



**HAL**  
open science

# Brain topography beyond parcellations: local gradients of functional maps

Elvis Dohmatob, Hugo Richard, Ana Luísa Pinho, Bertrand Thirion

## ► To cite this version:

Elvis Dohmatob, Hugo Richard, Ana Luísa Pinho, Bertrand Thirion. Brain topography beyond parcellations: local gradients of functional maps. *NeuroImage*, 2021, pp.117706. 10.1016/j.neuroimage.2020.117706 . hal-03122173v1

**HAL Id: hal-03122173**

**<https://inria.hal.science/hal-03122173v1>**

Submitted on 26 Jan 2021 (v1), last revised 19 May 2021 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Brain topography beyond parcellations: local gradients of functional maps

Elvis Dohmatob<sup>a,b</sup>, Hugo Richard<sup>a</sup>, Ana Luísa Pinho<sup>a</sup>, Bertrand Thirion<sup>a</sup>

<sup>a</sup>Inria, CEA, Université Paris-Saclay, Saclay, France

<sup>b</sup>Criteo AI Lab

---

## Abstract

Functional neuroimaging provides the unique opportunity to characterize brain regions based on their response to tasks or ongoing activity. As such, it holds the premise to capture brain spatial organization. Yet, the conceptual framework to describe this organization has remained elusive: on the one hand, parcellations build implicitly on a piecewise constant organization, i.e. flat regions separated by sharp boundaries; on the other hand, the recently popularized concept of functional gradient hints instead at a smooth structure. Noting that both views converge to a topographic scheme that pieces together local variations of functional features, we perform a quantitative assessment of local gradient-based models. Using as a driving case the prediction of functional Magnetic Resonance Imaging (fMRI) data—concretely, the prediction of task-fMRI from rest-fMRI maps across subjects—we develop a parcel-wise linear regression model based on a dictionary of reference topographies. Our method uses multiple random parcellations—as opposed to a single fixed parcellation—and aggregates estimates across these parcellations to predict functional features in left-out subjects. Our experiments demonstrate the existence of an optimal cardinality of the parcellation to capture local gradients of functional maps.

---

## Highlights

- Assessing the existence of local gradients in brain organization calls for multi-contrast analyses, such as task- *vs.* rest-fMRI
- Using about 200 parcels yields highest accuracy for local linear rest-to-task map prediction
- Combining results from multiple parcellations improves model accuracy
- Local linear models outperform whole-brain non-linear models
- Motor contrasts are less well predicted from resting-state activity than high-level contrasts.

## Keywords

Parcellation, functional mapping, prediction, model selection, functional gradients

## 1. Introduction

While *functional Magnetic Resonance Imaging* (fMRI) has, for a long time, been restricted to the role of providing a location-sensitive information of task effects from cognitive neuroscience experiments, it is now increasingly used as a marker for brain functional organization. fMRI contrast maps can indeed reveal the functional signature of brain regions (Genon et al., 2018), which is in turn linked to both connectivity and cytoarchitectural properties at the same areas (Cohen et al., 2008; Saygin et al., 2011; Glasser et al., 2016; Eickhoff et al., 2018a). An increasingly popular conceptualization refers to the continuous

topography that is inherent to these signatures as *functional gradients* (Margulies et al., 2016).

In this context, the concept of gradients requires further clarification: from a mathematical point of view, the gradient of a function is an operator that determines in which direction the values of the function *locally* vary most strongly. In the context of brain imaging, mapping *local gradients* thus amounts to studying how brain features vary spatially around a given location. Indeed, while sharp spatial transitions between regions have been reported in the literature (e.g. in Cohen et al. (2008)), other studies have emphasized smooth variations, a.k.a. local gradients (Xu et al., 2016b), that can also be described as a property of the underlying brain connectivity (Jbabdi et al., 2013). However the concept of functional gradients, as put forward in Margulies et al. (2016); Huntenburg et al. (2018), is instead a *global* or *large-scale* property. In these studies, key regions or networks, like the default-mode and the executive-control networks, are called *transmodal*. They exhibit some properties that make them maximally dissimilar from other regions referred to as *unimodal*, namely the primary-sensory and the motor regions. Therefore, the concept of gradient, used in this context, is justified by the smooth spatial interpolation in the characteristics of brain regions between the *poles*. Alternative conceptualizations of large-scale brain organization can be found in e.g. Toro and Burnod (2003) and Mesmoudi et al. (2013).

In this paper, we consider the question of *local gradients*, namely whether the local variability of brain characteristics represents a meaningful feature. Accounting for such variations is known to be important in primary sen-

sory regions (see e.g. [Schwarzkopf et al. \(2011\)](#)). Yet, it is not clear whether such description is quantitatively relevant in regions associated with higher-order cognitive domains. In order to assess empirically the significance of local gradients, we propose to frame the question as a learning problem: what do we learn by observing such local variations? In other words, what prediction can one make, upon the observation of these local variations? Concretely, we propose to test whether some *information* can be carried from a modality or a contrast to another ([Amunts et al., 2014](#)). In the present study, we thus propose to target the prediction of one type of fMRI data from another. Indeed, fMRI data come in two main instances: (1) task-fMRI, that maps the response to experimental conditions according to the classic psychological perspective; and (2) rest-fMRI, that maps instead functional similarities and the interplay between different regions through differences in the magnitude of their correlation ([Bzdok et al., 2015, 2016](#)). The feasibility of performing either cross-contrast or cross-modal fMRI-data predictions is supported by previous studies, e.g. the reconstruction of task-fMRI contrast maps from other contrast maps ([Thirion et al., 2014b](#)) or from resting-state signal components ([Tavor et al., 2016](#)). Such predictive frameworks are very useful, because they provide a sensitive way to capture the underlying structure of the signal, even in the presence of noise in the data.

There are several indications that local variations of brain activity constitute a meaningful feature. The fine topographic information conveyed by both task- and rest-fMRI opens the possibility to functionally characterize individuals through *fingerprints* ([Finn et al., 2015](#)). The recent surge of interest on the information conveyed by topographic maps ([Bijsterbosch et al., 2018](#)) indeed underlines the importance to capture local gradients accurately. Since both task- and rest-fMRI data can be used for that purpose, one shall consider the shared information carried by these fMRI-based modalities, as first observed in [Smith et al. \(2009\)](#). The analysis of the consistency between task- and rest- patterns was noticed in [Biswal et al. \(1995\)](#), where it actually served to demonstrate that the information carried by resting-state signals was consistent with that conveyed by task-fMRI.

Interestingly, most of the recent advances in the use of task-fMRI and rest-fMRI have been formulated under the framework of brain parcellations. The homogeneity of the information conveyed by these fMRI-based modalities is used to divide and further analyze the brain into functionally consistent territories ([Thirion et al., 2006](#); [Cohen et al., 2008](#); [Yeo et al., 2011](#); [Glasser et al., 2016](#); [Schaefer et al., 2017](#); [Eickhoff et al., 2018b](#)). It is noteworthy that brain parcellations are very useful, as they provide data-adaptive dimension reductions that are necessary to compare brain features, estimate connectivity or decode brain activity (see e.g. [Schaefer et al. \(2017\)](#); [Dadi et al. \(2019\)](#); [Mensch et al. \(2017\)](#)). Nevertheless, their capacity to faithfully represent the data is limited, because par-

celling the brain into sub-regions assumes that the corresponding signals are piece-wise constant. Contrariwise, the concept of *local gradient* hints at a smooth structure featuring the spatial transitions between the territories. Therefore, it is timely to enrich parcel-based description with a model of signal variation, which corresponds to a *local gradient* model.

Local topographic information is strikingly subject-dependent, hence one has to define a correspondence model between subjects. In the context of spatial signal components derived from resting-state fMRI functional alignment is often performed using e.g. a dual-regression approach ([Nickerson et al., 2017](#)). Other methods based either on the *hyperalignment* ([Haxby et al., 2011a](#)) or on the *shared response model* ([Chen et al., 2015](#)), also capture inter-individual differences through linear mappings. While it is well-known that data-derived parcellations are more effective than atlases to capture brain signals ([Dadi et al., 2020](#)), there is no consensus on the definition of population-level parcellations ([Thirion et al., 2014a](#)), nor on the appropriate number of parcels. Averaging predictions over parcellations — or equivalently, marginalizing the effect of parcellations — is an attractive alternative to mitigate the impact of arbitrary parcellation choices [Da Mota et al. \(2013\)](#); [Hoyos-Idrobo et al. \(2018\)](#).

In summary, the joint analysis of rest- and task-fMRI offers the appropriate setting to combine the benefit of parcellations (local models) and gradient approaches (local contrasts). Yet, this conceptual framework requires some quantitative assessment:

- How to efficiently compute the individual maps of functional topographies from different fMRI-based modalities, in particular from rest-fMRI?
- How does a global non-linear rest-to-task model compare with a collection of local linear models (see [Fig.1\(a\)](#))?
- What are the tradeoffs in terms of spatial resolution or number of parcels in regard to mapping task-based contrasts from task-related activity (see [Fig.1\(b\)](#))?
- Is there any loss incurred by relying on one given parcellation or should one combine the results obtained from several parcellations, instead (see [Fig.1\(c\)](#))?

*Contributions.* To address these questions, we present herein three main contributions: 1) We propose a computationally efficient framework for group models of rest-fMRI. Topographic information is obtained by Dictionary learning, an alternative to independent components analysis (ICA). The approach then relies on dual regression to draw correspondences across individuals ([Nickerson et al., 2017](#)), but also relates to the popular shared response model for hyperalignment ([Chen et al., 2015](#)). 2) Among a range of parcellation cardinalities, corresponding to different resolutions, we identify an optimal regime of resolution for parcellation (about 200 parcels) and show that it is more accurate than non-linear neural network predictor fitting

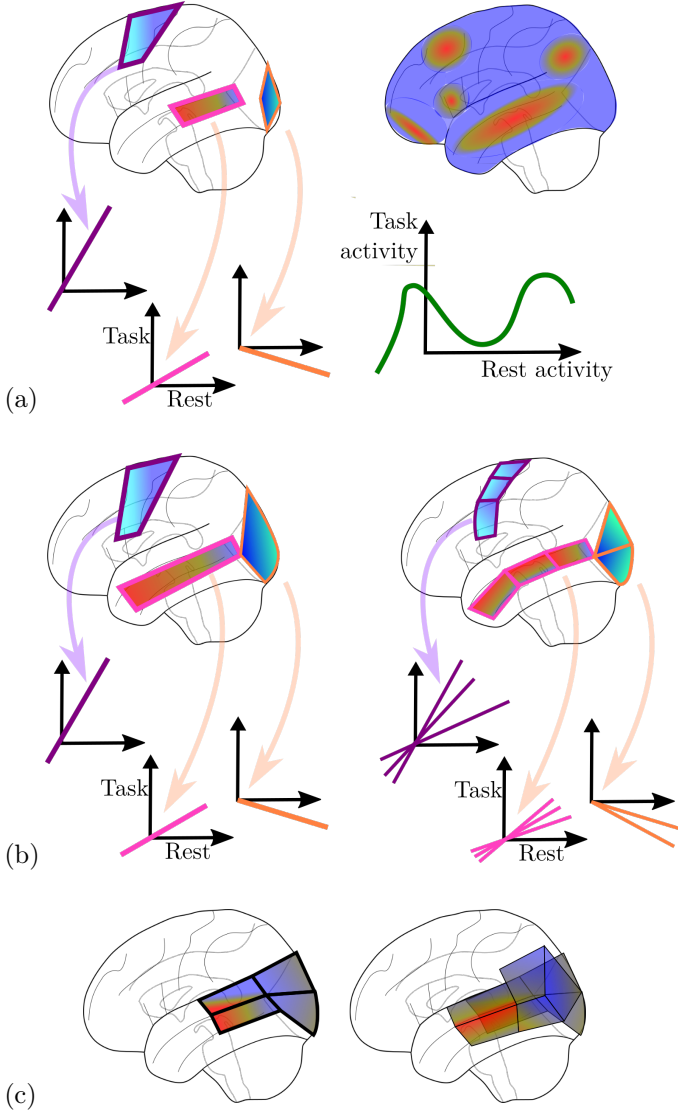


Figure 1: **Challenges in the definition of local gradients of brain function:** Do local linear models, a.k.a. gradients (a - left), capture the joint variation of task-fMRI and rest-fMRI features as well as —or better than— a brain-wide non-linear model that would relate these two types of data (a - right)? (b) What is the relevant resolution of such local-gradient models? and, thus, what is the required number of parcels to describe these local variations? (c) Is a unique parcellation sufficient to capture the local-joint variations of task- and rest-fMRI features, or should one consider instead multiple and overlapping parcellations to capture these variations?

whole-brain data. 3) We show that aggregating results obtained from different parcellations reduces prediction variance and, thus, enhances predictive power. The gains are particularly important for models learned on small groups of subjects but they are also observed on larger samples, i.e. hundreds of subjects. This means that this approach is useful in neuroimaging studies, most of which are based on limited cohorts. Our results are obtained from the *Human Connectome Project* (HCP) dataset as well as from a smaller-scale dataset named *Individual Brain Charting*. Altogether they show the gains brought by this approach in a small-sample dataset.

## 2. Methods

We start by describing the modeling approach to capture local gradients. Data and experiments are presented in section 3.

### 2.1. Setting

*Notations.* We denote matrices with bold capital letters, vectors with bold small letters, scalars or indexes with standard small letters. For any matrix  $\mathbf{A}$ ,  $\|\mathbf{A}\|_{\text{Fro}}$  will denote its Frobenius norm (the square root of the sum of its squared coefficients).  $\langle \cdot, \cdot \rangle$  denotes the scalar product between vectors or matrix multiplication.

*The data.* We consider a setting in which rest- and task-fMRI datasets are observed in a populations of subjects, as in Shafto et al. (2014); van Essen et al. (2012); Gordon et al. (2017). We denote the rest-fMRI datasets  $\mathbf{X}_s \in \mathbb{R}^{p \times n_s}$ , as  $p \times n_s$  matrices, where  $s$  indexes the subjects,  $p$  is the number of voxels in the brain and  $n_s$  is the number of 3D volumes of a run. Task-fMRI datasets are assumed to be in the form of statistical parametric maps, displaying some contrasts of interest. For a given subject  $s$ , we denote them  $\mathbf{Y}_s \in \mathbb{R}^{p \times c}$ , where  $c$  is the number of task-related contrasts studied.

### 2.2. Deriving Resting-state topographies

The problem amounts to defining local mappings  $f : \mathbf{X}_s \rightarrow \mathbf{Y}_s$  that are common across subjects. This is very ill-posed *a priori*, given that resting state signals are random in time and, hence, have no consistency across individuals. We resort to the standard multivariate fMRI hypothesis, namely the fact that these signals are the random and time-varying expression of stable structures, called *brain networks*:  $\mathbf{X}_s \approx \mathbf{D}_s \mathbf{C}_s + \text{noise}$ , where  $\mathbf{D}_s \in \mathbb{R}^{p \times k}$  are individual topographies or networks of interest, while  $\mathbf{C}_s$  encodes the random time course of the subjects resting-state activity. The first challenge is thus to estimate  $\mathbf{D}_s$  from  $\mathbf{X}_s$  in a consistent manner across individuals, namely ensuring that the first component of  $\mathbf{D}_s$  corresponds to the same network across all subjects, and so on. For this, we first estimate a common set of topographic components  $\mathbf{D}$  across the population using a dictionary model, then we

adapt this population-level set of topographies to each individual by dual regression, obtaining the individual counterpart  $\mathbf{D}_s$  of  $\mathbf{D}$ . The whole procedure is summarized in Fig. 2. We refer the reader to the Appendix for more details about the derivation of this model: Appendix A.1 presents the generative model underlying the analysis procedures that we present; Appendix A.2 yields the derivation of the  $\mathbf{D}$  group-level topographies; and Appendix A.3 yields the derivation of corresponding individual topographies with dual regression. We also relate this algorithm to alternative approaches based on the shared response model and hyperalignment.

### 2.3. The link function: local linear models

We now develop our model for predicting subject-specific activation maps  $\mathbf{Y}_s$  from resting-state dictionaries  $\mathbf{D}_s$ . The aim is to learn a simple local model, one per region, to predict task-evoked activation from the topographic information captured from resting-state time courses. These local linear models are meant to capture *local gradients* of brain activity. These region-specific predictive models are limited to the class of linear models. There are various advantages with this approach: beyond the fact that linear models are interpretable, they can be combined easily in ensembles by simply averaging the coefficients. For each training subject  $s$ , a linear model  $\langle \cdot, \mathbf{W}_s \rangle$  is fitted for predicting  $\mathbf{Y}_s$  from  $\mathbf{D}_s$ , i.e.  $\mathbf{Y}_s \approx \mathbf{D}_s \mathbf{W}_s$ . For each subject, the features  $\mathbf{D}_s$  are normalized to have zero mean and unit variance over all voxels. This implicitly handles between-subjects differences, and facilitates *transfer-learning* from one subject to another at test time.

The underlying regions are obtained from a parcellation of the brain (see Thirion et al. (2014a) for a review of data-driven parcellation methods in fMRI analysis). For each parcel  $\mathcal{P}_m$  of a parcellation  $\mathcal{P}$ , we solve a multi-output ridge regression problem:

$$\underset{\mathbf{W} \in \mathbb{R}^{k \times c}}{\text{minimize}} \frac{1}{2} \|\mathbf{Y}_s|_{\mathcal{P}_m} - \mathbf{D}_s|_{\mathcal{P}_m} \mathbf{W}\|_{\text{Fro}}^2 + \frac{1}{2} \lambda \|\mathbf{W}\|_{\text{Fro}}^2, \quad (1)$$

where  $\mathbf{D}_s|_{\mathcal{P}_m}$  denotes the spatial features of subject  $s$  limited to voxels in parcel  $\mathcal{P}_m$ . Each **voxel** of the brain is assigned to exactly one parcel.

### 2.4. The learning algorithm

The estimation problem (1) is solved in each parcel and training subject to obtain the coefficients for predicting each individual subject map for the functional contrasts considered. Moreover, as there is no unique way to parcel the brain, the algorithm can be run on different parcellations concurrently; these different estimators are then aggregated to improve prediction accuracy. The computational procedure is described in Algorithm 1 in Appendix A.4. We emphasize that this is a parallel algorithm: the computation is performed independently across parcels and, thus, it becomes trivial to run a parallel algorithm and benefit from many-cores machines.

### 2.5. Inference / prediction algorithm

At prediction time, these different models are applied on held-out subjects and their results are aggregated by averaging. Such a “divide-and-conquer” approach allows one to learn complementary aspects of the topographic information of the data, reducing the variance of the individual component models. This is a well-known statistical property of bagging ensembles (see e.g. Breiman (1996); Hoyos-Idrobo et al. (2018)). The inference can be done by making a single pass of Algorithm 2 in Appendix A.4.

## 3. Data and Experiments

### 3.1. Setup

Our experiments were conducted on task-fMRI data of 500 subjects from the HCP dataset (van Essen et al., 2012; Barch et al., 2013). This dataset was developed as an attempt to assess fundamental cognitive domains that sample the diversity of neural systems, namely language processing (including semantic and phonological processing), working-memory, social cognition, relational processing, motor responses, reward and emotion processing. We used the preprocessed data from the HCP500 release.

*Task-fMRI data.* We consider task activation maps obtained from General Linear Models (GLMs) (Friston et al., 1994) that show the activation of each brain voxel to different functional contrasts, for each subject. For instance, the working memory task refers to 19 functional activation maps –each containing  $p \sim 2 \times 10^5$  voxels– per subject. These maps correspond to diverse combinations of the initial 8 conditions of that task. For each subject  $s$ , this gives an output matrix  $\mathbf{Y}_s \in \mathbb{R}^{p \times c}$ , where  $c$  is the total number of contrasts considered. In our experiments, we considered all HCP tasks, giving a total of  $c = 47$  functional contrasts. We used the fixed-effects maps across RL and LR runs.

*Rest-fMRI data.* Rest-fMRI acquired in the same subjects consist of 4 sequences  $n_s = 1200$  3D volumes of  $p = 2 \times 10^5$  voxels per volume, for very subject, forming an  $p \times n_s$  matrix  $\mathbf{X}_s$ . The information extraction described in section 2.2 was then applied to transform each  $\mathbf{X}_s$  into low-dimensional functional connectivity features  $\mathbf{D}_s \in \mathbb{R}^{p \times k}$ , with  $k = 100$ .

Note that the dual regression was performed on the four runs within each subject, and the resulting individual components were then averaged across these four estimates. This is meant to mitigate the possible impact of noise per run (LR/RL phase encoding-related distortions etc.).

*Procedure.*  $N_{\text{train}} = 200$  subjects were used in Algorithm 1 to fit an ensemble of models. As a baseline method, we used parcellations in which each parcel comprised about 4000 voxels, for a total of about 60 parcels. The parcellation was obtained using Ward algorithm with spatial

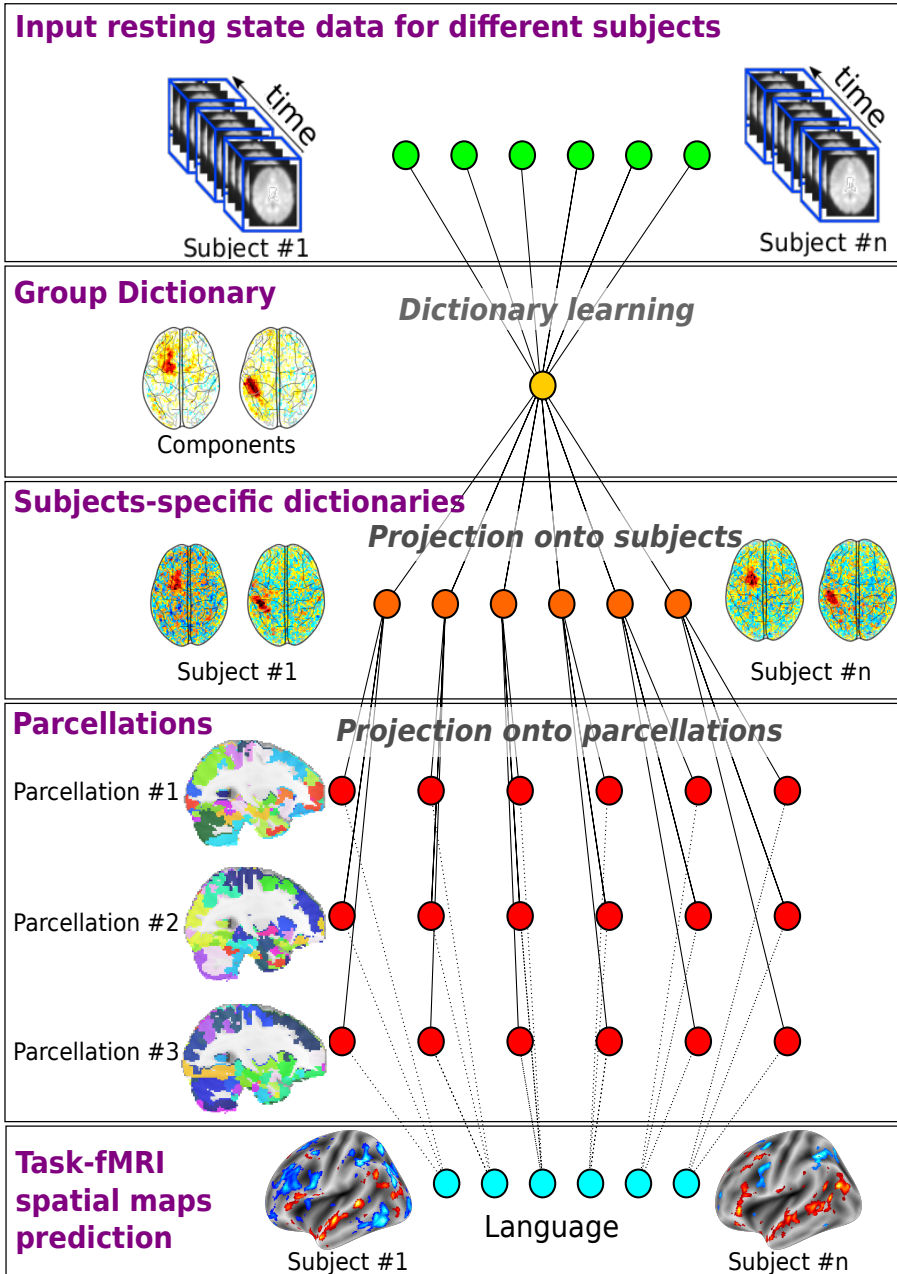


Figure 2: **Linking resting-state topographic information with contrast maps** Resting-state data from each subject  $\mathbf{X}_s \in \mathbb{R}^{p \times n_s}$  are pulled into a dictionary-learning algorithm that estimates reference spatial components  $\mathbf{D} \in \mathbb{R}^{p \times k}$ , which in turn form the linear basis of the core brain networks. These spatial maps are then *projected* on each subjects space via spatio-temporal regression. The output are  $k$  spatial maps  $\mathbf{D}_s^1, \dots, \mathbf{D}_s^k \in \mathbb{R}^p$  encoding each of the  $p$  voxels of in each subject as  $k$  coefficients in a common space. For each subject  $s$ , these spatial maps  $\mathbf{D}_s$  are linked to the voxel-wise task-evoked activation map  $\mathbf{Y}_s$  through a predictive model, by fitting a parcel-wise linear regression model.

connectivity constraints (Thirion et al., 2014a). Note that the number of parcels is a user-specified parameter, hence it can vary freely. While this algorithm is deterministic, i.e. it returns a fixed parcellation for a given input dataset (in the present case, the input dataset is rest-fMRI data), some randomness can be introduced by varying the input data, e.g. by Bootstrap (Hoyos-Idrobo et al., 2018).  $N_{\text{test}} = 200$  subjects were held out for evaluating the models predictions, computed via Algorithm 2.

### 3.2. Individual Brain Charting (IBC) dataset

We reproduced the analysis on the IBC dataset. Although this dataset has a limited number of participants (twelve) data from many tasks were collected for each of them, thus providing a comprehensive coverage of cognitive functions. The task-fMRI data related to this dataset are described in Pinho et al. (2018); they are available on NeuroVault, under the collection 4438 (<https://neurovault.org/collections/4438>). In total, 53 functional contrasts were studied, namely those listed in Pinho et al. (2018). Unlike HCP, these contrasts are independent; they include the tasks of the HCP protocol, plus another battery assessing brain responses for mental arithmetics, spatial abilities, social cognition, emotional responses and a *rapid-serial-visual-presentation* language localizer.

The dataset also comprises resting-state data, with 4 acquisitions obtained on a 3T Siemens Prisma scanner. The imaging parameters for resting-state fMRI are different from those of task-fMRI: 2.2mm (vs 1.5mm) isotropic resolution, Multiband factor of 6 (vs 3), TR=.76s (vs 2.0s); the phase encoding direction was AP then PA. Four runs were acquired per subject in two sessions (one AP and one PA acquisition per session). The duration of each rest-fMRI data run is 14 minutes 12 seconds, amounting to 1120 scans. Acquisition details for task-fMRI data are available in Pinho et al. (2018) and a detailed documentation of the data is available at <https://project.inria.fr/IBC/data>. For the sake of the analysis presented here, all images (task and rest) were resampled to 3mm resolution after motion estimation, motion correction, coregistration to the individual anatomical scan and spatial normalization to MNI space. The preprocessing steps were performed with the SPM12 software, called through Nipype and Pypreprocess, using the routines described in Pinho et al. (2018). These data were preprocessed as follows: band-pass filtering to the  $[.01, .1]Hz$  range, 5mm smoothing, centering and standardization of the resulting time series. Raw data for this dataset are available in <https://openneuro.org/datasets/ds002685>. Here we use a leave-one-subject-out cross-validation due to the limited number of subjects.

### 3.3. Experiments

First, we describe quantitatively the results of task-fMRI maps reconstruction. We then introduce quantitative assessment of prediction accuracy, which allows to

compare reconstruction quality across contrasts, then across parcellations (with a varying number of parcels). Using these quantitative tools, we study the impact of aggregating results from multiple parcellations and compare local linear models with a global (brain-wide) non-linear model, i.e. with a multi-layer perception (MLP). For each training subject, such a model is fitted. The input to the MLP is the subject’s spatial maps  $\mathbf{X}_s$ . As voxels are samples in this learning problem, there are  $2.2 \times 10^5$  samples per subject. The input dimension is  $k = 100$  (the number of spatial components in the dictionary). There are respectively two hidden layers of shape  $100 \times 50$  and  $50 \times c$  per task, where  $c \geq 1$  is the number of contrasts in the task (e.g the HCP Language task yields 3 contrasts –STORY, MATH, STORY-MATH– and so has  $c = 3$ ).

*Prediction accuracy.* To quantify the performance of the proposed model, we resort to three statistics,  $R_{\text{contrast}}^2$ ,  $R_{\text{whole brain}}^2$  and  $R_{\text{voxel}}^2$ . Let us denote  $\mathcal{S}$  the set of test subjects. First, we can compute, per contrast and voxel, a score summarizing the goodness of individual fit between the predicted  $\hat{\mathbf{Y}}_s(c)$  (via Algorithm 2) and the true task-activation map  $\mathbf{Y}_s(c)$  for a given functional contrast  $c$ :

$$R_{\text{contrast}}^2(c) := 1 - \frac{\sum_{s \in \mathcal{S}} \|\hat{\mathbf{Y}}_s(\cdot, c) - \mathbf{Y}_s(\cdot, c)\|_{\text{Fro}}^2}{\sum_{s \in \mathcal{S}} \|\mathbf{Y}_s(\cdot, c) - \bar{\mathbf{Y}}_s(\cdot, c)\|_{\text{Fro}}^2},$$

where  $\bar{\mathbf{Y}}_s$  is the mean map of voxel-wise functional activation across test subjects. Second, we compute for each test subject the whole-brain  $R^2$ -score (across voxels), and average it across test subjects:

$$R_{\text{whole brain}}^2(c) := \text{mean}_{s \in \mathcal{S}} \left( 1 - \frac{\|\hat{\mathbf{Y}}_s(\cdot, c) - \mathbf{Y}_s(\cdot, c)\|_{\text{Fro}}^2}{\|\mathbf{Y}_s(\cdot, c) - \bar{\mathbf{Y}}_s(\cdot, c)\|_{\text{Fro}}^2} \right)$$

Third, we can measure, for each voxel  $v$ , the accuracy of the prediction across contrasts, averaged across subjects:

$$R_{\text{voxel}}^2(v) := \text{mean}_{s \in \mathcal{S}} \left( 1 - \frac{\|\hat{\mathbf{Y}}_s(v, \cdot) - \mathbf{Y}_s(v, \cdot)\|_{\text{Fro}}^2}{\|\mathbf{Y}_s(v, \cdot) - \bar{\mathbf{Y}}_s(v, \cdot)\|_{\text{Fro}}^2} \right)$$

In short,  $R_{\text{contrast}}^2$  represents the goodness of fit of the model for a given contrast and voxel,  $R_{\text{whole brain}}^2$  represents the goodness of fit of the model for a given contrast, while  $R_{\text{voxel}}^2$  represent the capacity of the model to fit the signal at a given location across contrasts. The maximum possible value for each  $R^2$  statistic is 1, indicating perfect prediction of brain activation. Negative values indicate predictions worse than those of the oracle that returns the mean activity in that voxel (note that this mean value is unknown, and can be different between training and test data). This explains why these  $R^2$  statistics can be negative.

*Permutation test.* To assess the significance of the statistics, it is possible to perform a non-parametric test: repeating the computation  $B$  times after randomizing the relationship between  $\mathbf{X}_s$  and  $\mathbf{Y}_s$ , and then comparing the

observed value to the distribution of the  $B$  randomized statistics. We did this for  $R^2_{\text{whole brain}}$ , to obtain a baseline accuracy of the per-contrast fit. This was corrected for multiple comparison by using the  $T_{max}$  approach (Westfall and Young, 1993): tabulating the distribution of the maximum statistic across contrasts for  $B$  randomizations, and comparing the actual values to a quantile of the distribution.

### 3.4. Tools used

The experiments are based on custom scripts that rely on the Nilearn library (v. 0.5.0) for the core structures and visualization tools Abraham et al. (2013b). The Ridge regression and multi-layer perceptron used as task-to-rest fitting models are those of Scikit-learn v. 0.20 Pedregosa et al. (2011). Note that the RidgeCV estimator with default parameters was used, meaning that the regularization parameter was set by nested 5-fold cross-validation, among an initial grid of 3 values [0.1, 1, 10]. Experiments were run initially under Python 2.7 on Linux workstations and later reproduced under Python 3.5. Analysis code is available under the following repository: [http://github.com/hbp-brain-charting/public\\_analysis\\_code/tree/master/papers\\_scripts/gradients](http://github.com/hbp-brain-charting/public_analysis_code/tree/master/papers_scripts/gradients)

## 4. Results

### 4.1. Qualitative metrics

*Predictions capture inter-subject variability.* A successful prediction of subject-specific task-evoked activation maps should topographically match the signs and magnitudes of the subject’s true activation maps, thus recovering individual functional topography in spite of inter-subject variability. In Fig. 3, for a couple of subjects and functional contrasts, we plot the true activation maps against the predicted activation maps. We can see that the predicted activation maps systematically match the true maps very well, which qualitatively reproduces the main findings in Tavor et al. (2016). We outline some ROIs for the *STORY-MATH* contrast showing that individual topographic features are indeed recovered by the estimation procedure.

*Prediction accuracy as a region-specific, rather than task-specific characteristic.* Activation maps (as obtained with a GLM) usually highlight only a restricted subset of the brain that is implicated in the cognitive function under study. These areas can be activated or deactivated above chance, which is reflected in high absolute statistical values (e.g.  $|z| \geq 3$ ).

Is prediction accuracy task-specific, i.e. does the prediction of activation maps from resting-state features perform well in voxels that are on average activated in all subjects? Or is it region-specific, i.e. does a given region tend to be better predicted across contrasts? Figure 4 shows that  $R^2_{\text{contrast}}$  prediction accuracy is actually tied to regions, with mild dependence upon the task. In particular, there is no strong statistical relationship between the

mean activation score and prediction accuracy maps. We observe in general a higher explained variance in parietal-frontal regions across contrasts, extending into the dorsal pathway and lateral occipito-temporal regions. Contrariwise, somato-sensory and motor regions are not well predicted by the model; this is particularly the case for motor regions (see the left-hand contrast of Figure 3).

### 4.2. Quantitative evaluation

*Improved prediction with multiple parcellations.* Fig. 5 shows the across-subject average  $R^2_{\text{whole brain}}$ -score of the prediction (with error-bars over different test subjects) per functional contrast. We can see that prediction-accuracy is variable both across subjects and functional contrasts, capturing between 0% and 50% of the signal. These values are significant at  $p < 0.05$  level, corrected for multiple comparisons, as per a permutation test (the corresponding  $R^2_{\text{whole brain}}$  being 0.005). The most striking feature is the relative weakness of the motor tasks in that prediction; this hints at a specific case related to task/rest activity relationships, that is further discussed in section 5.

In Fig. 6, we plot  $R^2_{\text{whole brain}}$  scores for predictions on test subjects as a function of the training sample size, i.e. the number of subjects in the training set, that goes from 1 to 300 in steps of 10. We do so with a family of models, ranging from a model based on a single parcellation (reddest curve) to models aggregating an increasing number of parcellations (up to 20 parcellations). We clearly see that using multiple parcellations improves performance, in a fashion that is not negligible with respect to the average scores obtained (see e.g. Fig. 8). Indeed, accuracy gains are particularly pronounced in the small-sample regime and they diminish with increasing sample size. For a sample size of  $N = 31$  subjects—the typical median size of fMRI studies—the gain is more than 3%. Yet, when  $N$  grows to hundreds, it remains superior or equal to 1.5%.

*Best predicted regions.* We further studied which brain regions were best fitted by the predictive model across all contrasts. We obtained this information, but considering the cross-subject and cross-contrast average of  $R^2_{\text{voxel}}$  maps. The result is shown in Fig. 7: the Intra-Parietal Sulcus bilaterally, pre-SMA, lateral occipital regions around MT/V5 cortex bilaterally, and bilaterally, spots in the superior frontal Gyrus, medial frontal Gyrus and insula. This topography does not correspond with the first functional gradient of Margulies et al. (2016), that outlined the default mode network. Instead, it outlines the visual pathways and dorsal/ventral attention networks, also known as executive control and salience networks. The corresponding  $R^2_{\text{voxel}}$  map for the IBC dataset (see below) is displayed in Figure B.10 in the appendix.

### 4.3. Model complexity

*Bias-variance tradeoff.* For the experiments described in this section, the training set to learn the reference model



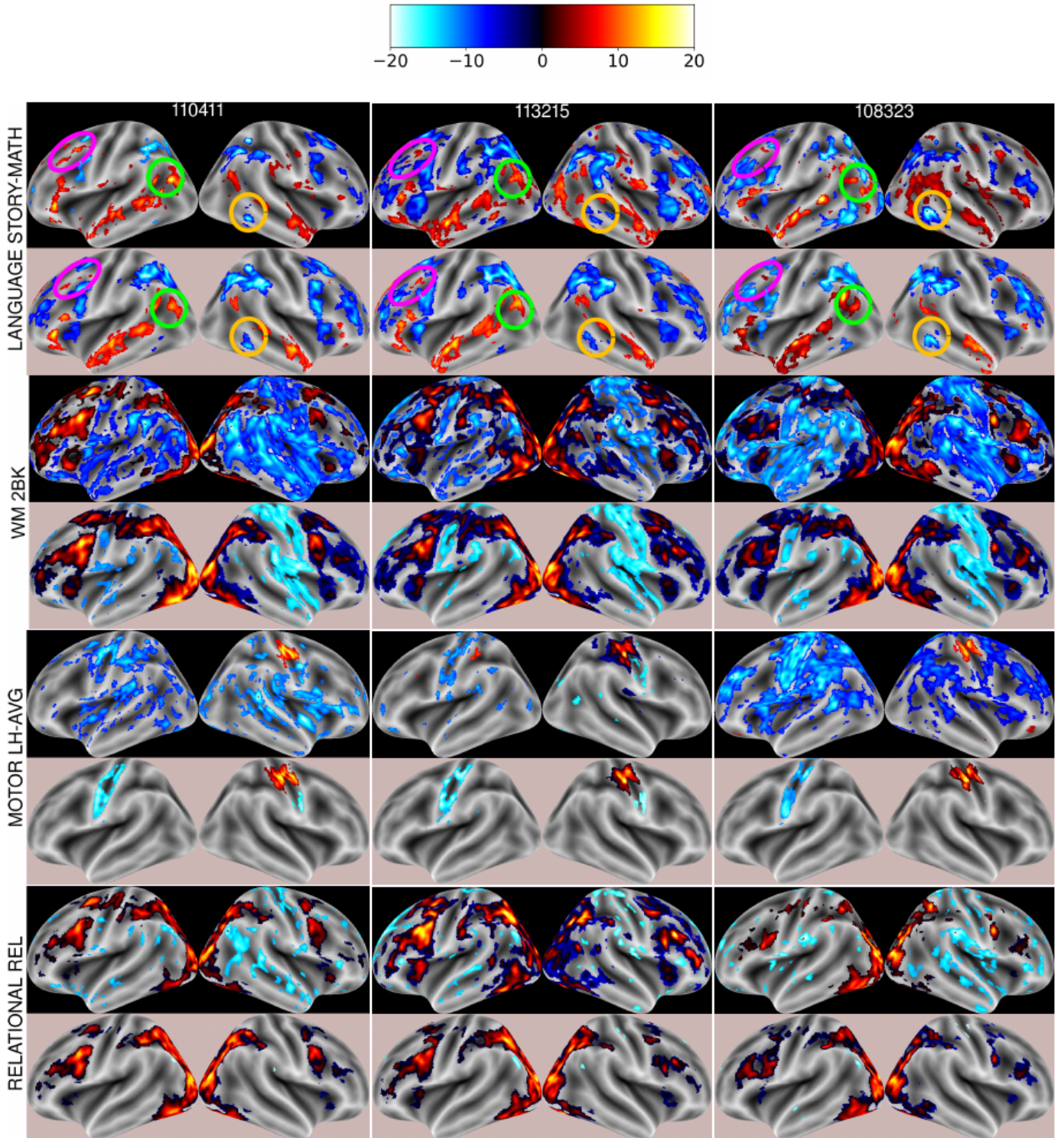
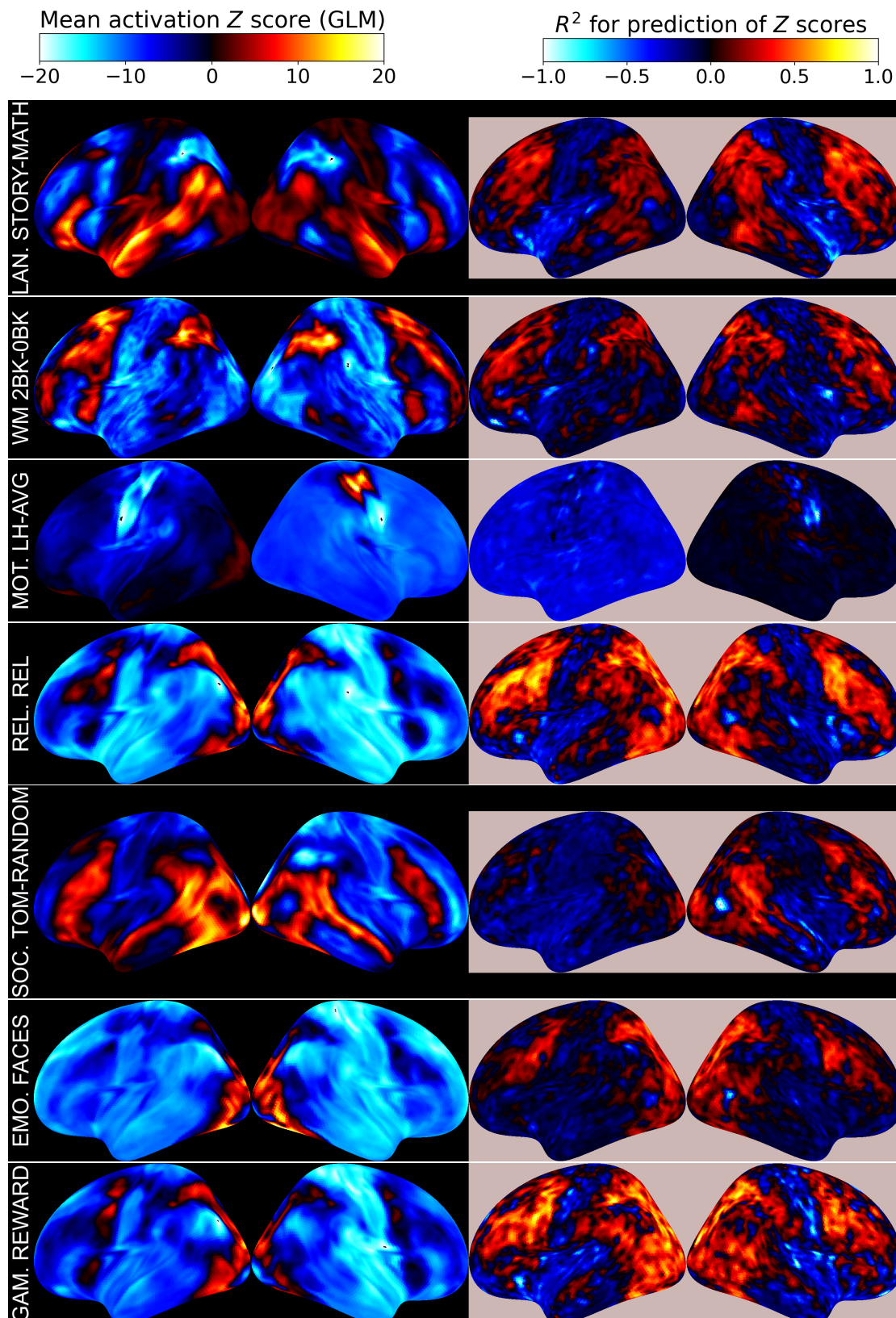


Figure 3: **True vs. predicted contrast maps.** Each other row is the true activation  $\mathbf{z}$  map (black background) for a particular subject with the predicted map shown under (misty background). Each row corresponds to a different contrast. We see that the predicted maps are close to the subject-specific map both in magnitude, sign, and topography. For sake of clarity, only a few contrasts are displayed. On the first row, 3 regions are outlined to draw the attention on some individual patterns that are well-recovered by the predictive model. These results are obtained from  $N_{train} = 200$  subjects, using a parcellation with 256 parcels.

Figure 4: **Which voxels are the most predictable per contrast?** For each functional contrast, we plot the map of mean  $z$  scores (black background). We also plot  $R^2_{contrast}$ , comparing the prediction true values and predicted values of the  $z$  scores in the voxel across all the test subjects (misty background). One can see that the model performs best on voxels that are actively involved in the underlying functional process (high  $R^2_{contrast}$ ), while “noise” voxels are not predicted above chance. For sake of clarity, only a few contrasts are displayed here. These results are obtained from  $N_{train} = 200$  subjects, using a parcellation with 256 parcels.



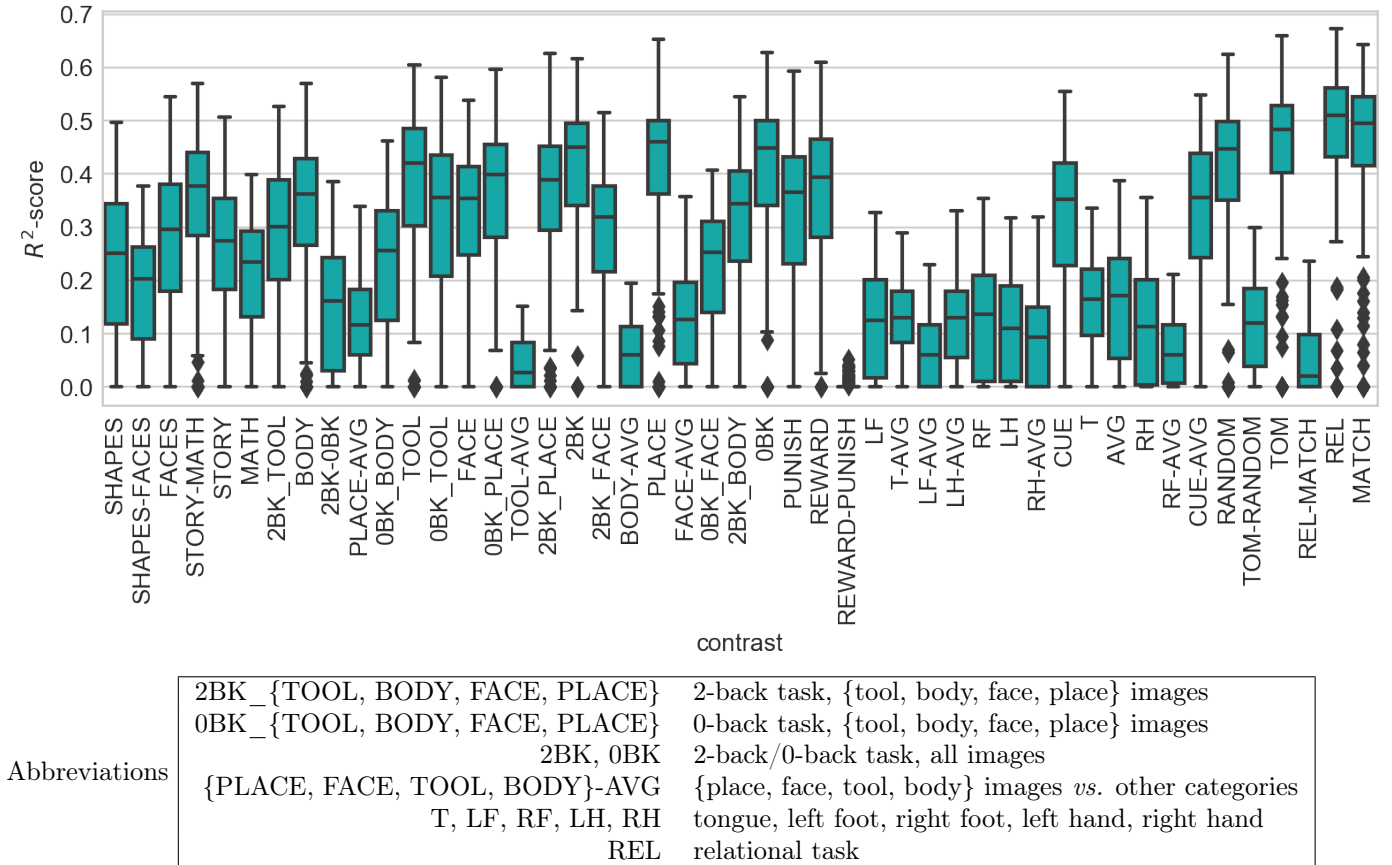


Figure 5: **Prediction scores for different functional contrasts.** We can see that the subject-specific activation in response to some functional contrasts  $R^2_{\text{whole brain}}$  is predicted well-above chance in average, across brain regions. Up to 50% of the signal can indeed be captured by resting-state gradients. The level of success per-contrast may be indicative of either the signal-to-noise ratio of the corresponding map or the clear specification of the functional contrast. These results are obtained from  $N_{\text{train}} = 200$  subjects, using a parcellation with 256 parcels. The  $p < 0.05$  multiplicity-corrected  $R^2$  score is .005. Note that the values displayed here are not directly comparable to those displayed in Fig. 4, as the underlying  $R^2$  statistics do not rely on the same baseline.

is reduced to  $N_{\text{train}} = 60$  subjects. This is because model selection is a very expensive process that requires a large number of model fits. Doing it on the full set would be prohibitive, and it is likely that the amount of data from 60 subjects represents a sufficient sampling of variability in brain structure and organization to allow unbiased inference. In our proposed model, the estimator is piecewise linear, i.e. there is one linear estimator trained on each parcel of a given parcellation. Thus, the number of parcels directly controls the complexity of the model; as the number of parcels is increased, the overall model gains in capacity (degrees of freedom). This additional flexibility allows some local specificity in mapping connectivity gradients to functional gradients. Since the number of voxels per parcel and, thus, the amount of data available to fit the model vary inversely with the total number of parcels used, one can anticipate that increasing the number of parcels can become detrimental beyond a certain level. There are two reasons for this collapse: *i*) with a lot of small parcels, the number of voxels is too small to correctly estimate the model (1); *ii*) small parcel models are more sensitive to between subject-variability, given that

the parcellation scheme used here is fixed across subjects.

Using the average predictive accuracy across contrasts and voxels  $\text{mean}_{c \in C} R^2_{\text{whole brain}}$  as a metric for model selection, we show in Figure 8 (left) that this model selection curve reaches an optimal value: 200 parcels yields the highest prediction accuracy. Figure 8 (middle) is more explicit since it reports relative accuracy values without the subject effect, thus displaying tighter spread around the mean behavior.

*Benchmark against a non-linear model fitted on whole-brain.* We show the gain in  $R^2_{\text{whole brain}}$  score compared to a multi-layer perceptron (MLP) or neural network, i.e. a non-linear predictor, fitted in the *whole brain*. Fig. 8 (middle) shows that the non-linear model performs well, but is outperformed by the parcel-wise linear model in a wide parameters region ( $20 < k < 1000$  parcels).

#### 4.4. Reproduction on a smaller scale dataset

We eventually reproduced the analysis on the IBC dataset ( $N=12$ ). We assessed *i*) whether significantly accurate predictions were possible on small groups of subjects; and

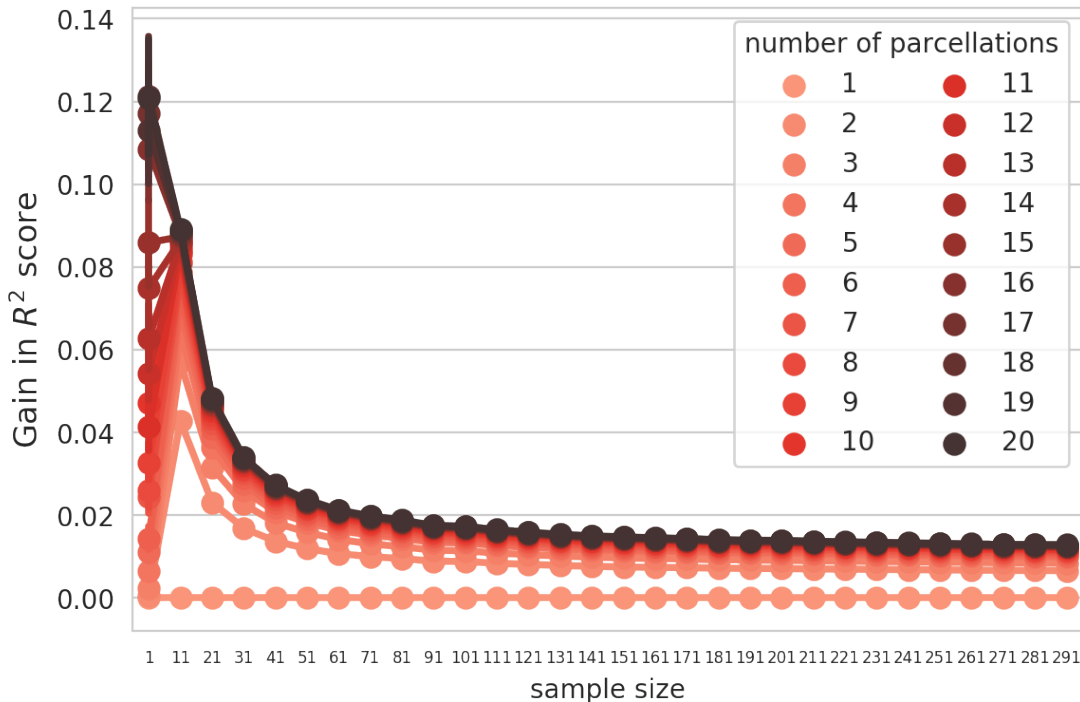


Figure 6: **Prediction accuracy gains brought by multiple parcellations.** The curves display the difference between the accuracy obtained using a variable number of parcellations and then averaging the predictions, compared to using a single fixed parcellation (baseline method). We observe that using multiple parcellations tends to improve performance. The corresponding gains (1.5% to 10%) are non-negligible in front of the average accuracy reported in Fig. 8. This means that the prediction problem is dominated by variance, such as that induced by physiological and acquisition noise, which is reduced by bagging. Accordingly, the gain brought by multiple atlases is particularly pronounced in the small-sample regime and it diminishes as the sample size increases.

*ii*) whether the multiple parcellation scheme proposed in this article was beneficial in this particular prediction. This was indeed the case for all contrasts studied in Fig. 9(top):  $R^2_{\text{whole brain}}$  was better than chance, and significant ( $p < 0.05$ , FWER-corrected) for almost all contrasts. The impact of averaging is displayed for three representative tasks (among 12) in Figure 9(bottom): averaging across multiple parcellations systematically improves the outcome.

## 5. Discussion

It is well known that, to a large extent, function can be ascribed to brain territories, in particular to cortical areas. The concept of *local functional gradients* leads us one step further, namely that spatial variations of functional properties are an important feature. The relevance of such variations is evidenced by their consistency across fMRI-based modalities (Jbabdi et al., 2013; Glasser et al., 2016; Margulies et al., 2016). However, the analysis of subtle variations faces the low signal-to-noise ratio of functional imaging and, hence, it requires a well-suited analytical framework that is powerful enough to overcome noise, and has a clear metric to assess model goodness-of-fit. Here, we choose multivariate prediction and focus on the problem

of predicting task-fMRI maps from resting-state topographies (Tavor et al., 2016).

Based on these premises, we have contributed several improvements in regard to the previously proposed procedure:

- First, we have found that such prediction is successful, not only in large populations of subjects as in Tavor et al. (2016), but also in a smaller group of 12 subjects. Reaching more than 50% explained variance in some regions is indeed a noticeable performance, that brings us close to the noise ceiling of the data, i.e. the maximum amount of signal that can be captured by deterministic models. Even 25% explained variance in out-of-samples settings is already an impressive performance in the realm of brain imaging (see e.g. Elliott et al. (2020)). This successful prediction spans many different functional contrasts and brain regions. It is visible through qualitative adjustments of brain maps to the particular subject under consideration, as shown in Fig.3. Importantly, it is particularly strong for regions supporting high-level functions like dorsal-attentional and executive networks.
- Second, in the framework of local linear predictions, we clearly outlined a bias/variance tradeoff in the choice of the parcellation used to generate local

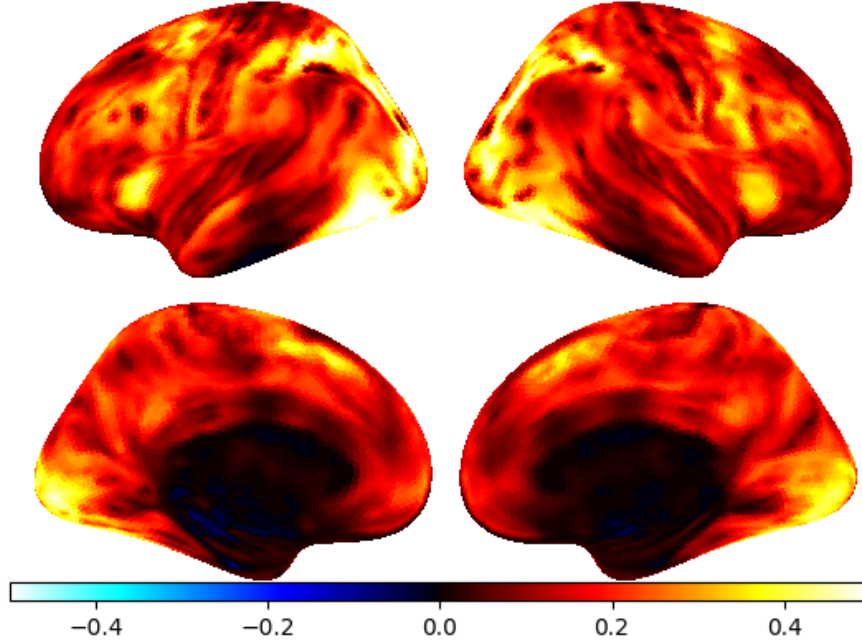


Figure 7: **Overall prediction accuracy per voxel (across contrasts) on the HCP dataset.** This map displays  $R^2_{\text{voxel}}$  statistic obtained from the HCP dataset. Some brain regions that are overall best fitted by the model: the Intra-Parietal Sulcus bilaterally, pre-SMA, lateral occipital regions around MT/V5 cortex bilaterally, and bilaterally, spots in the superior frontal Gyrus, medial frontal Gyrus and insula. This topography does not correspond with that of the Default Mode Network; instead, it outlines the visual pathways and dorsal/ventral attention networks. These results are obtained from  $N_{\text{train}} = 200$  subjects, using a parcellation with 256 parcels.

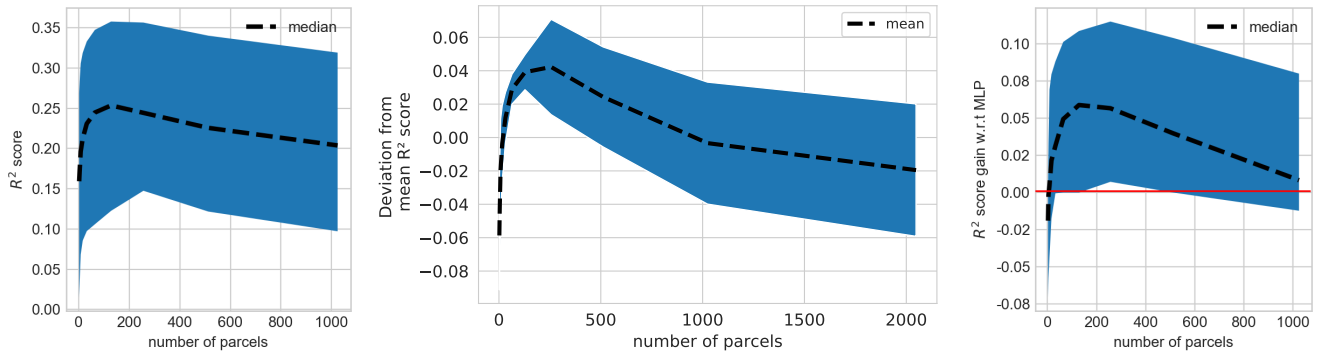


Figure 8: **Analysis of Model complexity.** **Left:** Absolute value of the mean accuracy score  $\text{mean}_{c \in C} R^2_{\text{whole brain}}$  across subjects, as a function of the parcellation cardinality. As the number of parcels directly controls the complexity of the overall model, the curve shows a standard bias/variance compromise, with an optimal region near 200 parcels. **Middle:** Relative prediction accuracy as a function of the number of regions; this curves displays tighter confidence intervals, as the random subject effect is removed. **Right:** Gain in  $R^2$  score compared to an MLP fitted on the entire brain (i.e without parcellation); in its optimal region, the piece-wise linear model outperforms the global non-linear model. All the results were obtained using  $N_{\text{train}} = 60$  subjects.

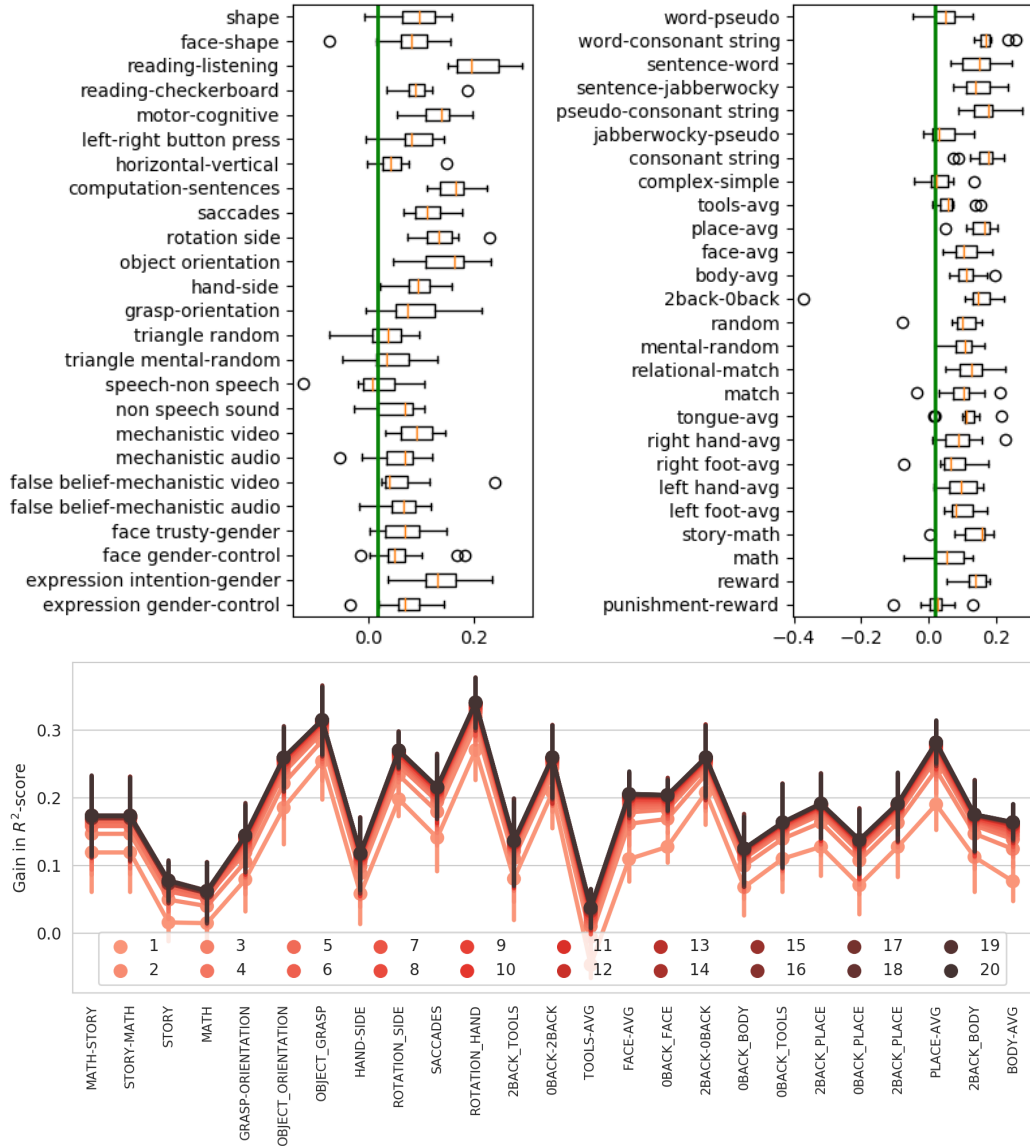


Figure 9: **Reproduction on a small-scale study.** We reused the same modeling approach in the IBC dataset comprising 12 subjects. Prediction-accuracy results, measured as a ratio of explained variance, are displayed on the top for the 12 tasks in [Pinho et al. \(2018\)](#). The vertical green bar indicates the across-contrast-corrected  $p < 0.05$  performance obtained by a permutation test. The effect of increasing the number of parcellations is illustrated for three tasks in the lower panel: a language-localizer (the *RSVP-Language* task in [Pinho et al. \(2018\)](#)), working-memory task, and spatial-localizer task, named *ARCHI Spatial*. Error-bars represent cross-validation splits. We observe that accurate results are obtained for almost all conditions of all tasks, and that averaging across multiple parcellations systematically improves the outcome.

model instances. The simplicity and the computational efficiency of the linear model lead to the clear result of approximately 200 regions. Moreover, in that regime, the locally linear model outperforms a whole-brain non-linear model, obtained by a multi-layer perceptron, confirming the relevance of localized variation model to capture distributed activity patterns.

- Third, we introduced for the first time, in this context, a multiple-parcellation estimator and showed that it improves prediction accuracy and reduces variance. For a sample size of  $N=11$ , simply bagging results from different parcellations brings an increase of 0.07 in average  $R^2$ . For  $N=31$ , i.e. the typical median size of neuroimaging studies, the gain remains above 0.03, which is not negligible. It is thus especially useful in the low-sample regime, which is well illustrated in the IBC-dataset use case with few subjects.

Regarding the comparison with the Multi-Layer Perceptron (MLP), let us point out that the model has been grossly optimized for the numbers of units but not for all architecture aspects (number of layers etc.), as performing such optimization on a given dataset unavoidably results in overfit. The performance of complex non-linear models is a topic of investigation in the field [He et al. \(2020\)](#), but the current consensus is that noisy and sample-limited data used in neuroimaging may not benefit as much from such models as e.g. computer vision applications.

*On quantitative analysis of local gradients.* Recent studies have embraced the goal to describe individualized brain features ([DiNicola et al., 2020](#); [Gordon et al., 2017](#); [Braga et al., 2019](#); [Pinho et al., 2018](#)). Nevertheless, the present study departs from those by taking a quantitative point of view, in order to establish the significance of the organization described in front of the noise level. Indeed, the low test-retest reproducibility of fMRI-derived brain maps ([Elliott et al., 2020](#)) warns against purely qualitative description of brain organization that are not backed up by quantitative arguments. The ultimate demonstration that local functional gradients capture some key brain-organization feature is that they *predict features that were not part of the learning data* ([Varoquaux and Thirion, 2014](#); [Varoquaux and Poldrack, 2019](#)). This is why our contribution is meant to ground the analysis of local brain activity gradients on quantitative analysis and model selection.

*Implications for brain parcellations.* The results presented here provide two important insights on brain parcellations. On the one hand, parcellations are *useful* to create location-specific models of brain organization, as the association between imaging characteristics may not easily hold across all brain territories. In that respect, this paper brings a novel piece of evidence, namely that local linear models — besides being much more computationally efficient — are also more powerful than a global (brain-wide) nonlinear

model. On the other hand, parcellations yield biased estimators of brain characteristics, that do not fit accurately all individual topographic information. Therefore, bagging multiple models obtained from different parcellations, that break the piece-wise nature of this model, brings significant improvements. Likewise, we do not expect any other standard population-level brain parcellation scheme to yield markedly different outcome, although it is expected that functional parcellations outperform anatomical ones ([Dadi et al., 2020](#)). The parcellation used in the present work gives the flexibility to accurately control the number of regions. Whether individualized parcellations yield different results is an interesting question for future research. The good small-sample properties of this estimator are useful, as small cohorts remain quite prevalent in brain imaging.

Another important point concerns parcellation resolution, that we studied here from the point of view of cardinality: as pointed by a reviewer, there is no reason to assume that there would be one optimal resolution that can explain all the cognitive processes covered in the paper; this is why modern parcellation schemes tend to span multiple cardinalities, as there is no one-fits-all model (see e.g. [Dadi et al. \(2020\)](#); [Schaefer et al. \(2017\)](#)). Note however that unlike previous work, our point is not handle parcels as homogeneous entities, but to study within parcel functional differences (gradients), leading to different regimes of optimality for the model.

*Low task/rest activity correspondence in motor regions?.* A result from Figures 5 and 7 is the relative weakness of explained variance in the motor regions. It is noteworthy that [Glasser et al. \(2016\)](#) defined region-level boundaries mostly from the variation of resting-state activity along the cortical surface and it was decided to keep the primary motor cortex as one region, violating the intuition of strong functional specificity in these regions ([Penfield and Boldrey, 1937](#)). The present study sheds some lights on this discrepancy. Although functional specialization is considered a well-grounded feature, it turns out that it is not well explained by differences in resting-state activity in these regions. This relative task/rest discrepancy certainly deserves more investigation, as it represents an opportunity to improve our understanding about the system-level structure of ongoing activity.

*Potential limitations.* While drawing conclusions from this study, one should remain aware of some limitations. The present study focuses on the correspondence of information provided by two fMRI-based modalities: rest-fMRI and task-fMRI. The reason is that those are easier to handle and good at identifying nuances between neighboring regions. Nevertheless, these observations may be confounded by some common artifacts (e.g. physiological effects, signals dropouts, distortions) and, therefore, results may not fully reflect actual brain properties. Cross-validation with other modalities — e.g. anatomical connectivity obtained

from diffusion MRI (Saygin et al., 2011) and anatomical properties (such as myelination index and approximation thereof (Glasser and Van Essen, 2011))— will be necessary to conclude on the nature and significance of functional gradients.

A second and less prominent issue is the fact that current analysis rests on fixed, instead of subject-adaptive parcellations. The reason is that such parcellations, while conceptually attractive, are hard to define and to validate in an objective fashion (Glasser et al., 2016). It is likely however that linear mappings are less compromised by the slight misfits of the parcellations than standard models that build on piecewise-constant approximations of brain characteristics. This limitation is further alleviated by the use of bagging across parcellations, making the final model resilient to parcellation mis-specifications.

Third, when taking the perspective of cognitive mapping, one should remember that not all functional contrasts are equally well mapped to brain territories, resulting in less consistent or more irregular mappings across individuals. Here, we focused on the limited set of contrasts made available in the datasets that we used. Probing more general families of contrasts will require multiple datasets, which goes beyond the scope of the present paper. An important related question is what cognitive domains are best predicted by resting-state features.

Worthy of note is the fact that the definition of individual topographies from resting-state might require specific methods. We use herein *Dictionary Learning*; yet, the popular *Independent Components Analysis*, with the same number of components, yields similar results (Bzdok et al., 2016; Dohmatob et al., 2016), although we observed slightly lower accuracy (about 3%). Indeed, *Sparse Dictionary Learning* and *Independent Components Analysis* are deeply related (Daubechies et al., 2009), but the former yields sharper components.

Fourth, as noted by one reviewer, the family structure of the HCP sample was not taken into account: it is thus possible that members of the same family were included in the train and test set respectively. The family information was not available to the authors when the experiments were performed. We conjecture that this has little to no impact on the results for two reasons: all the models combine equal contributions from all the HCP population, mitigating any bias related to one genetically similar sample. Second, the IBC dataset comprises unrelated individuals yet shows similar effects.

*Future work.* Given the above limitations, the present work calls for extensions to more functional datasets and modalities, to ascertain and clarify the role of functional gradients and their relevance to brain cognitive architecture.

Another direction of research consists in revisiting the parcellation tools used for brain mapping. Actually, the present work presents evidence that traditional partitions of brain domains into discrete territories fall short of capturing the properties of these territories. If one instead

considers parcels as a model of brain regions, their definition could include a (spatial) *modulation* by the degree of certainty on their spatial definition. Such objects have sometimes been called *probabilistic parcellations*. The present work establishes that the complex random nature of valid brain parcellations can be captured by sampling multiple parcellations, as done previously in Hoyos-Idrobo et al. (2018); Da Mota et al. (2013). This approach is particularly powerful in brain mapping. Note that the marginal benefit of adding parcellations decreases with the number of parcellations considered (Hoyos-Idrobo et al., 2018), so the right number typically results from a computational cost/ statistical benefit tradeoff; 20 seems to be enough in the above results.

### Conclusion

In the recent years, neuroimaging has moved away from the classical framework (that merely reports sparse-activation peaks associated with functional contrasts) in two ways: first, by focusing on the interactions among these regions, through the study of functional connectivity; and second, by assessing spatial variations of brain activation, through functional-gradients analysis.

The present paper draws the attention about the underlying difficulty of this endeavor due to the limited signal-to-noise ratio of this type of data. FMRI-based topographic mapping calls for powerful models to capture the signals and variations of interest. Specifically, predictive modeling is a useful guide for model selection; it yields statistically optimal procedures, making it possible to draw high-level conclusions on functional brain organization, e.g. in cross-species comparison (Xu et al., 2019).

Making sense of the observed gradients will ultimately require the use of dedicated datasets that overcome the low signal-to-noise ratio of each acquisition, using multiple acquisitions in fixed cohorts of participants (e.g. developing deep phenotyping framework) Gordon et al. (2017); Pinho et al. (2018). In particular, the higher prediction accuracy found for *multimodal* regions calls for more studies. Varying the conditions and modalities will be essential to formally validate the relevance of local and global brain-organization gradients.

### Acknowledgment

This project has received funding from the European Union’s Horizon 2020 Framework Program for Research and Innovation under Grant Agreement No 945539 (Human Brain Project SGA3) and the KARAIB AI chair (ANR-20-CHIA-0025-01). Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.



## References

- Abraham, A., Dohmatob, E., Thirion, B., Samaras, D., Varoquaux, G., 2013a. Extracting brain regions from rest fMRI with Total-Variation constrained dictionary learning, in: MICCAI - 16th International Conference on Medical Image Computing and Computer Assisted Intervention - 2013, Springer.
- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Muller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2013b. Machine Learning for Neuroimaging with Scikit-Learn. *Frontiers in Neuroscience*, 15URL: <https://hal.inria.fr/hal-01093971>, doi:10.3389/fninf.2014.00014.
- Amunts, K., Hawrylycz, M.J., Van Essen, D.C., Van Horn, J.D., Harel, N., Poline, J.B., De Martino, F., Bjaalie, J.G., Dehaene-Lambertz, G., Dehaene, S., et al., 2014. Interoperable atlases of the human brain. *Neuroimage* 99, 525–532.
- Barch, D., Burgess, G., Harms, M., Petersen, S., Schlaggar, B., Corbetta, M., Glasser, M., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A., Consortium, W.M., 2013. Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* 80, 169–89. doi:10.1016/j.neuroimage.2013.05.033.
- Bijsterbosch, J.D., Woolrich, M.W., Glasser, M.F., Robinson, E.C., Beckmann, C.F., Van Essen, D.C., Harrison, S.J., Smith, S.M., 2018. The relationship between spatial configuration and functional connectivity of brain regions. *eLife* 7, e32992. doi:10.7554/eLife.32992.
- Biswal, B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 34, 537–541.
- Braga, R.M., Van Dijk, K.R.A., Polimeni, J.R., Eldaief, M.C., Buckner, R.L., 2019. Parallel distributed networks resolved at high resolution reveal close juxtaposition of distinct regions. *J. Neurophysiol.* 121, 1513–1534.
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140. URL: <https://doi.org/10.1023/A:1018054314350>, doi:10.1023/A:1018054314350.
- Bzdok, D., Eickenberg, M., Grisel, O., Thirion, B., Varoquaux, G., 2015. Semi-supervised factored logistic regression for high-dimensional neuroimaging data, in: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., pp. 3348–3356.
- Bzdok, D., Varoquaux, G., Grisel, O., Eickenberg, M., Poupon, C., Thirion, B., 2016. Formal models of the network co-occurrence underlying mental operations. *PLoS Comput Biol* 12.
- Chen, P.H.C., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., Ramadge, P.J., 2015. A reduced-dimension fMRI shared response model, in: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 28*. Curran Associates, Inc., pp. 460–468.
- Cohen, A.L., Fair, D.A., Dosenbach, N.U.F., Miezin, F.M., Dierker, D., Van Essen, D.C., Schlaggar, B.L., Petersen, S.E., 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage* 41, 45–57. URL: <https://pubmed.ncbi.nlm.nih.gov/18367410>, doi:10.1016/j.neuroimage.2008.01.066. 18367410[pmid].
- Da Mota, B., Fritsch, V., Varoquaux, G., Banaschewski, T., Barker, G.J., Bokde, A.L.W., Bromberg, U., Conrod, P.J., Gallinat, J., Garavan, H., Martinot, J.L., Nees, F., Paus, T., Pausova, Z., Rietschel, M., Smolka, M.N., Ströhle, A., Frouin, V., Poline, J.B., Thirion, B., 2013. Randomized parcellation based inference. *NeuroImage*, epub ahead of printdoi:10.1016/j.neuroimage.2013.11.012.
- Dadi, K., Rahim, M., Abraham, A., Chyzyk, D., Milham, M., Thirion, B., Varoquaux, G., 2019. Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, 115–134doi:10.1016/j.neuroimage.2019.02.062.
- Dadi, K., Varoquaux, G., Machlouzariides-Shalit, A., Gorgolewski, K.J., Wassermann, D., Thirion, B., Mensch, A., 2020. Fine-grain atlases of functional modes for fMRI analysis. *NeuroImage* URL: <https://hal.archives-ouvertes.fr/hal-02496156>. working paper or preprint.
- Daubechies, I., Roussos, E., Takerkart, S., Benharrosh, M., Golden, C., D’Ardenne, K., Richter, W., Cohen, J.D., Haxby, J., 2009. Independent component analysis for brain fMRI does not select for independence. *Proceedings of the National Academy of Sciences* 106, 10415–10422. doi:10.1073/pnas.0903525106, arXiv:<https://www.pnas.org/content/106/26/10415.full.pdf>.
- DiNicola, L.M., Braga, R.M., Buckner, R.L., 2020. Parallel distributed networks dissociate episodic and social functions within the individual. *J. Neurophysiol.* 123, 1144–1179.
- Dohmatob, E., Mensch, A., Varoquaux, G., Thirion, B., 2016. Learning brain regions via large-scale online structured sparse dictionary-learning, in: *Neural Information Processing Systems (NIPS)*, Barcelona, Spain.
- Eickhoff, S.B., Constable, R.T., Yeo, B.T.T., 2018a. Topographic organization of the cerebral cortex and brain cartography. *Neuroimage* 170, 332–347.
- Eickhoff, S.B., Yeo, B.T.T., Genon, S., 2018b. Imaging-based parcellations of the human brain. *Nature Reviews Neuroscience* 19, 672–686. URL: <https://doi.org/10.1038/s41583-018-0071-7>, doi:10.1038/s41583-018-0071-7.
- Elliott, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramrakha, S., Sison, M.L., Moffitt, T.E., Caspi, A., Hariri, A.R., 2020. What is the test-retest reliability of common task-functional MRI measures? new empirical evidence and a meta-analysis. *Psychological Science* 0, 0956797620916786. URL: <https://doi.org/10.1177/0956797620916786>, doi:10.1177/0956797620916786, arXiv:<https://doi.org/10.1177/0956797620916786>. PMID: 32489141.
- Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience* 18, 1664.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping* 2, 189–210.
- Genon, S., Reid, A., Langner, R., Amunts, K., Eickhoff, S.B., 2018. How to characterize the function of a brain region. *Trends in Cognitive Sciences* 22, 350–364. URL: <https://doi.org/10.1016/j.tics.2018.01.010>, doi:10.1016/j.tics.2018.01.010.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al., 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171.
- Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by t1- and t2-weighted MRI. *Journal of Neuroscience* 31, 11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011, arXiv:<https://www.jneurosci.org/content/31/32/11597.full.pdf>.
- Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Grattton, C., Sun, H., Hampton, J.M., Coalson, R.S., Nguyen, A.L., McDermott, K.B., Shimony, J.S., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., Nelson, S.M., Dosenbach, N.U., 2017. Precision functional mapping of individual human brains. *Neuron* 95, 791–807.e7. doi:<https://doi.org/10.1016/j.neuron.2017.07.011>.

- Haxby, J., Guntupalli, J.S., Connolly, A., Halchenko, Y., Conroy, B., Gobbini, M., Hanke, M., Ramadge, P., 2011a. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–16. doi:[10.1016/j.neuron.2011.08.026](https://doi.org/10.1016/j.neuron.2011.08.026).
- Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011b. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416.
- He, T., Kong, R., Holmes, A.J., Nguyen, M., Sabuncu, M.R., Eickhoff, S.B., Bzdok, D., Feng, J., Yeo, B.T., 2020. Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage* 206, 116276. URL: <http://www.sciencedirect.com/science/article/pii/S1053811919308675>, doi:<https://doi.org/10.1016/j.neuroimage.2019.116276>.
- Hoyos-Idrobo, A., Varoquaux, G., Schwartz, Y., Thirion, B., 2018. Frem – scalable and stable decoding with fast regularized ensemble of models. *NeuroImage* 180, 160 – 172. doi:<https://doi.org/10.1016/j.neuroimage.2017.10.005>. new advances in encoding and decoding of brain signals.
- Huntenburg, J.M., Bazin, P.L., Margulies, D.S., 2018. Large-Scale Gradients in Human Cortical Organization. *Trends Cogn. Sci. (Regul. Ed.)* 22, 21–31.
- Jbabdi, S., Sotiropoulos, S., Behrens, T., 2013. The topographic connectome. *Current opinion in neurobiology* 23. doi:[10.1016/j.conb.2012.12.004](https://doi.org/10.1016/j.conb.2012.12.004).
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research* 11.
- Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences* 113, 12574–12579.
- Mensch, A., Mairal, J., Bzdok, D., Thirion, B., Varoquaux, G., 2017. Learning neural representations of human cognition across many fmri studies, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), *Advances in Neural Information Processing Systems* 30. Curran Associates, Inc., pp. 5883–5893.
- Mensch, A., Mairal, J., Thirion, B., Varoquaux, G., 2016. Dictionary learning for massive matrix factorization, in: *ICML, ACM*.
- Mesmoudi, S., Perlberg, V., Rudrauf, D., Messe, A., Pinsard, B., Hasboun, D., Cioli, C., Marrelec, G., Toro, R., Benali, H., Burnod, Y., 2013. Resting state networks’ corticotopy: the dual intertwined rings architecture. *PLoS ONE* 8, e67444.
- Nickerson, L.D., Smith, S.M., Öngür, D., Beckmann, C.F., 2017. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Frontiers in Neuroscience* 11, 115. doi:[10.3389/fnins.2017.00115](https://doi.org/10.3389/fnins.2017.00115).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* URL: <https://hal.inria.fr/hal-00650905>.
- Penfield, W., Boldrey, E., 1937. Somatic Motor and Sensory Representation in the Cerebral Cortex of Man as Studied by Electrical Stimulation. *Brain* 60, 389–443. URL: <https://doi.org/10.1093/brain/60.4.389>, doi:[10.1093/brain/60.4.389](https://doi.org/10.1093/brain/60.4.389).
- Pinho, A.L., Amadon, A., Ruest, T., Fabre, M., Dohmatob, E., Degenhien, I., Ginisty, C., Desmidt, S., Becuwe, S., Laurier, L., Doublé, C., Martins, B., Pinel, P., Eger, E., Varoquaux, G., Pallier, C., Dehaene, S., Hertz-Pannier, L., Thirion, B., 2018. Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Scientific Data* 5, 180105.
- Saygin, Z., Osher, D., Koldewyn, K., Reynolds, G., Gabrieli, J., Saxe, R., 2011. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nature neuroscience* 15, 321–7. doi:[10.1038/nn.3001](https://doi.org/10.1038/nn.3001).
- Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T., 2017. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral Cortex* 28, 3095–3114.
- Schwarzkopf, D.S., Song, C., Rees, G., 2011. The surface area of human V1 predicts the subjective experience of object size. *Nat Neurosci* 14, 28–30.
- Shafiq, M., Tyler, L., Dixon, M., Taylor, J., Rowe, J., Cusack, R., Calder, A., Marslen-Wilson, W., Duncan, J., Dalgleish, T., Henson, R., Brayne, C., Matthews, F., 2014. The cambridge centre for ageing and neuroscience (cam-can) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC neurology* 14, 204. doi:[10.1186/s12883-014-0204-1](https://doi.org/10.1186/s12883-014-0204-1).
- Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., et al., 2009. Correspondence of the brain’s functional architecture during activation and rest. *Proceedings of the National Academy of Sciences* 106, 13040–13045.
- Smith, S.M., Hyvärinen, A., Varoquaux, G., Miller, K.L., Beckmann, C.F., 2014. Group-pca for very large fmri datasets. *NeuroImage* , 738–749.
- Tavor, I., Jones, O.P., Mars, R., Smith, S., Behrens, T., Jbabdi, S., 2016. Task-free mri predicts individual differences in brain activity during task performance. *Science* 352.
- Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J., 2006. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Hum brain map* 27, 678.
- Thirion, B., Varoquaux, G., Dohmatob, E., Poline, J.B., 2014a. Which fMRI clustering gives good brain parcellations? *Frontiers in Neuroscience* 8, 13.
- Thirion, B., Varoquaux, G., Grisel, O., Poupon, C., Pinel, P., 2014b. Principal Component Regression predicts functional responses across individuals, in: *MICCAI, Springer, Boston, United States*.
- Toro, R., Burnod, Y., 2003. Geometric atlas: modeling the cortex as an organized surface. *NeuroImage* 20, 1468–1484.
- van Essen, D., et al., 2012. The human connectome project: A data acquisition perspective. *NeuroImage* 62.
- Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B., 2011. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity, in: *Information Processing in Medical Imaging, Gábor Székely, Horst Hahn. Springer, Kaufbeuren, Germany. pp. 562–573. URL: https://hal.inria.fr/inria-00588898, doi:10.1007/978-3-642-22092-0\_46*.
- Varoquaux, G., Poldrack, R., 2019. Predictive models avoid excessive reductionism in cognitive neuroimaging. *Current Opinion in Neurobiology* 55. URL: <https://hal.archives-ouvertes.fr/hal-01856412>, doi:[10.1016/j.conb.2018.11.002](https://doi.org/10.1016/j.conb.2018.11.002).
- Varoquaux, G., Sadaghiani, S., Pinel, P., Kleinschmidt, A., Poline, J.B., Thirion, B., 2010. A group model for stable multi-subject ica on fmri datasets. *NeuroImage* 51.
- Varoquaux, G., Thirion, B., 2014. How machine learning is shaping cognitive neuroimaging. *GigaScience* 3, 28. URL: <https://hal.inria.fr/hal-01094737>, doi:[10.1186/2047-217X-3-28](https://doi.org/10.1186/2047-217X-3-28).

Westfall, P.H., Young, S.S., 1993. Resampling-based multiple testing: examples and methods for P-value adjustment. John Wiley & Sons, New York; Chichester.

Xu, J., Potenza, M., Calhoun, V., Zhang, R., Yip, S., Wall, J., Pearlson, G., Worchunsky, P., Garrison, K., Moran, J., 2016a. Large-scale functional network overlap is a general property of brain functional organization: Reconciling inconsistent fmri findings from general-linear-model-based analyses. *Neuroscience and biobehavioral reviews* 71, 83–100. doi:10.1016/j.neubiorev.2016.08.035.

Xu, T., Nenning, K.H., Schwartz, E., Hong, S.J., Vogelstein, J.T., Fair, D.A., Schroeder, C.E., Margulies, D.S., Smallwood, J., Milham, M.P., Langs, G., 2019. Cross-species functional alignment reveals evolutionary hierarchy within the connectome. bioRxiv doi:10.1101/692616.

Xu, T., Opitz, A., Craddock, R.C., Wright, M.J., Zuo, X.N., Milham, M.P., 2016b. Assessing Variations in Areal Organization for the Intrinsic Brain: From Fingerprints to Reliability. *Cerebral Cortex* 26, 4192–4211. URL: <https://doi.org/10.1093/cercor/bhw241>, doi:10.1093/cercor/bhw241, arXiv:<https://academic.oup.com/cercor/article-pdf/26/11/4192/>

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

## Appendix A. Derivation of individual topographies from Resting-state fMRI

Here we describe in detail the model used. In [Appendix A.1](#), we provide an overview of the model, which is necessary to bridge concepts with the analysis performed. Then in section [Appendix A.2](#), we describe the algorithmic solution for the derivation of group-level topographies, a major ingredient of our method. In section [Appendix A.3](#), we describe how population topographies can be related to individual topographies, considering also state of the art solutions. Finally, in section [Appendix A.4](#), we describe the parallel training and prediction algorithms used.

*Notations.* In what follows,  $[[N]]$  will denote the set of integers  $\{1, \dots, N\}$ . We denote matrices with bold capital letters, vectors with bold small letters, scalars or indexes with standard small letters.  $\mathcal{N}$  will denote the multi-dimensional Gaussian distribution. For any matrix  $\mathbf{A}$ ,  $\mathbf{A}^T$  will denote its transpose and  $\mathbf{A}^\dagger$  its pseudo-inverse;  $\|\mathbf{A}\|_{\text{Fro}}$  will denote its Frobenius norm (the square root of the sum of its squared coefficients). Finally,  $\mathbf{I}_k$  will denote the identity matrix in dimension  $k$ .

### Appendix A.1. A random effects model of brain topographies

*A group Rest-fMRI model.* We use a simple linear mixing model to capture the per-voxel intrinsic activity via a small number of topographies  $\mathbf{D}_s \in \mathbb{R}^{p \times k}$ , so that  $\mathbf{X}_s \approx \mathbf{D}_s \mathbf{C}_s + \text{noise}$ .  $k$  is the model dimensionality, and the  $k$ -by- $n_s$  matrix  $\mathbf{C}_s$  encodes the time course of the subjects resting-state activity. One can formulate a prior on these signals, namely  $\mathbf{C}_s \sim \mathcal{N}(\mathbf{0}, \alpha \mathbf{I}_{n_s})$ .

Crucially, we need the representational space to be consistent across subjects, so that the ensuing predictive model generalizes well. To achieve this, we impose that resting-state data from all subjects be tied to a common shared space defined by a reference dictionary of spatial components  $\mathbf{D} \in \mathbb{R}^{p \times k}$ . Each subject-specific dictionary  $\mathbf{D}_s$  is thus a deformation of the reference dictionary  $\mathbf{D}$ . Additionally, we constrain the group-level dictionary  $\mathbf{D}$  to have sparse columns. This will force the voxels in each component of the reference dictionary to be organized in spatially compact domains, consistently with the view of the brain consisting of partially overlapping networks ([Xu et al., 2016a](#)). We denote these components—the columns of  $\mathbf{D}$  as  $\mathbf{d}^j$ . Altogether, we propose the following generative model

#### Multi-subject topographic model

$\mathbf{X} = \mathbf{D}\mathbf{C} + \text{"noise"}$  (common representation),

where  $\forall j \|\mathbf{d}^j\|_1 \leq r$  with  $r > 0$

(spatial sparsity prior on reference dictionary),

$\forall s \in [[N]] :$

$\mathbf{D}_s = (\mathbf{I}_k + \mathbf{E}_s)\mathbf{D}$

(subject-specific random effects),

$\mathbf{X}_s = \mathbf{D}_s \mathbf{C}_s + \text{"noise"}$

(subject-specific representation),

$\mathbf{C}_s \sim \mathcal{N}(\mathbf{0}, \alpha \mathbf{I})$

(prior on the individual loadings),

$\mathbf{Y}_s = f(\mathbf{D}_s) + \text{"noise"}$

(spontaneous activity predicts task-evoked activity)

where  $\mathbf{X} = [\mathbf{X}_1 | \dots | \mathbf{X}_N] \in \mathbb{R}^{p \times \nu}$  (resp.  $\mathbf{C} = [\mathbf{C}_1 | \dots | \mathbf{C}_N] \in \mathbb{R}^{k \times \nu}$ ) is a horizontal stacking of resting-state data (resp. resting-state time courses) from all the subjects, and  $\nu = \sum_{s=1}^N n_s$ .  $\|\mathbf{d}^j\|_1 := \sum_{v \in \text{voxels}} |d_v^j|$  is the  $\ell_1$  norm of the  $j$  component of the reference dictionary, and the constraint  $\|\mathbf{d}^j\|_1 \leq r$ , for  $r$  sufficiently small, imposes sparsity. Only a few voxels per component get a non-zero value. Such a choice, enforcing sparse marginal on the distribution of spatial components, solves the indeterminacy of the decomposition ([Daubechies et al., 2009](#)).

For each subject  $s$  and each voxel  $v$ , we thus define a  $k$ -dimensional encoding  $\mathbf{d}_{s,v} \in \mathbb{R}^k$  of the voxel’s time-series  $\mathbf{x}_{s,v} \in \mathbb{R}^{n_s}$  with explicit correspondences within the group-level space.

Our core assumption is that these *loadings* are a local marker of brain function. Based on this, we postulate the existence of link function  $f : \mathbb{R}^k \rightarrow \mathbb{R}^c$  that maps voxel-wise resting-state topographies to task-evoked brain activity. Following our core intuitions on the existence of local gradients of activity, we identify this link function within the span of predictive models that are piece-wise linear over brain regions. This assumption makes it possible to fit an instance of  $f$  per subject and brain region, yielding efficient training and prediction. Importantly, unlike Margulies et al. (2016), the mapping is from resting-state derived topographies to task-induced activations, not spatial coordinates to task-induced activations.

### Appendix A.2. Estimation of group-level resting-state topographies

Several models have been proposed for group resting-state modeling (Smith et al., 2014; Varoquaux et al., 2010, 2011; Abraham et al., 2013a; Mensch et al., 2016; Dohmatob et al., 2016), yet some of them are costly, as they require the iterative estimation of group-level and individual parameters. Since the resting-state time-series data are large (for example, 1200 3D volumes of  $2 \times 10^5$  voxels in each subject of the HCP dataset (van Essen et al., 2012)), a decomposition method that scales well is required. Incremental *Principal Component Analysis/ICA*-based methods (Smith et al., 2014; Varoquaux et al., 2010) have been designed for that purpose. Yet, the *structured online dictionary-learning* (Mairal et al., 2010; Dohmatob et al., 2016), which was recently proposed, has been shown to better leverage abundant data (Mensch et al., 2016; Dohmatob et al., 2016). Using structured online dictionary-learning, the shared dictionary  $\mathbf{D} \in \mathbb{R}^{p \times k}$  and the subject temporal dynamics  $\mathbf{C}_s \in \mathbb{R}^{k \times n_s}$  are obtained by solving

$$\min_{\mathbf{C}_1, \dots, \mathbf{C}_N, \mathbf{D}^1, \dots, \mathbf{D}^k} \frac{1}{N} \sum_{s=1}^N \left( \frac{1}{2} \|\mathbf{X}_s - \mathbf{D}\mathbf{C}_s\|_{\text{Fro}}^2 + \frac{\alpha}{2} \|\mathbf{C}_s\|_{\mathbf{K}}^2 \right)$$

subject to  $\|\mathbf{D}^j\|_1 \leq 1 \forall j \in [k]$

where  $\|\mathbf{A}\|_{\text{Fro}} := \sqrt{\text{tr}(\mathbf{A}^T \mathbf{A})}$  defines the Frobenius (a.k.a. Hilbert-Schmidt) norm of matrix  $\mathbf{A}$ . This problem can be efficiently solved using online block-coordinate descent. The update of a component of the dictionary —with all the other components held constant— is done via a simple projection onto the  $\ell_1$  unit-ball in  $\mathbb{R}^p$ . With the shared dictionary fixed, an a subject’s resting-state data  $\mathbf{X}_s$  is encoded by solving for  $\mathbf{C}$  via Tikhonov (generalized Ridge) regression:

$$\mathbf{C}_s \in \text{argmin}_{\mathbf{C}_s \in \mathbb{R}^{n_s \times k}} \frac{1}{2} \|\mathbf{X}_s - \mathbf{D}\mathbf{C}_s\|_{\text{Fro}}^2 + \frac{\alpha}{2} \|\mathbf{C}_s\|_{\mathbf{K}}^2.$$

All one needs is to compute a singular value decomposition of the current dictionary  $\mathbf{D}$ . The process of optimizing with respect to the shared dictionary and, then, with respect to the loadings is alternated until convergence. Technical details on online structured dictionary-learning can

be found in (Dohmatob et al., 2016; Mensch et al., 2016; Mairal et al., 2010).

### Appendix A.3. Group-level modeling of individual topographies from Rest-fMRI

The estimation of group-level resting-state topographies is performed using an optimized dictionary learning approach described in Section Appendix A.2. It formally corresponds to a *concatenated sparse dictionary learning* method. Then, for any given subject  $s$ ,  $\mathbf{D}_s$  is the individual counterpart of  $\mathbf{D}$ . We assume  $\mathbf{D}$  to be under-complete – i.e  $k \ll \min(\min_{s \in [S]} n_s, p)$ – and therefore full-rank (i.e  $\mathbf{D}^T \mathbf{D}$  is invertible). Under this hypothesis, we can estimate  $\mathbf{C}_s$  by standard linear regression, leading to:

$$\mathbf{C}_s = \mathbf{D}^\dagger \mathbf{X}_s,$$

Following the popular dual-regression approach (Nickerson et al., 2017), it is tempting to define

$$\mathbf{D}_s \in \text{argmin}_{\mathbf{D}_s \in \mathbb{R}^{k \times p}} \frac{1}{2} \|\mathbf{X}_s - \mathbf{D}_s \mathbf{C}_s\|_{\text{Fro}}^2,$$

leading to  $\mathbf{D}_s = \mathbf{X}_s \mathbf{C}_s^\dagger$ . Plugging the first equation into the second leads to

$$\mathbf{D}_s = \mathbf{X}_s \mathbf{X}_s^\dagger \mathbf{D},$$

Meaning that  $\mathbf{D}_s$  is simply the orthogonal projection of  $\mathbf{D}$  onto  $\text{span}(\mathbf{X}_s)$ , i.e. the space spanned by the brain maps observed in  $\mathbf{X}_s$ . This clarifies the concept of *multiplicative perturbation* introduced previously in the definition of  $\mathbf{D}_s$ . This perturbation is simply a linear projection on the observation of a subject  $s$ . To our knowledge, this relationship has not been established previously.

*Relation to Shared Response Model (SRM).* It turns out that the *Shared Response Model* (SRM) framework (Chen et al., 2015) and the “dual regression” (DR) scheme (Nickerson et al., 2017), that we have presented, are closely related. Note that the SRM can be seen as an efficient implementation of the “hyperalignment” method Haxby et al. (2011b), in which individual data are projected to a template using Procrustes (orthogonal) transformation. Indeed, Chen et al. (2015) considers the following problem

$$\begin{aligned} & \text{minimize} \quad \frac{1}{N} \sum_{s=1}^N \|\mathbf{X}_s - \mathbf{D}_s \mathbf{C}\|_{\text{Fro}}^2 \\ & \text{over } \mathbf{D}_s \in \mathbb{R}^{p \times k}, \mathbf{C} \in \mathbb{R}^{k \times n_s}, \\ & \text{subject to } \mathbf{D}_s^T \mathbf{D}_s = \mathbf{I}_k, \forall s \in [1 \dots N]. \end{aligned}$$

Ignoring the orthonormality constraints “ $\mathbf{D}_s^T \mathbf{D}_s = \mathbf{I}_k$ ” this problem is analogous to the above spatial-temporal regression problem. (A.1) is usually solved via an alternating minimization scheme. Viz,

- **Update rotations (orthonormal Procrustes analysis):**

$$\mathbf{D}_s^{(t+1)} = \mathbf{U}_s^{(t)} \mathbf{V}_s^{(t)T} \quad \forall s \in [1 \dots N],$$

where  $\mathbf{U}_s^{(t)} \mathbf{\Sigma}_s \mathbf{V}_s^{(t)T}$  is the Singular Value decomposition of the  $p$ -by- $k$  matrix  $\mathbf{X}_s \mathbf{C}^{(t)\dagger}$ .

- **Update shared-dictionary:**

$$\mathbf{C}^{(t+1)} = \frac{1}{N} \sum_{s=1}^N \mathbf{C}_s^{(t+1)},$$

where  $\mathbf{C}_s^{(t+1)} := \mathbf{D}^{(t+1)T} \mathbf{X}_s$ .

However, dual-regression is much more attractive due to its low cost, as it is non-iterative. Indeed in comparison, hyperalignment performs a Singular Value Decomposition per subject per iteration and becomes prohibitive for large problems. Note that hyperalignment is usually done region-wise (i.e locally) to limit the cost of the fit.

#### Appendix A.4. Algorithms for learning and inference

Algorithm 1 describes how learning proceeds based on individual topographies ( $\mathbf{D}_s$ ),  $s = 1..N_{\text{train}}$ . Then Algorithm 2 describes the inference procedure that was used.

---

**Algorithm 1 Massively parallel training of the predictive model:** for a collection of parcellation  $\mathcal{P}^q$ ,  $q \in [Q]$ , the model learns to predict task-related maps from resting-state maps. It results in one global-brain estimator per parcellation.

**Require:** Data from  $N_{\text{train}}$  subjects. For each subject  $s$  we have pre-computed spatial features  $\mathbf{D}_s \in \mathbb{R}^{p \times k}$ . A set of brain parcellations  $(\mathcal{P})^q$ ,  $q \in [Q]$ .

**Ensure:** For every parcellation  $\mathcal{P}^q = \bigcup_m \mathcal{P}_m^q$ , and every parcel  $\mathcal{P}_m^q$  thereof, a fitted estimator  $f_m^q$  for predicting the task-evoked activity  $\mathbf{Y}_{s'}|_{\mathcal{P}_m^q}$  of the voxels in that parcel from the spatial features  $\mathbf{D}_{s'}|_{\mathcal{P}_m^q}$ .

- 1: **parallel for** each subject  $s \in \text{train}$  **do**
  - 2:     **parallel for** each parcellation  $\mathcal{P}^q$  **do**
  - 3:         **parallel for** each parcel  $\mathcal{P}_m^q \in \mathcal{P}^q$  **do**
  - 4:             **Fit** a predictive model  $\mathbf{D}_s^T|_{\mathcal{P}_m} \xrightarrow{f_{s,m}^q} \mathbf{Y}_s|_{\mathcal{P}_m^q}$
  - 5:             **end pararell for**
  - 6:             **Bagging:**  $f_m^q \leftarrow \frac{1}{N_{\text{train}}} \sum_{s=1}^{N_{\text{train}}} f_{s,m}^q$
  - 7:         **end pararell for**
  - 8:     **end pararell for**
- 

## Appendix B. Best predicted regions from the IBC dataset

Figure B.10 shows the average prediction accuracy per voxel, across contrasts and subjects, for the IBC dataset.

---

**Algorithm 2 Massively parallel prediction:** the aggregation mechanism is used to obtain one improved prediction that does not reflect merely one parcellation, but an ensemble thereof.

**Require:** Data from  $N_{\text{test}}$  subjects. For each subject  $s$ , we have pre-computed spatial features  $\mathbf{D}_s \in \mathbb{R}^{p \times k}$ . Also, a collection of bagged fitted models  $f_m^q$ , one per parcel per parcellation  $q \in [Q]$  (see Algorithm 1).

**Ensure:** Predictions  $\mathbf{Y}_s \in \mathbb{R}^{p \times c}$ , for each test subject  $s$ .

- 1: **parallel for** each subject  $s \in \text{test}$  **do**
  - 2:      $\mathbf{Y}_s \leftarrow \mathbf{0} \in \mathbb{R}^{p \times c}$
  - 3:     **parallel for** each parcellation  $\mathcal{P}^q = \bigcup_m \mathcal{P}_m^q$  **do**
  - 4:         **parallel for** each parcel  $\mathcal{P}_m^q \in \mathcal{P}^q$  **do**
  - 5:             **Predict**  $\mathbf{Y}_s^q|_{\mathcal{P}_m^q} \leftarrow \mathbf{Y}_s|_{\mathcal{P}_m} + f_m^q(\mathbf{D}_s^T|_{\mathcal{P}_m^q})$
  - 6:             **end pararell for**
  - 7:     **end pararell for**
  - 8:      $\mathbf{Y}_s \leftarrow \frac{1}{Q} \mathbf{Y}_s^q$
  - 9: **end pararell for**
- 

The best-fitted regions are the ventral visual pathway, the language network and the Intra-Parietal Sulcus bilaterally. The scores are typically weaker than the those of the HCP dataset (compare with Figure 7) due to the smaller number of subjects, and the set of best-fitted regions are slightly different, reflecting the different functional protocols used.

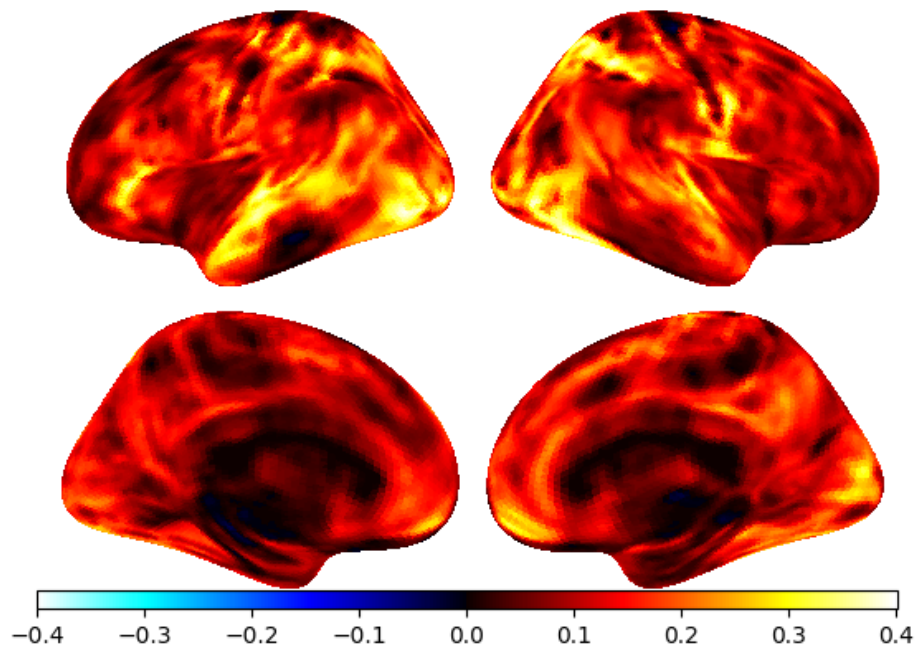


Figure B.10: **Average prediction accuracy per voxel (across contrasts) on the IBC dataset.** This map displays  $R^2_{\text{voxel}}$  statistic obtained from the IBC dataset. This outlines the brain regions that are overall best fitted by the model: the ventral visual pathway, the language network and the Intra-Parietal Sulcus bilaterally.