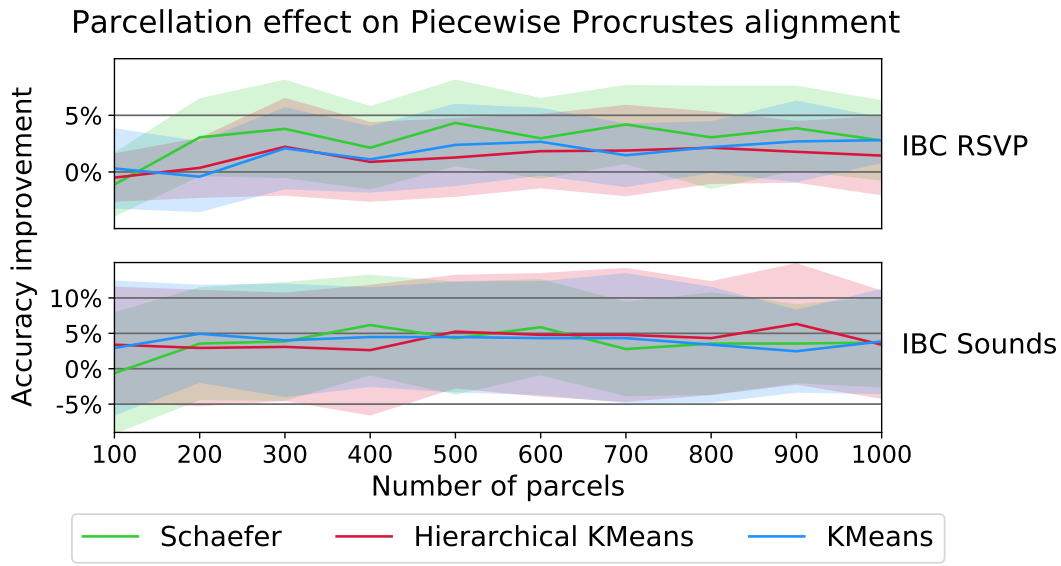

Figure 6.8: Effect of parcellation type and resolution on Piecewise Procrustes decoding accuracy improvement over anatomical alignment. We consider the impact of parcellation type (the *a priori* Schaefer atlas or learned directly on the data with k-means or hierarchical k-means) and resolution (from 100 to 1000 parcels). Results are shown for the IBC RSVP and IBC Sounds decoding tasks. Each line represents the average accuracy improvement for piecewise Procrustes over standard, anatomical-only alignment, and the confidence band represents the range of accuracy improvements seen across all IBC subjects. Accuracy improvements are calculated by subtracting anatomical-only inter-subject decoding accuracy scores for the same leave-one-subject-out cross-validation fold. We see that parcellation type and resolution show limited impact on accuracy gains.

We plot piecewise Procrustes accuracy improvements for these three parcellation methods and ten resolutions in Figure 6.8. Here, we only show the IBC Sounds and IBC RSVP decoding tasks to ease in interpretation. Overall, we observe on these two tasks that the type and resolution of parcellation used does not have a strong impact on accuracy improvements above anatomical-only alignment. We therefore suggest that *piecewise alignment* can be used with confidence that the parcellation choice won't strongly impact its results.

6.5.4 *Grid-search of Piecewise SRM hyperparameters*

As a piecewise implementation of SRM is a novel contribution from this work, we had no prior knowledge on how to properly set hyperparameters from this method (clustering type and resolution as well as the number of components to set for each local SRM).

For the type of clustering, we limited ourselves to a pre-computed parcellation (the Schaefer atlas) available at various resolutions. This is based on the intuitions acquired on Procrustes (see Section 6.5.3) that the parcellation type was not of utmost importance to decoding results. We ran a cross-validation on the two remaining parameters. We used Schaefer atlas at resolution: [100,300,500,700] while our number of

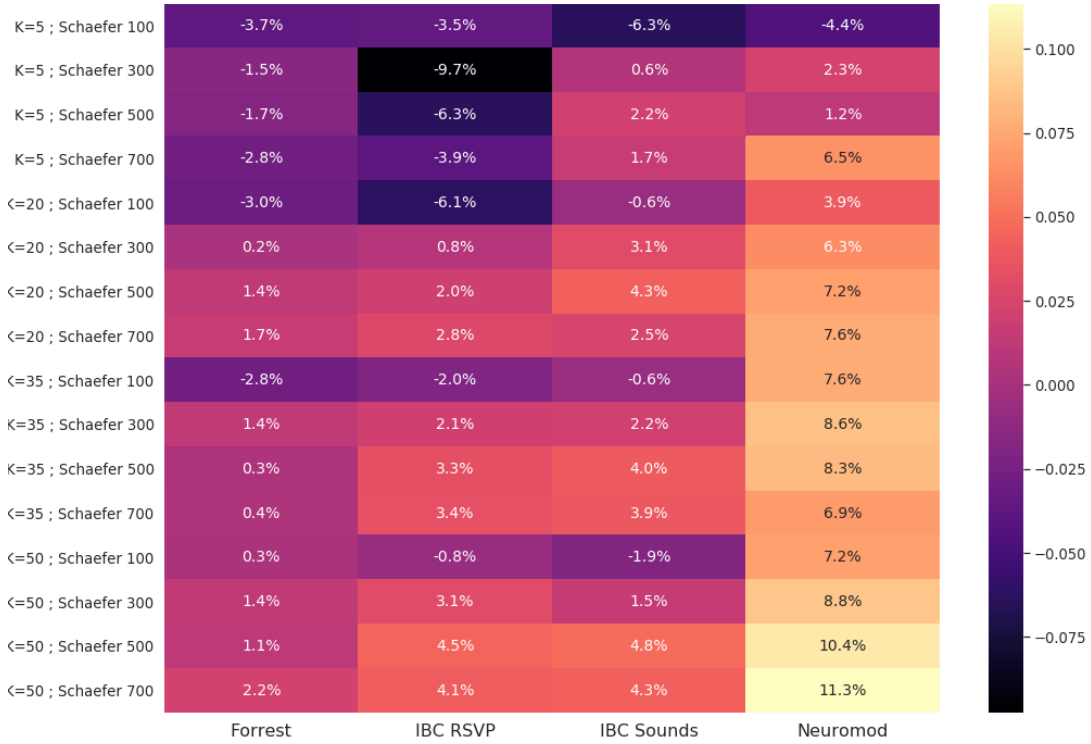

Figure 6.9: Grid search of Piecewise SRM hyperparameters impact on decoding accuracy across datasets. We considered a grid of 4 parcellations (Schaefer atlas at resolution 100, 300, 500 and 700) and 4 different values of k (number of SRM components) for each model fitted on a parcel. We ran our inter-subject decoding pipeline on four inter-subject decoding task (in columns, among those used in the main benchmark). We report here the decoding accuracy improvement over anatomical baseline across datasets for each set of parameter (in line). Although we didn't have the computational means to run an extensive grid-search, we can already conclude that high-resolution parcellations (and thus more fitted local SRMs) yield a higher decoding gain as long as they come with enough component. Decoding accuracy is also positively linked with K, probably up to a plateau that we did not clearly reach with our limited grid. For the main benchmark, we retained the last line model ($K = 50$, Schaefer 700).

components ranged from [5,25,35,50]. Figure 6.9 present the results of this cross-validation, that led us to chose Schaefer atlas 700 and 50 components as hyperparameters for our main experiments.

6.5.5 *Functional alignment is not merely smoothing*

Gaussian smoothing is a common preprocessing step in neuroimaging group studies, which reconciles dissimilar subject-level signals by smoothing over inter-individual variability. Our qualitative results (Section 6.3.3) show that best performing alignment methods do not

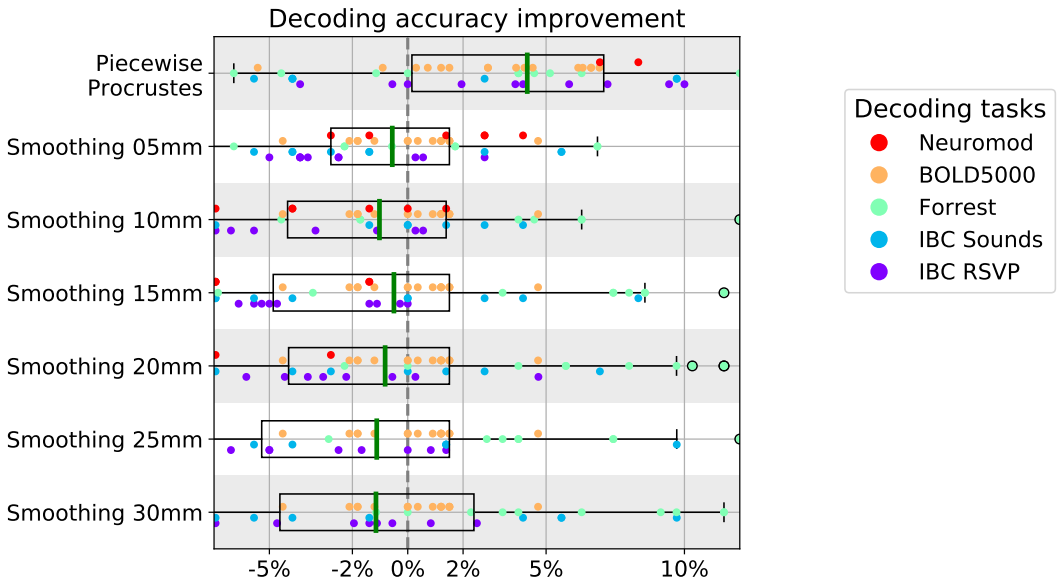

Figure 6.10: Decoding accuracy does not improve after Gaussian smoothing over anatomical alignment. For six smoothing kernels fwhm, we show inter-subject decoding accuracy scores after subtracting anatomical-only inter-subject decoding accuracy for the same leave-one-subject-out cross-validation fold. Each dot represents a single subject, and subjects are colored according to their decoding task. We also show differences in decoding accuracy scores for the reference functional alignment method piecewise Procrustes, again as compared to anatomical-only alignment. Box plots describe the distribution of values for a smoothing kernel or an alignment method, and the green line indicates the median. We see that Gaussian smoothing does not show the same pattern of decoding accuracy differences as the reference functional alignment method.

seem to smooth the signal across voxels, but instead preserve the signal specificity while matching its geometry with the target subject functional topography. Specifically, we compared decoding gains from six different Gaussian smoothing kernels to those obtained through the reference method piecewise Procrustes alignment.

The results displayed in Figure 6.10 clearly support previous findings (Guntupalli et al., 2016) that smoothing does not improve inter-subject decoding performance—and therefore recover mutual information—in the same way that functional alignment.

6.5.6 *Impact of the data representation and resolution*

Oosterhof et al., 2011 argued that functional alignment benefits from working with a representation of the fMRI signal on the cortical surface (Coalson, Van Essen, and Glasser, 2018). Relatedly, we would also expect that the resolution of the data representation—whether in the surface or the volume—will impact the quality of the alignment learned.

To assess the dependence of our 3mm volumetric results presented in the main text on sampling parameters, we replicated our inter-

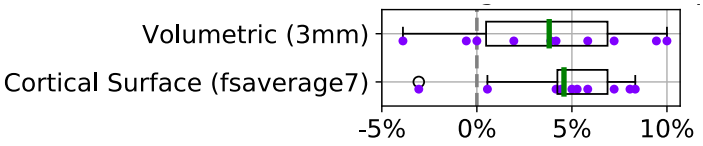

Figure 6.11: Comparing piecewise Procrustes accuracy improvements across volumetric and surface data representations. For the IBC RSVP task, we compare piecewise Procrustes decoding accuracy scores to anatomical-only alignment. Each dot represents an IBC subject, where their difference score is calculated by subtracting the inter-subject decoding accuracy for anatomical-only alignment from the piecewise Procrustes alignment accuracy score for the same leave-one-subject-out cross-validation fold; i.e., where they are the left-out subject. We compare these difference scores, as calculated using data in the volume (3mm resolution), to data on the high-resolution cortical surface (*fsaverage7*). Box plots describe the distribution of values for a data representation, and the green line indicates the median. We see that the high-resolution surface representation yields a moderate gain of decoding accuracy, compared to 3mm isotropic volumetric representation.

subject decoding framework with the IBC RSVP language task data on a high-resolution cortical surface representation(*fsaverage7*)(obtained through `freesurfer` surfacic projection of full-resolution raw images in their respective subject space, later on mapped to the common surfacic template). This surface mesh includes 168k cortical nodes per hemisphere, which we divided into 350 parcels per hemisphere using *Schaefer* atlas at scale 700.

We provide results for the inter-subject decoding accuracy gains seen with the reference functional alignment method of piecewise Procrustes over standard, anatomical-only alignment. We had to limit to this setting because (i) replicating this analysis on every dataset would represent an important amount of processing work, and (ii) working on other methods than piecewise Procrustes on this very large data is computationally prohibitive.

The results displayed in Figure 6.11 show that although decoding gains are a little higher using high-resolution surface-based representation, they remain in the same range as the volume-based representation. This shows that a 10-fold higher resolution can help match more precisely topographies across subjects (and reduce the decoding variance as a consequence), but no important marginal gains can be expected from it. In the end, the signal available for use is bounded by the same rough limitations: test-retest reliability in each subject.

6.5.7 *IBC alignment data explained*

In this work, 53 contrasts were pulled together are used as alignment for IBC dataset. The contrasts are common to all subjects and taken from both HCP and ARCHI protocol. In order, they are labelled: *audio left button press, audio right button press, video left button press, video right button press, horizontal checkerboard, vertical checkerboard, audio sentence, video sentence, audio computation, video computation, saccades,*

rotation hand, rotation side, object grasp, object orientation, mechanistic audio, mechanistic video, triangle mental, triangle random, false belief audio, false belief video, speech sound, non speech sound, face gender, face control, face trusty, expression intention, expression gender, expression control, shape, face, punishment, reward, left hand, right hand, left foot, right foot, tongue, cue, story, math, relational, match, mental, random, oback body, 2back body, oback face, 2back face, oback tools, 2back tools, oback place, 2back place.

To know more, please visit the relevant IBC documentation.

6.5.8 *fMRIprep preprocessing*

Results included in this manuscript come from preprocessing performed using *fMRIprep* (Esteban et al., 2019), which is based on *Nipype* 1.5.0 (Gorgolewski et al., 2011).

ANATOMICAL DATA PREPROCESSING The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.2.0 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR_002823, Zhang, Brady, and Smith, 2001). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009, RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym).

FUNCTIONAL DATA PREPROCESSING For each subject's BOLD runs (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated by aligning and averaging 1 single-band references (SBRefs). A Bo-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with opposing phase-encoding directions, with `3dQwarp` Cox and Hyde, 1997 (AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with six degrees of freedom.

Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatio-temporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al., 2002).

First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIprep*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIprep*.

All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964).

Many internal operations of *fMRIprep* use *Nilearn* 0.6.2 (Abraham et al., 2014b, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIprep*'s documentation.

7 CONCLUSION

In this thesis, we have proposed a new local functional alignment method using Optimal Transport and a new Piecewise aggregation scheme to perform whole-brain alignment. Combined with the templating method we proposed, this enables to perform functional alignment at common resolution on a group of subjects to reduce inter-subject variability. We also introduced a Piecewise Shared Response Model which is another instance of the functional alignment concept using aggregated latent factor models and compared it with an original Intra-subject alignment scheme. We produced a rigorous and thorough benchmark of several alignment methods using inter-subject decoding on five decoding tasks from four datasets. This allowed us to measure the ability of alignment methods to transfer signal across subjects while preserving signal specificity in a range of cognitive tasks considered in the field. This benchmark enables us to conclude that several alignment methods indeed hold the potential to handle inter-subject variability. The alignment methods compared were able to recover more than half of decoding accuracy lost to variability. We hope that this benchmark will help neuroscientists to get a clearer account of available alignment methods performance to be expected—and their running cost—as well as methodological choices to be made. Hopefully, this will help disseminate alignment that indeed could be a useful analysis step in many studies to strengthen group statistical conclusions, help in precise brain-mapping endeavors and ultimately benefit translational research. We especially focused on releasing efficient and well-documented open-source implementations of these ideas, as we believe this is key to foster their broader use.

PRACTICAL CHALLENGES

The biggest challenge we encountered during this thesis is data availability and disparity. While we feel lucky to have been able to investigate this topic (intensively relying on datasets that were acquired just prior or during this thesis work) we want to point that the scarceness of such resources and their design were important constraints to design and realize our experiments. Many interesting ideas will remain impossible to test and use in this field until bigger deep phenotyping datasets become available. Similarly, a significant portion of this thesis work was spent on screening and analyzing numerous datasets in order to find the rare appropriate ones for our experiments. We

acknowledge this is a common problem in fMRI studies and one that is hard to solve. Additionally, fMRI data are heavy and alignment is a data-driven process relying on in-depth individual characterization and thus requires a significant amount of computation, let alone running those methods thousands of time for benchmarking purposes. We were lucky to benefit from INRIA's and Parietal's important computing resources and parallel computing and software expertise, which helped alleviate this problem.

A second important challenge during this thesis was the conceptual inconsistency observed in the literature pertaining to variability and ways to handle it. As a matter which is hard to study and model, several related conceptual frameworks coexisted without being related or compared in any meaningful ways. As a consequence, several dozens of methodological advances have been proposed without being properly compared with each other. Alignment methods were introduced with metrics and baselines focused on algorithmic validation, and the overall foundations for methodological developments in this field are unsteady. This constrained our work to focus heavily on empirical validation, as a necessary foundation for later well-motivated developments.

FUTURE DIRECTIONS

HIGH RESOLUTION Although Chapter 6 provides a comprehensive benchmark of several functional alignment methods, there are several dimensions which we hope future work will better address to improve our understanding of those methods. Notably, we did not thoroughly investigate how alignment performance is impacted by image resolution and whether data are represented on the surface or the volume. Using volumetric images downsampled to a standard resolution of 3mm isotropic enabled us to make fair comparisons across datasets at a reasonable computational cost. We also show in Section 6.5.6 that results from piecewise Procrustes alignment on the IBC dataset hold in a higher resolution, surface-based setting. Nonetheless, other functional alignment methods might show different patterns of performance in this setting or at different resolution levels. Moreover, applying these methods on high-resolution images is an exciting perspective to better understand how brain function details vary across subjects. To progress in this direction, a stronger focus on developing computationally efficient methods will be needed. The use of high-resolution parcellations— combined with efficient implementations of piecewise Optimal Transport or a piecewise Shared Response Model—seem to be particularly promising directions.

FUNCTIONAL CHARACTERIZATION OF BRAIN REGIONS We have not examined either the impact of alignment data on the learned transformations or whether this impact varies across cortex. That is, we could further ask whether certain kinds of stimuli may produce more accurate functional alignments for specialized functional regions. In general, the surveyed functional alignment methods view each subject alignment image as a sample, and the resulting transformation is trained to match corresponding samples across subjects. If some training images lack stable signal in a given Region of Interest (ROI), functional alignment methods are unlikely to learn meaningful transformations in this region. To improve our usability of alignment techniques, we ought to better understand the impact of alignment data used. More generally, our alignment proves that parcels can be characterized reasonably well by a certain number of functional dimensions, and understanding which dimensions are relevant for each parcel is an exciting brain mapping endeavor.

INTER-DATASET ALIGNMENT While we have shown that inter-subject variability could be handled in a given dataset context, an important question to get closer to translational purposes is whether they can also help to link subjects that were not imaged in exactly the same context. Or to put it differently, how can matching based on functional characterization help deal with site-effect on top of inter-individual variability. Our benchmarking paradigm based on inter-subject decoding could be extended to address this kind of question. A compelling first step towards generalizability would be to integrate information from one deep-phenotyping dataset with another. If learned models are to be generalizable to broader subsets of the population, they should first generalize to another, similarly-sampled cohort. An important limitation is that it is especially rare to have synchronized alignment data in common across different datasets. We are thus currently exploring options to learn alignment over various cohorts of subjects by bridging some functional characterization data.

TEMPLATE-BASED MODELS Latent factor models such as SRM provide an interesting way to map individual topographies inside a common group framework. Importantly, they enable to transfer information across subjects. But, beyond this pragmatic view of a common intermediate representation for computational purposes and learning, it remains an important goal to derive some group model with statistical significance that would embody the general population, or a chosen homogenous sub-group functional organization principles precisely. Although our attempts did not solve this problem yet, this is an important future direction for brain imaging statistical analysis.

WHAT DID NOT WORK

During this thesis, we also explored several ideas that did not improve results.

7.0.1 *Kernel Truncation for high-dimensional optimal transport*

To go beyond the piecewise aggregation of local optimal transport alignments, we tried to apply kernel-truncation (Schmitzer, 2019), in order to build a multi-scale optimal transport implementation (Feydy et al., 2019). This method derives a coarse, low-resolution matching of histograms in a first step and then iteratively focuses on higher-resolution low potential regions (regions where cost is below a given level) to try to refine the matching where it can be improved. In theory, this would enable to recover optimal whole-brain alignment across subjects at a reasonable computational cost while getting rid of potential discontinuity effects introduced by parcellation techniques. In practice, the numerical stability of this method entails a complex fine-tuning in order to reach the precision of matching required in our use-case, and we did not manage to show increased metrics using those alignments.

7.0.2 *Wasserstein barycenter to build templates informed by signal geometry*

As developed in Section 5.3 we tried to use Wasserstein Barycenters to improve initialization of our iterative templating scheme, by taking subjects' signals geometry into account. However, we did not manage yet to extend our use of signal dissimilarity to build a transport cost matrix—that proved very successful for pairwise alignments— to a group setting, nor to apply spatial transport in a convincing way.

7.0.3 *Hyperalignment for MEG data*

In a joint work with Hugo Richard, we tried to use comparable techniques to improve Magnetoencephalography (MEG) decoding using functional alignment methods. Although the spatial resolution is much lower, which defeats part of alignment's purpose, we wanted to see if combining sensors signals across subjects would help in finding better decoding patterns. We tried to learn alignment. Although we tried several state-of-the-art methods on this problem, they didn't bring improvement.

RESOURCES CREATED DURING THIS THESIS

7.0.4 *Software contributions*

We leave after this thesis several reference implementations for the state-of-the art methods that we proposed and benchmarked, including those of Piecewise Procrustes, Piecewise Optimal Transport and Piecewise Shared Response Model. Those implementations are packaged in the *fmralign* module: https://github.com/Parietal-INRIA/fmralign, with an emphasis on performance and documentation to improve ease-of-use. With these tools, any neuroscientist using Python ecosystem tools can easily input images from his or her subjects and get back their aligned subjects with a *Scikit-learn* API. We hope these contributions will foster functional alignment adoption. We also released additional alignment tutorials in https://github.com/neurodatascience/fmralign-tutorials.

To evaluate methodological contributions, we also designed a benchmark, for which we try to open-source as many components as possible to ease replication and further benchmarking endeavors. We make code available as well as data derivatives for the IBC dataset in the https://github.com/neurodatascience/fmralign-benchmark module.

Besides contributions directly related to my thesis work, I contributed to the *Nilearn* package: https://github.com/nilearn/nilearn during my entire PhD. Nilearn is a widely used Python package for statistical analysis and predictive modelling of neuroimaging data. I primarily contributed to the "Decoding" module of this package, as well as to improving its documentation and friendliness for new contributors. I also took part in the organization of several teaching and coding events for Nilearn. Additionally, I was briefly involved in the maintenance of *pypreprocess* https://github.com/neurospin/pypreprocess.

7.0.5 *Other projects*

During my PhD, I also got interested in various fMRI related projects. To develop my machine learning skills on related subjects, I led a team of fellow PhD students from INRIA Parietal team in a Kaggle challenge participation (https://www.kaggle.com/c/trends-assessment-prediction). This competition was aimed at improving translational use of resting state fMRI data in order to predict variables of clinical interest, including age. In our final solution, we used Riemannian geometry as well as state-of-the-art estimators and hierarchical schemes to predict those targets from our source data, which ranked us in the top 2% contenders of this challenge. Besides machine learning, this project helped me improve a lot my

ability to design code and infrastructure for efficient scientific collaboration. This was also a fun and collaborative way to explore a scientific problem at the edge of many core topics: machine learning, statistics, cognition, and the brain.

I also took part in the analysis of several fMRI datasets. Two of those analysis were included in publications: one for Individual Brain Charting fourth dataset extension (Torre et al., in preparation), and one for an extension of Optimal Transport with an additional group penalization (Janati et al., 2019, 2020).

REFERENCES

Abraham, Alexandre, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, and Gael Varoquaux (2013). "Extracting brain regions from rest fMRI with total-variation constrained dictionary learning." In: *International conference on medical image computing and computer-assisted intervention*. Springer, pp. 607–615.

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux (Feb. 2014a). "Machine learning for neuroimaging with scikit-learn." en. In: *Front. Neuroinform.* 8, p. 14.

– (2014b). "Machine learning for neuroimaging with scikit-learn." In: *Frontiers in neuroinformatics* 8, p. 14.

Aglieri, V., B. Cagna, P. Belin, and S. Takerkart (2019). *InterTVA. A multimodal MRI dataset for the study of inter-individual differences in voice perception and identification*. https://openneuro.org/datasets/ds001771/versions/1.0.2.

Al-Wasity, Salim, Stefan Vogt, Aleksandra Vuckovic, and Frank E Pollick (2020). "Hyperalignment of motor cortical areas based on motor imagery during action observation." In: *Scientific reports* 10.1, pp. 1–12.

Alaux, Jean, Edouard Grave, Marco Cuturi, and Armand Joulin (2018). "Unsupervised hyperalignment for multilingual word embeddings." In: *arXiv preprint arXiv:1811.01124*.

Alberton, Bianca AV, Thomas E Nichols, Humberto R Gamba, and Anderson M Winkler (2020). "Multiple testing correction over contrasts for brain imaging." In: *NeuroImage* 216, p. 116760.

Ashburner, John and Karl J Friston (2005). "Unified segmentation." In: *Neuroimage* 26.3, pp. 839–851.

– (2011). "Diffeomorphic registration using geodesic shooting and gauss newton optimisation." In: *Neuroimage , 55(3):954–967*.

Avants, Brian B., Charles L. Epstein, Murray Grossman, and James C. Gee (2008). "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain." In: *Medical Image Analysis* 12.1, pp. 26–41. URL: http://dblp.uni-trier.de/db/journals/mia/mia12.html#AvantsEGG08.

Bach, Francis R and Michael I Jordan (2005). "A probabilistic interpretation of canonical correlation analysis." In:

Baldassano, Christopher, Janice Chen, Asieh Zadbood, Jonathan W Pillow, Uri Hasson, and Kenneth A Norman (2017). "Discovering event structure in continuous narrative perception and memory." In: *Neuron* 95.3, pp. 709–721.

Barch, D. M., G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, and C. Feldt (2013). "Function in the human connectome: task-fMRI and individual differences in behavior." In: *Neuroimage* 80, pp. 169–189.

Bartels, Andreas, Semir Zeki, and Nikos K Logothetis (2008). "Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain." In: *Cerebral cortex* 18.3, pp. 705–717.

Bazeille, Thomas, Elizabeth Dupre, Hugo Richard, Jean-Baptiste Poline, and Bertrand Thirion (under review). "An empirical evaluation of functional alignment using inter-subject decoding." In: URL: https://www.biorxiv.org/content/10.1101/2020.12.07.415000v2.

Bazeille, Thomas, Hugo Richard, Hicham Janati, and Bertrand Thirion (2019). "Local Optimal Transport for Functional Brain Template Estimation." In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 237–248.

Beg, M Faisal, Michael I Miller, Alain Trouvé, and Laurent Younes (2005). "Computing large deformation metric mappings via geodesic flows of diffeomorphisms." In: *International journal of computer vision* 61.2, pp. 139–157.

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T Liu (2007). "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI." In: *Neuroimage* 37.1, pp. 90–101.

Bellec, Pierre and Julie Boyle (2019). "Bridging the gap between perception and action: the case for neuroimaging, AI and video games." In:

Bellec, Pierre, Pedro Rosa-Neto, Oliver C. Lyttelton, Habib Benali, and Alan C. Evans (2010). "Multi-level bootstrap analysis of stable clusters in resting-state fMRI." In: *NeuroImage* 51.3, pp. 1126 –1139. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2010.02.082. URL: http://www.sciencedirect.com/science/article/pii/S1053811910002697.

Benamou, Jean-David, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré (2015). "Iterative Bregman projections for regularized transportation problems." In: *SIAM Journal on Scientific Computing* 37.2, A1111–A1138.

Bennett, Craig M and Michael B Miller (2010). "How reliable are the results from functional magnetic resonance imaging?" In: *Annals of the New York Academy of Sciences* 1191.1, pp. 133–155.

Benton, Adrian, Huda Khayrallah, Biman Gujral, Dee Ann Reisinger, Sheng Zhang, and Raman Arora (2017). "Deep generalized canonical correlation analysis." In: *arXiv preprint arXiv:1702.02519*.

Bilenko, Natalia Y and Jack L Gallant (Nov. 2016). "Pyrcca: Regularized Kernel Canonical Correlation Analysis in Python and Its Applications to Neuroimaging." en. In: *Front. Neuroinform.* 10, p. 49.

Blank, Idan A, Swathi Kiran, and Evelina Fedorenko (2017). "Can neuroimaging help aphasia researchers? Addressing generalizability, variability, and interpretability." In: *Cognitive neuropsychology* 34.6, pp. 377–393.

Botvinik-Nezer, Rotem, Felix Holzmeister, Colin F Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A Mumford, R Alison Adcock, et al. (2020). "Variability in the analysis of a single neuroimaging dataset by many teams." In: *Nature*, pp. 1–7.

Boyle, J. A., B. Pinsard, A. Boukhdhir, S. Belleville, S. Brambatti, J. Chen, J. Cohen-Adad, A. Cyr, A. Fuente, P. Rainville, and P. Bellec (2020). *The Courtois project on neuronal modelling: 2020 data release*. https://docs.cneuromod.ca. Presented at the 26th annual meeting of the Organization for Human Brain Mapping.

Braver, Todd S., Alexander Kizhner, Rongxiang Tang, Michael C. Freund, and Joset A. Etzel (2020). "The Dual Mechanisms of Cognitive Control (DMCC) Project." In: *bioRxiv*. DOI: 10.1101/2020.09.18.304402. eprint: https://www.biorxiv.org/content/early/2020/09/20/2020.09.18.304402.full.pdf. URL: https://www.biorxiv.org/content/early/2020/09/20/2020.09.18.304402.

Brown, Gregory G, Daniel H Mathalon, Hal Stern, Judith Ford, Bryon Mueller, Douglas N Greve, Gregory McCarthy, James Voyvodic, Gary Glover, Michele Diaz, et al. (2011). "Multisite reliability of cognitive BOLD data." In: *Neuroimage* 54.3, pp. 2163–2175.

Buckner, Randy L., Fenna M. Krienen, Angela Castellanos, Julio C. Diaz, and B. T. Thomas Yeo (2011). "The organization of the human cerebellum estimated by intrinsic functional connectivity." In: *Journal of Neurophysiology* 106.5. PMID: 21795627, pp. 2322–2345. DOI: 10.1152/jn.00339.2011. URL: https://doi.org/10.1152/jn.00339.2011.

Busch, Erica L, Lukas Slipski, Ma Feilong, J Swaroop Guntupalli, Matteo Visconti di Oleggio Castello, Jeremy F Huckins, Samuel A Nastase, M Ida Gobbini, Tor D Wager, and James V Haxby (2021). "Hybrid Hyperalignment: A single high-dimensional model of shared information embedded in cortical patterns of response and functional connectivity." In: *NeuroImage* 233, p. 117975.

Button, Katherine S, John PA Ioannidis, Claire Mokrysz, Brian A Nosek, Jonathan Flint, Emma SJ Robinson, and Marcus R Munafò (2013). "Power failure: why small sample size undermines the reliability of neuroscience." In: *Nature Reviews Neuroscience* 14.5, p. 365.

Caballero-Gaudes, César and Richard C Reynolds (2017). "Methods for cleaning the BOLD fMRI signal." In: *Neuroimage* 154, pp. 128–149.

Casey, BJ, Tariq Cannonier, May I Conley, Alexandra O Cohen, Deanna M Barch, Mary M Heitzeg, Mary E Soules, Theresa Teslovich, Danielle V Dellarco, Hugh Garavan, et al. (2018). "The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites." In: *Developmental cognitive neuroscience* 32, pp. 43–54.

Chang, Luke J, Peter J Gianaros, Stephen B Manuck, Anjali Krishnan, and Tor D Wager (June 2015). "A Sensitive and Specific Neural

Signature for Picture-Induced Negative Affect." en. In: *PLoS Biol.* 13.6, e1002180.

Chang, Nadine, John A Pyles, Austin Marcus, Abhinav Gupta, Michael J Tarr, and Elissa M Aminoff (May 2019). "BOLD5000, a public fMRI dataset while viewing 5000 visual images." en. In: *Sci Data* 6.1, p. 49.

Chen, Chaoqi, Weiping Xie, Wenbing Huang, Yu Rong, Xinghao Ding, Yue Huang, Tingyang Xu, and Junzhou Huang (2019). "Progressive feature alignment for unsupervised domain adaptation." In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 627–636.

Chen, Po-Hsuan, Janice Chen, Yaara Yeshurun, Uri Hasson, James Haxby, and Peter J Ramadge (2015). "A Reduced-Dimension fMRI Shared Response Model." In: *Advances in Neural Information Processing Systems 28*. Ed. by C Cortes, N D Lawrence, D D Lee, M Sugiyama, and R Garnett. Curran Associates, Inc., pp. 460–468.

Chen, Po-Hsuan, J Swaroop Guntupalli, James V Haxby, and Peter J Ramadge (2014). "Joint SVD-Hyperalignment for multi-subject FMRI data alignment." In: *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, pp. 1–6.

Chen, Po-Hsuan, Xia Zhu, Hejia Zhang, Javier S. Turek, Janice Chen, Theodore L. Willke, Uri Hasson, and Peter J. Ramadge (2016). *A Convolutional Autoencoder for Multi-Subject fMRI Data Aggregation*. arXiv: 1608.04846 [stat.ML].

Chevalier, Jérôme-Alexis, Tuan-Binh Nguyen, Joseph Salmon, Gaël Varoquaux, and Bertrand Thirion (2021). "Decoding with confidence: Statistical control on decoder maps." In: *NeuroImage* 234, p. 117921.

Chiba, Toshinori, Tetsufumi Kanazawa, Ai Koizumi, Kentarou Ide, Vincent Taschereau-Dumouchel, Shuken Boku, Akitoyo Hishimoto, Miyako Shirakawa, Ichiro Sora, Hakwan Lau, et al. (2019). "Current status of neurofeedback for post-traumatic stress disorder: a systematic review and the possibility of decoded neurofeedback." In: *Frontiers in human neuroscience* 13, p. 233.

Chizat, Lenaic, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard (2018). "Scaling algorithms for unbalanced optimal transport problems." In: *Mathematics of Computation* 87.314, pp. 2563–2609.

Churchland, Paul M (1998). "Conceptual similarity across sensory and neural diversity: the Fodor/Lepore challenge answered." In: *J. Philos.* 95.1, pp. 5–32.

Coalson, Timothy S, David C Van Essen, and Matthew F Glasser (2018). "The impact of traditional neuroimaging methods on the spatial localization of cortical areas." In: *Proceedings of the National Academy of Sciences* 115.27, E6356–E6365.

Coifman, Ronald R and Stéphane Lafon (2006). "Diffusion maps." In: *Applied and computational harmonic analysis* 21.1, pp. 5–30.

Conroy, Bryan R, Benjamin D Singer, J Swaroop Guntupalli, Peter J Ramadge, and James V Haxby (2013). "Inter-subject alignment of human cortical anatomy using functional connectivity." In: *NeuroImage* 81, pp. 400–411.

Coste, Clio P, Sepideh Sadaghiani, Karl J Friston, and Andreas Kleinschmidt (2011). "Ongoing brain activity fluctuations directly account for intertrial and indirectly for intersubject variability in Stroop task performance." In: *Cerebral cortex* 21.11, pp. 2612–2619.

Cox, Robert W. and James S. Hyde (1997). "Software tools for analysis and visualization of fMRI data." In: *NMR in Biomedicine* 10.4-5, pp. 171–178. DOI: 10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.

Cuturi, Marco (2013). "Sinkhorn Distances: Lightspeed Computation of Optimal Transport." In: *Advances in Neural Information Processing Systems 26*. Ed. by C J C Burges, L Bottou, M Welling, Z Ghahramani, and K Q Weinberger. Curran Associates, Inc., pp. 2292–2300.

Dadi, Kamalaker, Gaël Varoquaux, Antonia Machlouzarides-Shalit, Krzysztof J Gorgolewski, Demian Wassermann, Bertrand Thirion, and Arthur Mensch (2020). "Fine-grain atlases of functional modes for fMRI analysis." In: *NeuroImage* 221, p. 117126.

Davatzikos, Christos, Feng Xu, Yang An, Yong Fan, and Susan M Resnick (2009). "Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index." In: *Brain* 132.8, pp. 2026–2035.

Demertzi, Athena, Georgios Antonopoulos, Lizette Heine, Henning U Voss, Julia Sophia Crone, Carlo de Los Angeles, Mohamed Ali Bahri, Carol Di Perri, Audrey Vanhaudenhuyse, Vanessa Charland-Verville, et al. (2015). "Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients." In: *Brain* 138.9, pp. 2619–2631.

Dockès, Jérôme, Russell A Poldrack, Romain Primet, Hande Gözükan, Tal Yarkoni, Fabian Suchanek, Bertrand Thirion, and Gaël Varoquaux (2020). "NeuroQuery, comprehensive meta-analysis of human brain mapping." In: *eLife* 9. Ed. by Christian Büchel, Thomas Yeo, and Tor D Wager, e53385. ISSN: 2050-084X. DOI: 10.7554/eLife.53385. URL: https://doi.org/10.7554/eLife.53385.

Dohmatob, Elvis, Gael Varoquaux, and Bertrand Thirion (2018). "Inter-subject registration of functional images: do we need anatomical images?" In: *Frontiers in neuroscience* 12, p. 64.

Dubois, Julien and Ralph Adolphs (2016). "Building a science of individual differences from fMRI." In: *Trends in cognitive sciences* 20.6, pp. 425–443.

Eickhoff, Simon B, Angela R Laird, Christian Grefkes, Ling E Wang, Karl Zilles, and Peter T Fox (2009). "Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-

effects approach based on empirical estimates of spatial uncertainty." In: *Human brain mapping* 30.9, pp. 2907–2926.

Elliott, Maxwell L, Annchen R Knodt, David Ireland, Meriwether L Morris, Richie Poulton, Sandhya Ramrakha, Maria L Sison, Terrie E Moffitt, Avshalom Caspi, and Ahmad R Hariri (2020). "What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis." In: *Psychological Science* 31.7, pp. 792–806.

Esteban, Oscar, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit S Ghosh, Jessey Wright, Joke Durnez, Russell A Poldrack, and Krzysztof J Gorgolewski (Jan. 2019). "fMRIPrep: a robust preprocessing pipeline for functional MRI." en. In: *Nat. Methods* 16.1, pp. 111–116.

Evans, Alan C, Andrew L Janke, D Louis Collins, and Sylvain Baillet (2012). "Brain templates and atlases." In: *Neuroimage* 62.2, pp. 911–922.

Farahani, Abolfazl, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia (2020). "A Brief Review of Domain Adaptation." In: *arXiv preprint arXiv:2010.03978*.

Fedorenko, Evelina, Michael K Behr, and Nancy Kanwisher (2011). "Functional specificity for highlevel linguistic processing in the human brain." In: *Proceedings of the National Academy of Sciences* 108.39, pp. 16428–16433.

Feydy, Jean, Pierre Roussillon, Alain Trouvé, and Pietro Gori (2019). "Fast and scalable optimal transport for brain tractograms." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 636–644.

Finn, Emily S, Enrico Glerean, Arman Y Khojandi, Dylan Nielson, Peter J Molfese, Daniel A Handwerker, and Peter A Bandettini (2020). "Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging." In: *NeuroImage* 215, p. 116828.

Fonov, VS, AC Evans, RC McKinstry, CR Almli, and DL Collins (2009). "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood." In: *NeuroImage* 47, Supplement 1, S102. DOI: 10.1016/S1053-8119(09)70884-5.

Friston, Karl J, John Ashburner, Christopher D Frith, J-B Poline, John D Heather, and Richard SJ Frackowiak (1995). "Spatial registration and normalization of images." In: *Human brain mapping* 3.3, pp. 165–189.

Friston, Karl J, P Fletcher, Oliver Josephs, ANDREW Holmes, MD Rugg, and Robert Turner (1998). "Event-related fMRI: characterizing differential responses." In: *Neuroimage* 7.1, pp. 30–40.

Friston, Karl J, Steven Williams, Robert Howard, Richard SJ Frackowiak, and Robert Turner (1996). "Movement-related effects in fMRI time-series." In: *Magnetic resonance in medicine* 35.3, pp. 346–355.

Ganin, Yaroslav, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky (2016). "Domain-adversarial training of neural networks." In: *The journal of machine learning research* 17.1, pp. 2096–2030.

Giedd, Jay N, Armin Raznahan, Aaron Alexander-Bloch, Eric Schmitt, Nitin Gogtay, and Judith L Rapoport (2015). "Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development." In: *Neuropsychopharmacology* 40.1, pp. 43–49.

Gilron, Roee, Jonathan Rosenblatt, Oluwasanmi Koyejo, Russell A Poldrack, and Roy Mukamel (2017). "What's in a pattern? Examining the type of signal multivariate analysis uncovers at the group level." In: *NeuroImage* 146, pp. 113–120.

Glasser, Matthew F, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, Stephen M Smith, and David C Van Essen (Aug. 2016). "A multi-modal parcellation of human cerebral cortex." en. In: *Nature* 536.7615, pp. 171–178.

Glover, Gary H (2011). "Overview of functional magnetic resonance imaging." In: *Neurosurgery Clinics* 22.2, pp. 133–139.

Goense, Jozien, Yvette Bohraus, and Nikos K Logothetis (2016). "fMRI at high spatial resolution: implications for BOLD-models." In: *Frontiers in computational neuroscience* 10, p. 66.

Goodall, Colin (1991). "Procrustes methods in the statistical analysis of shape." In: *Journal of the Royal Statistical Society: Series B (Methodological)* 53.2, pp. 285–321.

Gordon, Evan M, Timothy O Laumann, Adrian W Gilmore, Dillan J Newbold, Deanna J Greene, Jeffrey J Berg, Mario Ortega, Catherine Hoyt-Drazen, Caterina Gratton, Haoxin Sun, et al. (2017). "Precision functional mapping of individual human brains." In: *Neuron* 95.4, pp. 791–807.

Gorgolewski, Krzysztof J, Gael Varoquaux, Gabriel Rivera, Yannick Schwarz, Satrajit S Ghosh, Camille Maumet, Vanessa V Sochat, Thomas E Nichols, Russell A Poldrack, JeanBaptiste Poline, et al. (2015). "NeuroVault.org: a web based repository for collecting and sharing unthresholded statistical maps of the human brain." In: *Frontiers in neuroinformatics* 9.

Gorgolewski, Krzysztof, Christopher D Burns, Cindee Madison, Dav Clark, Yaroslav O Halchenko, Michael L Waskom, and Satrajit S Ghosh (2011). "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python." In: *Frontiers in neuroinformatics* 5, p. 13.

Gower, John C, Garmt B Dijksterhuis, et al. (2004). *Procrustes problems*. Vol. 30. Oxford University Press on Demand.

Gramfort, Alexandre, Gabriel Peyré, and Marco Cuturi (2015). "Fast Optimal Transport Averaging of Neuroimaging Data." In: *CoRR* abs/1503.08596. arXiv: 1503.08596. URL: http://arxiv.org/abs/1503.08596.

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

Gratton, Caterina, Timothy O Laumann, Ashley N Nielsen, Deanna J Greene, Evan M Gordon, Adrian W Gilmore, Steven M Nelson, Rebecca S Coalson, Abraham Z Snyder, Bradley L Schlaggar, et al. (2018). "Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation." In: *Neuron* 98.2, pp. 439–452.

Greve, Douglas N and Bruce Fischl (2009). "Accurate and robust brain image alignment using boundary-based registration." In: *NeuroImage* 48.1, pp. 63–72. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2009.06.060.

Grosenick, Logan, Brad Klingenberg, Kiefer Katovich, Brian Knutson, and Jonathan E Taylor (2013). "Interpretable whole-brain prediction analysis with GraphNet." In: *NeuroImage* 72, pp. 304–321.

Güçlü, Umut and Marcel A. J. van Gerven (2015). "Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream." In: *Journal of Neuroscience* 35.27, pp. 10005–10014. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.5023-14.2015. eprint: http://www.jneurosci.org/content/35/27/10005.full.pdf. URL: http://www.jneurosci.org/content/35/27/10005.

Guntupalli, J Swaroop, Ma Feilong, and James V Haxby (Apr. 2018). "A computational model of shared fine-scale structure in the human connectome." en. In: *PLoS Comput. Biol.* 14.4, e1006120.

Guntupalli, J Swaroop, Michael Hanke, Yaroslav O Halchenko, Andrew C Connolly, Peter J Ramadge, and James V Haxby (June 2016). "A Model of Representational Spaces in Human Cortex." en. In: *Cereb. Cortex* 26.6, pp. 2919–2934.

Hamilton, Liberty S and Alexander G Huth (2020). "The revolution will not be controlled: natural stimuli in speech neuroscience." In: *Language, Cognition and Neuroscience* 35.5, pp. 573–582.

Hanke, Michael, Nico Adelhöfer, Daniel Kottke, Vittorio Iacovella, Ayan Sengupta, Falko R Kaule, Roland Nigbur, Alexander Q Waite, Florian Baumgartner, and Jörg Stadler (2016). "A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation." In: *Scientific data* 3.1, pp. 1–15.

Hanke, Michael, Florian J Baumgartner, Pierre Ibe, Falko R Kaule, Stefan Pollmann, Oliver Speck, Wolf Zinke, and Jörg Stadler (2014). "A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie." In: *Scientific data* 1.1, pp. 1–18.

Hanke, Michael, Richard Dinga, Christian Häusler, J Swaroop Guntupalli, Michael Casey, Falko R Kaule, and Jörg Stadler (2015). "High-resolution 7-Tesla fMRI data on the perception of musical genres–an extension to the studyforrest dataset." In: *F1000Research* 4.174, p. 174.

Hanke, Michael, Yaroslav O Halchenko, Per B Sederberg, Stephen José Hanson, James V Haxby, and Stefan Pollmann (Jan. 2009). "PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data." en. In: *Neuroinformatics* 7.1, pp. 37–53.

Hardoon, David R, Sandor Szedmak, and John Shawe-Taylor (2004). "Canonical correlation analysis: An overview with application to learning methods." In: *Neural computation* 16.12, pp. 2639–2664.

Hasson, Uri, Yuval Nir, Ifat Levy, Galit Fuhrmann, and Rafael Malach (2004). "Intersubject synchronization of cortical activity during natural vision." In: *science* 303.5664, pp. 1634–1640.

Haxby, James V, Andrew C Connolly, and J Swaroop Guntupalli (2014). "Decoding neural representational spaces using multivariate pattern analysis." In: *Annual review of neuroscience* 37, pp. 435–456.

Haxby, James V, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini (2001). "Distributed and overlapping representations of faces and objects in ventral temporal cortex." In: *Science* 293.5539, pp. 2425–2430.

Haxby, James V, J Swaroop Guntupalli, Andrew C Connolly, Yaroslav O Halchenko, Bryan R Conroy, M Ida Gobbini, Michael Hanke, and Peter J Ramadge (Oct. 2011). "A common, high-dimensional model of the representational space in human ventral temporal cortex." en. In: *Neuron* 72.2, pp. 404–416.

Haxby, James V, J Swaroop Guntupalli, Samuel A Nastase, and Ma Feilong (June 2020). "Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies." In: *Elife* 9, e56601.

Haynes, John-Dylan (2015). "A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives." In: *Neuron* 87.2, pp. 257–270.

Haynes, John-Dylan and Geraint Rees (2006). "Decoding mental states from brain activity in humans." In: *Nature Reviews Neuroscience* 7.7, pp. 523–534.

Heller, Ruth, Yulia Golland, Rafael Malach, and Yoav Benjamini (Nov. 2007). "Conjunction group analysis: An alternative to mixed/random effect analysis." In: *NeuroImage* 37, pp. 1178–85. DOI: 10.1016/j.neuroimage.2007.05.051.

Heun, Reiner, Frank Jessen, Uwe Klose, Michael Erb, DO Granath, Nikolaus Freymann, and Wolfgang Grodd (2000). "Interindividual variation of cerebral activation during encoding and retrieval of words." In: *European psychiatry* 15.8, pp. 470–479.

Hoerl, Arthur E and Robert W Kennard (1970). "Ridge regression: Biased estimation for nonorthogonal problems." In: *Technometrics* 12.1, pp. 55–67.

Holland, Dominic, Joshua M Kuperman, and Anders M Dale (2010). "Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging." In: *Neuroimage* 50.1, pp. 175–183.

Hoyos-Idrobo, Andrés, Gaël Varoquaux, Yannick Schwartz, and Bertrand Thirion (2018). "FReM–scalable and stable decoding with fast regularized ensemble of models." In: *NeuroImage* 180, pp. 160–172.

Huth, Alexander G, Shinji Nishimoto, An T Vu, and Jack L Gallant (2012). "A continuous semantic space describes the representation of thousands of object and action categories across the human brain." In: *Neuron* 76.6, pp. 1210–1224.

Hutton, Chloe, Andreas Bork, Oliver Josephs, Ralf Deichmann, John Ashburner, and Robert Turner (2002). "Image distortion correction in fMRI: a quantitative evaluation." In: *Neuroimage* 16.1, pp. 217–240.

Hyvärinen, Aapo and Erkki Oja (2000). "Independent component analysis: algorithms and applications." In: *Neural networks* 13.4-5, pp. 411–430.

Janati, Hicham, Thomas Bazeille, Bertrand Thirion, Marco Cuturi, and Alexandre Gramfort (2019). "Group level MEG/EEG source imaging via optimal transport: minimum Wasserstein estimates." In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 743–754.

– (2020). "Multi-subject MEG/EEG source imaging with sparse multi-task regression." In: *NeuroImage* 220, p. 116847.

Janati, Hicham, Marco Cuturi, and Alexandre Gramfort (2020). "Debiased sinkhorn barycenters." In: *International Conference on Machine Learning*. PMLR, pp. 4692–4701.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith (2002). "Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images." In: *NeuroImage* 17.2, pp. 825–841. ISSN: 1053-8119. DOI: 10.1006/nimg.2002.1132. URL: http://www.sciencedirect.com/science/article/pii/S1053811902911328.

Jiahui, Guo, Ma Feilong, Matteo Visconti di Oleggio Castello, J Swaroop Guntupalli, Vassiki Chauhan, James V Haxby, and M Ida Gobbini (Aug. 2020). "Predicting individual face-selective topography using naturalistic stimuli." en. In: *Neuroimage* 216, p. 116458.

Jiang, Di, Yuhui Du, Hewei Cheng, Tianzi Jiang, and Yong Fan (2013). "Groupwise spatial normalization of fMRI data based on multi-range functional connectivity patterns." In: *Neuroimage* 82, pp. 355–372.

Kaboli, Mohsen (2017). "A review of transfer learning algorithms." PhD thesis. Technische Universität München.

Kamitani, Yukiyasu and Frank Tong (2005). "Decoding the visual and subjective contents of the human brain." In: *Nature neuroscience* 8.5, pp. 679–685.

Kantorovitch, Leonid (1958). "On the translocation of masses." In: *Management Science* 5.1, pp. 1–4.

Koyejo, Oluwasanmi and Russell A Poldrack (2013). "Decoding cognitive processes from functional MRI." In: *NIPS Workshop on Machine Learning for Interpretable Neuroimaging*, pp. 5–10.

Kriegeskorte, Nikolaus and Jörn Diedrichsen (July 2019). "Peeling the Onion of Brain Representations." In: *Annu. Rev. Neurosci.* 42.1, pp. 407–432.

Kriegeskorte, Nikolaus, Rainer Goebel, and Peter Bandettini (Mar. 2006). "Information-based functional brain mapping." en. In: *Proc. Natl. Acad. Sci. U. S. A.* 103.10, pp. 3863–3868.

Kriegeskorte, Nikolaus, Marieke Mur, and Peter A Bandettini (2008). "Representational similarity analysis-connecting the branches of systems neuroscience." In: *Frontiers in systems neuroscience* 2, p. 4.

Lanczos, C. (1964). "Evaluation of Noisy Data." In: *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis* 1.1, pp. 76–85. ISSN: 0887-459X. DOI: 10.1137/0701007. URL: http://epubs.siam.org/doi/10.1137/0701007.

Langs, Georg, Polina Golland, Yanmei Tie, Laura Rigolo, and Alexandra J Golby (2010). "Functional Geometry Alignment and Localization of Brain Areas." en. In: *Adv. Neural Inf. Process. Syst.* 1, pp. 1225–1233.

Langs, Georg, Danial Lashkari, Andrew Sweet, Yanmei Tie, Laura Rigolo, Alexandra J. Golby, and Polina Golland (2011). "Learning an Atlas of a Cognitive Process in Its Functional Geometry." In: *Information Processing in Medical Imaging - 22nd International Conference, IPMI 2011, Kloster Irsee, Germany, July 3-8, 2011. Proceedings*, pp. 135–146. DOI: 10.1007/978-3-642-22092-0_12. URL: https://doi.org/10.1007/978-3-642-22092-0_12.

Langs, Georg, Andrew Sweet, Danial Lashkari, Yanmei Tie, Laura Rigolo, Alexandra J Golby, and Polina Golland (Dec. 2014). "Decoupling function and anatomy in atlases of functional connectivity patterns: language mapping in tumor patients." en. In: *Neuroimage* 103, pp. 462–475.

Laumann, Timothy O, Evan M Gordon, Babatunde Adeyemo, Abraham Z Snyder, Sung Jun Joo, Mei-Yen Chen, Adrian W Gilmore, Kathleen B McDermott, Steven M Nelson, Nico UF Dosenbach, et al. (2015). "Functional system and areal organization of a highly sampled individual human brain." In: *Neuron* 87.3, pp. 657–670.

Le Bihan, Denis, Jean-François Mangin, Cyril Poupon, Chris A Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat (2001). "Diffusion tensor imaging: concepts and applications." In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 13.4, pp. 534–546.

Liu, Ran, Cem Subakan, Aishwarya H Balwani, Jennifer Whitesell, Julie Harris, Sanmi Koyejo, and Eva L Dyer (2020). "A generative

modeling approach for interpreting population-level variability in brain structure." In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 257–266.

Lorbert, Alexander and Peter J Ramadge (2012). "Kernel hyperalignment." In: *Advances in Neural Information Processing Systems* 25, pp. 1790–1798.

Lu, Qihong, Po-Hsuan Chen, Jonathan W Pillow, Peter J Ramadge, Kenneth A Norman, and Uri Hasson (2018). "Shared representational geometry across neural networks." In: *arXiv preprint arXiv:1811.11684*.

Madan, Christopher R (2021). "Scan Once, Analyse Many: Using large open-access neuroimaging datasets to understand the brain." In: *Neuroinformatics*, pp. 1–29.

Mahowald, Kyle and Evelina Fedorenko (2016). "Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability." In: *Neuroimage* 139, pp. 74–93.

Majerus, Steve, Frédéric Péters, Marion Bouffier, Nelson Cowan, and Christophe Phillips (2018). "The dorsal attention network reflects both encoding load and top–down control during working memory." In: *Journal of Cognitive Neuroscience* 30.2, pp. 144–159.

Mandelkow, Hendrik, Jacco A de Zwart, and Jeff H Duyn (2016). "Linear discriminant analysis achieves high classification accuracy for the BOLD fMRI response to naturalistic movie stimuli." In: *Frontiers in human neuroscience* 10, p. 128.

Mantini, Dante, Uri Hasson, Viviana Betti, Mauro G Perrucci, Gian Luca Romani, Maurizio Corbetta, Guy A Orban, and Wim Vanduffel (2012). "Interspecies activity correlations reveal functional correspondence between monkey and human brain areas." In: *Nature methods* 9.3, pp. 277–282.

Mayberg, Helen S, Mario Liotti, Stephen K Brannan, Scott McGinnis, Roderick K Mahurin, Paul A Jerabek, J Arturo Silva, Janet L Tekell, Charles C Martin, Jack L Lancaster, et al. (1999). "Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness." In: *American journal of psychiatry* 156.5, pp. 675–682.

Mazziotta, John C., Arthur W. Toga, Alan Evans, Peter Fox, and Jack Lancaster (1995). "A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM)." In: *NeuroImage* 2.2, Part A, pp. 89–101. ISSN: 1053-8119. DOI: 10.1006/nimg.1995.1012.

Mensch, Arthur, Julien Mairal, Danilo Bzdok, Bertrand Thirion, and Gaël Varoquaux (2017). "Learning neural representations of human cognition across many fMRI studies." In: *Advances in Neural Information Processing Systems*, pp. 5883–5893.

Meyniel, Florent and Stanislas Dehaene (2017). "Brain networks for confidence weighting and hierarchical inference during probabilistic

learning." In: *Proceedings of the National Academy of Sciences* 114.19, E3859–E3868.

Miller, Michael B., Christa-Lynn Donovan, Craig M. Bennett, Elissa M. Aminoff, and Richard E. Mayer (2012). "Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals." In: *NeuroImage* 59.1. Neuroergonomics: The human brain in action and at work, pp. 83–93. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2011.05.060. URL: https://www.sciencedirect.com/science/article/pii/S1053811911005684.

Monge, Gaspard (1781). "Mémoire sur la théorie des déblais et des remblais." In: *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathéematique et de Physique pour la même année*, pp. 666–704.

Monti, Martin M (2011). "Statistical analysis of fMRI time-series: a critical review of the GLM approach." In: *Frontiers in human neuroscience* 5, p. 28.

Munkres, James (1957). "Algorithms for the assignment and transportation problems." In: *Journal of the society for industrial and applied mathematics* 5.1, pp. 32–38.

Naselaris, Thomas, Emily Allen, and Kendrick Kay (2021). "Extensive sampling for complete models of individual brains." In: *Current Opinion in Behavioral Sciences* 40, pp. 45–51.

Naselaris, Thomas, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant (2011). "Encoding and decoding in fMRI." In: *Neuroimage* 56.2, pp. 400–410.

Nastase, Samuel A, Valeria Gazzola, Uri Hasson, and Christian Keysers (2019). *Measuring shared responses across subjects using intersubject correlation*.

Nastase, Samuel A, Ariel Goldstein, and Uri Hasson (2020). "Keep it real: rethinking the primacy of experimental control in cognitive neuroscience." In: *NeuroImage* 222, p. 117254.

Nastase, Samuel A, Yun-Fei Liu, Hanna Hillman, Kenneth A Norman, and Uri Hasson (2020). "Leveraging shared connectivity to aggregate heterogeneous datasets into a common response space." In: *NeuroImage* 217, p. 116865.

Nenning, KarlHeinz, Hesheng Liu, Satrajit S Ghosh, Mert R Sabuncu, Ernst Schwartz, and Georg Langs (2017). "Diffeomorphic functional brain surface alignment: Functional demons." In: *NeuroImage*.

Ogawa, Seiji, Tso-Ming Lee, Alan R Kay, and David W Tank (1990). "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." In: *proceedings of the National Academy of Sciences* 87.24, pp. 9868–9872.

Olivetti, E., S. Veeramachaneni, and P. Avesani (2011). "Testing for Information with Brain Decoding." In: *2011 International Workshop*

on Pattern Recognition in NeuroImaging, pp. 33–36. DOI: 10.1109/PRNI.2011.14.

Oosterhof, Nikolaas N, Tobias Wiestler, Paul E Downing, and Jörn Diedrichsen (2011). "A comparison of volume-based and surface-based multi-voxel pattern analysis." In: *Neuroimage* 56.2, pp. 593–600.

O'Connor, David, Natan Vega Potler, Meagan Kovacs, Ting Xu, Lei Ai, John Pellman, Tamara Vanderwal, Lucas C Parra, Samantha Cohen, Satrajit Ghosh, et al. (2017). "The Healthy Brain Network Serial Scanning Initiative: a resource for evaluating inter-individual differences and their reliabilities across scan conditions and sessions." In: *Gigascience* 6.2, giw011.

Pajula, Juha and Jussi Tohka (2016). "How many is enough? Effect of sample size in inter-subject correlation analysis of fMRI." In: *Computational intelligence and neuroscience* 2016.

Paquola, Casey, Reinder Vos De Wael, Konrad Wagstyl, Richard AI Bethlehem, Seok-Jun Hong, Jakob Seidlitz, Edward T Bullmore, Alan C Evans, Bratislav Misic, Daniel S Margulies, et al. (2019). "Microstructural and functional gradients are increasingly dissociated in transmodal cortices." In: *PLOS Biology* 17.5, e3000284.

Pauling, Linus and Charles D Coryell (1936). "The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin." In: *Proceedings of the National Academy of Sciences* 22.4, pp. 210–216.

Paus, Tomáš (2013). "How environment and genes shape the adolescent brain." In: *Hormones and behavior* 64.2, pp. 195–202.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Penny, William D, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Peyré, Gabriel and Marco Cuturi (Mar. 2018). "Computational Optimal Transport." In: *arXiv e-prints*, arXiv:1803.00567, arXiv:1803.00567. arXiv: 1803.00567 [stat.ML].

Pike, G Bruce (2012). "Quantitative functional MRI: concepts, issues and future challenges." In: *Neuroimage* 62.2, pp. 1234–1240.

Pinel, Philippe, Bertrand Thirion, Sébastien Meriaux, Antoinette Jobert, Julien Serres, Denis Le Bihan, Jean-Baptiste Poline, and Stanislas Dehaene (2007). "Fast reproducible identification and large-scale databasing of individual functional cognitive networks." In: *BMC neuroscience* 8.1, p. 91.

Pinel, Philippe, Baudouin Forgeot d'Arc, Stanislas Dehaene, Thomas Bourgeron, Bertrand Thirion, Denis Le Bihan, and Cyril Poupon

(2019). "The functional database of the ARCHI project: Potential and perspectives." In: *Neuroimage* 197, pp. 527–543.

Pinho, Ana Luísa, Alexis Amadon, Murielle Fabre, Elvis Dohmatob, Isabelle Denghien, Juan Jesús Torre, Chantal Ginisty, Séverine Becuwe-Desmidt, Séverine Roger, Laurence Laurier, Véronique Joly-Testault, Gaelle Médiouni-Cloarec, Christine Doublé, Bernadette Martins, Philippe Pinel, Evelyn Eger, Gael Varoquaux, Christophe Pallier, Stanislas Dehaene, Lucie Hertz-Pannier, and Bertrand Thirion (2021). "Subject-specific segregation of functional territories based on deep phenotyping." In: *Human Brain Mapping* 42.4, pp. 841–870. DOI: https://doi.org/10.1002/hbm.25189. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.25189. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25189.

Pinho, Ana Luísa, Alexis Amadon, Baptiste Gauthier, Nicolas Clairis, André Knops, Sarah Genon, Elvis Dohmatob, Juan Jesús Torre, Chantal Ginisty, Séverine Becuwe-Desmidt, et al. (2020). "Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping." In: *Scientific Data* 7.1, pp. 1–16.

Pinho, Ana Luísa, Alexis Amadon, Torsten Ruest, Murielle Fabre, Elvis Dohmatob, Isabelle Denghien, Chantal Ginisty, Séverine Becuwe-Desmidt, Séverine Roger, Laurence Laurier, et al. (2018). "Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping." In: *Scientific data* 5.1, pp. 1–15.

Poldrack, Russell A (2017). "Precision neuroscience: Dense sampling of individual brains." In: *Neuron* 95.4, pp. 727–729.

Poldrack, Russell A, Deanna M Barch, Jason Mitchell, Tor Wager, Anthony D Wagner, Joseph T Devlin, Chad Cumba, Oluwasanmi Koyejo, and Michael Milham (2013a). "Toward open sharing of task-based fMRI data: the OpenfMRI project." In: *Frontiers in neuroinformatics* 7, p. 12.

– (2013b). "Toward open sharing of task-based fMRI data: the OpenfMRI project." In: *Frontiers in neuroinformatics* 7, p. 12.

Poldrack, Russell A, Timothy O Laumann, Oluwasanmi Koyejo, Brenda Gregory, Ashleigh Hover, Mei-Yen Chen, Krzysztof J Gorgolewski, Jeffrey Luci, Sung Jun Joo, Ryan L Boyd, et al. (2015). "Long-term neural and physiological phenotyping of a single human." In: *Nature communications* 6.1, pp. 1–15.

Rademacher, J, V S Caviness Jr, H Steinmetz, and A M Galaburda (July 1993). "Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology." en. In: *Cereb. Cortex* 3.4, pp. 313–329.

Richard, Hugo, Luigi Gresele, Aapo Hyvärinen, Bertrand Thirion, Alexandre Gramfort, and Pierre Ablin (2020). "Modeling shared responses in neuroimaging studies through multiview ica." In: *arXiv preprint arXiv:2006.06635*.

Richard, Hugo, Lucas Martin, Ana Luısa Pinho, Jonathan Pillow, and Bertrand Thirion (2019). "Fast shared response model for fMRI data." In: *arXiv preprint arXiv:1909.12537*.

Robinson, Emma C, Kara Garcia, Matthew F Glasser, Zhengdao Chen, Timothy S Coalson, Antonios Makropoulos, Jelena Bozek, Robert Wright, Andreas Schuh, Matthew Webster, et al. (2018). "Multimodal surface matching with higher-order smoothness constraints." In: *Neuroimage* 167, pp. 453–465.

Robinson, Peter N (2012). "Deep phenotyping for precision medicine." In: *Human mutation* 33.5, pp. 777–780.

Rodriguez-Vazquez, Bertha, Laura E Suarez, Golia Shafiei, Ross Markello, Casey Paquola, Patric Hagmann, Martijn Van Den Heuvel, Boris Bernhardt, R Nathan Spreng, and Bratislav Misic (2019). "Gradients of structure-function tethering across neocortex." In: *BioRxiv*, p. 561985.

Rubin, Timothy, Oluwasanmi O Koyejo, Michael N Jones, and Tal Yarkoni (2016). "Generalized correspondence-LDA models (GC-LDA) for identifying functional regions in the brain." In: *Advances in neural information processing systems* 29, pp. 1118–1126.

Sabuncu, M, E Bryan, Peter J Ramadge, and James V Haxby (Jan. 2010). "Function-based intersubject alignment of human cortical anatomy." In: *Cerebral Cortex* 20, pp. 130–140.

Salehi, Mehraveh, Abigail S Greene, Amin Karbasi, Xilin Shen, Dustin Scheinost, and R Todd Constable (2020). "There is no single functional atlas even for a single individual: Functional parcel definitions change with task." In: *NeuroImage* 208, p. 116366.

Sampaio-Baptista, Cassandra, Alexandre A Khrapitchev, Sean Foxley, Theresa Schlagheck, Jan Scholz, Saad Jbabdi, Gabriele C DeLuca, Karla L Miller, Amy Taylor, Nagheme Thomas, et al. (2013). "Motor skill learning induces changes in white matter microstructure and myelination." In: *Journal of Neuroscience* 33.50, pp. 19499–19503.

Saygin, Zeynep M, David E Osher, Elizabeth S Norton, Deanna A Youssoufian, Sara D Beach, Jenelle Feather, Nadine Gaab, John DE Gabrieli, and Nancy Kanwisher (2016). "Connectivity precedes function in the development of the visual word form area." In: *Nature neuroscience* 19.9, pp. 1250–1255.

Schaefer, Alexander, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and B T Thomas Yeo (Sept. 2018). "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI." en. In: *Cereb. Cortex* 28.9, pp. 3095–3114.

Schmitzer, Bernhard (2019). "Stabilized sparse scaling algorithms for entropy regularized transport problems." In: *SIAM Journal on Scientific Computing* 41.3, A1443–A1481.

Schoenauer-Sebag, Alice, Louise Heinrich, Marc Schoenauer, Michele Sebag, Lani F Wu, and Steve J Altschuler (2019). "Multi-domain adversarial learning." In: *arXiv preprint arXiv:1903.09239*.

Schrouff, J., J. M. Monteiro, L. Portugal, M. J. Rosa, C. Phillips, and J. Mourao-Miranda (Jan. 2018a). "Embedding Anatomical or Functional Knowledge in Whole-Brain Multiple Kernel Learning Models." In: *Neuroinformatics* 16.1, pp. 117–143.

Schrouff, Jessica, Julien Cremers, Gaëtan Garraux, Luca Baldassarre, Janaina Mourão-Miranda, and Christophe Phillips (2013). "Localizing and comparing weight maps generated from linear kernel machine learning models." In: *2013 International Workshop on Pattern Recognition in Neuroimaging*. IEEE, pp. 124–127.

Schrouff, Jessica, João M Monteiro, Liana Portugal, Maria J Rosa, Christophe Phillips, and Janaina Mourão-Miranda (2018b). "Embedding anatomical or functional knowledge in whole-brain multiple kernel learning models." In: *Neuroinformatics* 16.1, pp. 117–143.

Seghier, Mohamed L, François Lazeyras, Alan J Pegna, Jean-Marie Annoni, Ivan Zimine, Eugene Mayer, Christoph M Michel, and Asaid Khateb (2004). "Variability of fMRI activation during a phonological and semantic language task in healthy subjects." In: *Human brain mapping* 23.3, pp. 140–155.

Seghier, Mohamed L and Cathy J Price (2018). "Interpreting and utilising intersubject variability in brain function." In: *Trends in cognitive sciences* 22.6, pp. 517–530.

Smith, Stephen M, Christian F Beckmann, Narender Ramnani, Mark W Woolrich, Peter R Bannister, Mark Jenkinson, Paul M Matthews, and David J McGonigle (2005). "Variability in fMRI: A re-examination of inter-session differences." In: *Human brain mapping* 24.3, pp. 248–257.

Sonkusare, Saurabh, Michael Breakspear, and Christine Guo (2019). "Naturalistic stimuli in neuroscience: critically acclaimed." In: *Trends in cognitive sciences* 23.8, pp. 699–714.

Stelzer, Johannes, Gabriele Lohmann, Karsten Mueller, Tilo Buschmann, and Robert Turner (2014). "Deficient approaches to human neuroimaging." In: *Frontiers in Human Neuroscience* 8, p. 462.

Sui, Jing, Tülay Adali, Godfrey D Pearlson, and Vince D Calhoun (2009). "An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques." In: *Neuroimage* 46.1, pp. 73–86.

Sun, Xiaofei, Lin Shi, Yishan Luo, Wei Yang, Hongpeng Li, Peipeng Liang, Kuncheng Li, Vincent CT Mok, Winnie CW Chu, and Defeng Wang (2015). "Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions." In: *Biomedical engineering online* 14.1, pp. 1–17.

Tavor, Ido, O Parker Jones, Rogier B Mars, SM Smith, TE Behrens, and Saad Jbabdi (2016). "Task-free MRI predicts individual differences

in brain activity during task performance." In: *Science* 352.6282, pp. 216–220.

Thirion, Bertrand, Guillaume Flandin, Philippe Pinel, Alexis Roche, Philippe Ciuciu, and Jean-Baptiste Poline (2006a). "Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets." In: *Human brain mapping* 27.8, pp. 678–693.

– (Aug. 2006b). "Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fMRI datasets." en. In: *Hum. Brain Mapp.* 27.8, pp. 678–693.

Thirion, Bertrand, Philippe Pinel, Sébastien Mériaux, Alexis Roche, Stanislas Dehaene, and Jean-Baptiste Poline (2007). "Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses." In: *Neuroimage* 35.1, pp. 105–120.

Thirion, J-P (1998). "Image matching as a diffusion process: an analogy with Maxwell's demons." In: *Medical image analysis* 2.3, pp. 243–260.

Thompson, Paul M, Ole A Andreassen, Alejandro Arias-Vasquez, Carrie E Bearden, Premika S Boedhoe, Rachel M Brouwer, Randy L Buckner, Jan K Buitelaar, Kazima B Bulayeva, Dara M Cannon, et al. (2017). "ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide." In: *Neuroimage* 145, pp. 389–408.

Tom, Sabrina M, Craig R Fox, Christopher Trepel, and Russell A Poldrack (2007). "The neural basis of loss aversion in decision-making under risk." In: *Science* 315.5811, pp. 515–518.

Torre, Juan Jesús, Ana Luísa Pinho, Swetha Shankar, Alexis Amadon, Mani Saignavongs, Marcela Perrone-Bertolotti, Thomas Bazeille, Elvis Dohmatob, Isabelle Denghien, Chantal Ginisty, Séverine Becuwe-Desmidt, Séverine Roger, Yann Lecomte, Valérie Berland, Laurence Laurier, Véronique Joly-Testault, Gaelle Mediouni-Cloarec, Christine Doublé, Bernadette Martins, Jean-Philippe Lachaux, Patrick Bisset, Ayse Zeynep Enkavi, Ian Eisenberg, Russel Poldrack, Roberto Santoro, Elia Formisano, Gael Varoquaux, Stanislas Dehaene, Lucie Hertz-Pannier, and Bertrand Thirion (in preparation). "Individual Brain Charting fourth dataset extension, high-resolution fMRI data for cognitive mapping."

Turner, Benjamin O, Erick J Paul, Michael B Miller, and Aron K Barbey (2018). "Small sample sizes reduce the replicability of task-based fMRI studies." In: *Communications Biology* 1.1, pp. 1–10.

Tustison, Nicholas J, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee (2010). "N4ITK: improved N3 bias correction." In: *IEEE transactions on medical imaging* 29.6, pp. 1310–1320.

Van Essen, David C., Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, Kamil Ugurbil, and for the WU-Minn HCP Consortium (2013). "The WU-Minn Human Connectome Project: An overview." In: 80, pp. 62–79. DOI: 10.1016/j.neuroimage.2013.

05.041. URL: https://app.dimensions.ai/details/publication/pub.1022076283andhttp://europepmc.org/articles/pmc3724347?pdf=render.

Varoquaux, Gaël, Alexandre Gramfort, Fabian Pedregosa, Vincent Michel, and Bertrand Thirion (July 2011). "Multi-subject dictionary learning to segment an atlas of brain spontaneous activity." In: *Information Processing in Medical Imaging*. Vol. 6801. Lecture Notes in Computer Science. Gábor Székely, Horst Hahn. Kaufbeuren, Germany: Springer, pp. 562–573. DOI: 10.1007/978-3-642-22092-0_46. URL: https://hal.inria.fr/inria-00588898.

Varoquaux, Gaël, Pradeep Reddy Raamana, Denis A Engemann, Andrés Hoyos-Idrobo, Yannick Schwartz, and Bertrand Thirion (2017). "Assessing and tuning brain decoders: cross-validation, caveats, and guidelines." In: *NeuroImage* 145, pp. 166–179.

Varoquaux, Gaël, Sepideh Sadaghiani, Philippe Pinel, Andreas Kleinschmidt, Jean-Baptiste Poline, and Bertrand Thirion (2010). "A group model for stable multi-subject ICA on fMRI datasets." In: *Neuroimage* 51.1, pp. 288–299.

Varoquaux, Gaël, Yannick Schwartz, Russell A Poldrack, Baptiste Gauthier, Danilo Bzdok, Jean-Baptiste Poline, and Bertrand Thirion (2018). "Atlases of cognition with large-scale human brain mapping." In: *PLoS computational biology* 14.11, e1006565.

Vergara, Rodrigo C, Sebastián Jaramillo-Riveri, Alejandro Luarte, Cristóbal Moënne-Loccoz, Rómulo Fuentes, Andrés Couve, and Pedro E Maldonado (2019). "The energy homeostasis principle: neuronal energy regulation drives local network dynamics generating behavior." In: *Frontiers in computational neuroscience* 13, p. 49.

Virtanen, Pauli, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. (2020). "SciPy 1.0: fundamental algorithms for scientific computing in Python." In: *Nature methods* 17.3, pp. 261–272.

Wang, He, Jinping Sun, Dong Cui, Xin Wang, Jingna Jin, Ying Li, Zhipeng Liu, and Tao Yin (2021). "Quantitative assessment of inter-individual variability in fMRI-based human brain atlas." In: *Quantitative Imaging in Medicine and Surgery* 11.2, p. 810.

Wang, Weiran, Raman Arora, Karen Livescu, and Jeff Bilmes (2015). "On deep multi-view representation learning." In: *International conference on machine learning*. PMLR, pp. 1083–1092.

Welvaert, Marijke and Yves Rosseel (2013). "On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data." In: *PloS one* 8.11, e77089.

Woo, Choong-Wan, Luke J Chang, Martin A Lindquist, and Tor D Wager (2017). "Building better biomarkers: brain models in translational neuroimaging." In: *Nature neuroscience* 20.3, p. 365.

Wu, Anqi, Oluwasanmi Koyejo, and Jonathan W Pillow (2019). "Dependent relevance determination for smooth and structured sparse regression." In: *J. Mach. Learn. Res.* 20.89, pp. 1–43.

Wu, Anqi, Samuel A Nastase, Christopher A Baldassano, Nicholas B Turk-Browne, Kenneth A Norman, Barbara E Engelhardt, and Jonathan W Pillow (2021). "Brain kernel: a new spatial covariance function for fMRI data." In: *bioRxiv*.

Xu, Hao, Alexander Lorbert, Peter J Ramadge, J Swaroop Guntupalli, and James V Haxby (2012). "Regularized hyperalignment of multi-set fMRI data." In: *2012 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, pp. 229–232.

Xu, Ting, Karl-Heinz Nenning, Ernst Schwartz, Seok-Jun Hong, Joshua T Vogelstein, Alexandros Goulas, Damien A Fair, Charles E Schroeder, Daniel S Margulies, Jonny Smallwood, et al. (2020). "Cross-species functional alignment reveals evolutionary hierarchy within the connectome." In: *Neuroimage* 223, p. 117346.

Xu, Tonglin, Muhammad Yousefnezhad, and Daoqiang Zhang (2018a). "Gradient Hyperalignment for multi-subject fMRI data alignment." In: *Pacific Rim International Conference on Artificial Intelligence*. Springer, pp. 1058–1068.

– (2018b). "Gradient hyperalignment for multi-subject fmri data alignment." In: *Pacific Rim International Conference on Artificial Intelligence*. Springer, pp. 1058–1068.

Yeo, BT Thomas, Mert R Sabuncu, Tom Vercauteren, Nicholas Ayache, Bruce Fischl, and Polina Golland (2009). "Spherical demons: fast diffeomorphic landmark-free surface registration." In: *IEEE transactions on medical imaging* 29.3, pp. 650–668.

Yousefnezhad, Muhammad and Daoqiang Zhang (2017). "Deep Hyperalignment." In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Curran Associates, Inc., pp. 1604–1612. URL: http://papers.nips.cc/paper/6758-deep-hyperalignment.pdf.

Zhang, Y., M. Brady, and S. Smith (2001). "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm." In: *IEEE Transactions on Medical Imaging* 20.1, pp. 45–57. ISSN: 0278-0062. DOI: 10.1109/42.906424.

université PARIS-SACLAY

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

Titre : Estimation de modèle après des alignements arbitraires : application à l'imagerie médicale

Mots clés : Imagerie médicale, Cartographie cérébrale, Transport optimal, Estimation statistique

Résumé : Un objectif important des neurosciences cognitives est de comprendre l'organisation fonctionnelle du cerveau. Pour cela, elles s'appuient sur l'Imagerie par Résonance Magnétique fonctionnelle (IRMf), un outil puissant pour étudier le lien entre les fonctions cérébrales et les structures anatomiques sous-jacentes à une haute résolution spatiale. La variabilité fonctionnelle interindividuelle est un obstacle majeur qui limite la précision de la cartographie fonctionnelle du cerveau et la généralisation des résultats obtenus par les études d'imagerie cérébrale. Cette variabilité importante, observable entre des sujets effectuant la même tâche cognitive, va bien au-delà des variations de taille et de forme des structures anatomiques. Dans cette thèse, nous nous intéresserons à un ensemble de méthodes conçues pour traiter la variabilité fonctionnelle : l'alignement fonctionnel. Ces méthodes mettent en correspondances les signaux cérébraux de différents sujets sur la base de leur similarité fonctionnelle.

Dans un premier temps, nous présenterons les concepts et les techniques usuels pour la cartographie fonctionnelle cérébrale, ainsi que les difficultés induites par la variabilité fonctionnelle. Nous passerons également en revue les méthodes d'alignement fonctionnel existantes. Dans une deuxième partie, nous proposerons une nouvelle méthode d'alignement fonctionnel, basée sur le transport optimal—une théorie mathématique qui s'intéresse à la mise en correspondance de distributions de probabilité prenant en compte leur géométrie. Par ailleurs, les méthodes d'alignement fonctionnel sont appliquées localement, et de nombreux alignements locaux doivent être agrégés pour construire des alignements à l'échelle du cerveau entier. De plus, ces mises en correspondances sont définies entre deux sujets et il est nécessaire, pour aligner plusieurs sujets, de recourir à un « modèle fonctionnel de groupe » : une représentation commune sur laquelle tous les sujets peuvent être alignés. Cependant, les solutions proposées dans la littérature présentent de nombreux inconvénients et nous proposerons donc une méthode alternative d'agrégation ainsi qu'un algorithme pour concevoir des modèles de groupe. La troisième partie sera consacrée à une validation expérimentale approfondie de ces méthodes. En effet, l'alignement est rarement utilisé dans des études applicatives et sa capacité à compenser la variabilité fonctionnelle en pratique est peu documentée. Nous évaluerons sa capacité à améliorer la généralisation de modèles prédictifs à de nouveaux sujets. Dans cette configuration—dite de « décodage inter-sujet »—appliquée à quatre jeux de données, nous montrerons que les méthodes d'alignement ont le potentiel de compenser une part importante de la variabilité interindividuelle.

Title: Template estimation for arbitrary alignments: application to brain imaging

Keywords: Medical Imaging, Statistics, Brain Mapping, Optimal transport, Estimation

Abstract: An important goal of cognitive neurosciences is to understand the functional organization of the brain. It heavily relies on Functional Magnetic Resonance Imaging (fMRI), a powerful tool to investigate the link between brain function and anatomical structures at a high spatial-resolution. Functional inter-individual variability is a major obstacle limiting functional brain mapping precision and generalizability of results obtained in neuroimaging studies. This variability, observed across subjects performing the same task, goes far beyond anatomical variability in brain shape and size. In this work, we focus on a class of methods designed to address functional variability, namely functional alignment. These methods match subjects neural signals based on their functional similarity.

In a first part, we review standard functional brain mapping paradigms and techniques, as well as the challenges induced by functional variability. We additionally review existing functional alignment methods and related work, and discuss the current limitations of these approaches. In a second part, we develop a new functional alignment method, based on optimal transport—a mathematical theory interested in matching probability distributions while taking their geometry into account. Functional alignment methods are local, which means that many local alignments need to be aggregated to compose whole-brain alignments. Moreover, these methods derive pairwise matching and call for a "functional template", a common functional representation to which all subjects of a study can be aligned. To overcome limitations of existing solutions, we additionally introduce a new aggregation scheme as well as a principled template design procedure. In a third part, we turn to empirical validation of alignment performance. Indeed, these methods are seldom used in applied studies, and it is unclear to what extent they can address functional variability in typical cognitive studies. We investigate their performance to improve generalization of predictive models to new subjects. In this inter-subject decoding set-up, spanning four different datasets, we show that alignment methods hold real potential to recover an important share of prediction accuracy lost due to inter-subject variability.

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