NeuroImage 89 (2014) 203–215

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Randomized parcellation based inference

Benoit Da Mota [a,b,*,1], Virgile Fritsch [a,b,1], Gaël Varoquaux [a,b], Tobias Banaschewski [e,f], Gareth J. Barker [d], Arun L.W. Bokde [j], Uli Bromberg [g], Patricia Conrod [d,h], Jürgen Gallinat [i], Hugh Garavan [j,r], Jean-Luc Martinot [k,l], Frauke Nees [e,f], Tomas Paus [m,n,o], Zdenka Pausova [q], Marcella Rietschel [e,f], Michael N. Smolka [p], Andreas Ströhle [i], Vincent Frouin [b], Jean-Baptiste Poline [b,c], Bertrand Thirion [a,b,*], the IMAGEN consortium [2]

[a] Parietal Team, INRIA Saclay-Île-de-France, Saclay, France
[b] CEA, DSV, I²BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France
[c] Henry H. Wheeler Jr. Brain Imaging Center, University of California at Berkeley, USA
[d] Institute of Psychiatry, Kings College London, UK
[e] Central Institute of Mental Health, Mannheim, Germany
[f] Medical Faculty Mannheim, University of Heidelberg, Germany
[g] Universitaetsklinikum Hamburg Eppendorf, Hamburg, Germany
[h] Department of Psychiatry, Universite de Montreal, CHU Ste. Justine Hospital, Canada
[i] Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charité Universitätsmedizin Berlin, Germany
[j] Institute of Neuroscience, School of Medicine, Trinity College Dublin, Dublin, Ireland
[k] Institut National de la Santé et de la Recherche Médicale, INSERM CEA Unit 1000 "Imaging & Psychiatry", University Paris Sud, Orsay, France
[l] AP-HP Department of Adolescent Psychopathology and Medicine, Maison de Solenn, University Paris Descartes, Paris, France
[m] Rotman Research Institute, University of Toronto, Toronto, Canada
[n] School of Psychology, University of Nottingham, UK
[o] Montreal Neurological Institute, McGill University, Canada
[p] Neuroimaging Center, Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Germany
[q] The Hospital for Sick Children, University of Toronto, Toronto, Canada
[r] Department of Psychiatry and Psychology, University of Vermont, Burlington, USA

ARTICLE INFO

Article history:
Accepted 5 November 2013
Available online 19 November 2013

Keywords:
Group analysis
Parcellation
Reproducibility
Multiple comparisons
Permutations

ABSTRACT

Neuroimaging group analyses are used to relate inter-subject signal differences observed in brain imaging with behavioral or genetic variables and to assess risks factors of brain diseases. The lack of stability and of sensitivity of current voxel-based analysis schemes may however lead to non-reproducible results. We introduce a new approach to overcome the limitations of standard methods, in which active voxels are detected according to a consensus on several random parcellations of the brain images, while a permutation test controls the false positive risk. Both on synthetic and real data, this approach shows higher sensitivity, better accuracy and higher reproducibility than state-of-the-art methods. In a neuroimaging–genetic application, we find that it succeeds in detecting a significant association between a genetic variant next to the *COMT* gene and the BOLD signal in the left thalamus for a functional Magnetic Resonance Imaging contrast associated with incorrect responses of the subjects from a *Stop Signal Task* protocol.

Introduction

Analysis of brain images acquired on a group of subjects makes it possible to draw inferences on regionally-specific anatomical properties of the brain, or its functional organization. The major difficulty with such studies lies in the inter-subject variability of brain shape and vasculature. In functional studies, a task-related variability of subject performance is also observed. The standard-analytic approach is to register and normalize the data in a common reference space. However a perfect voxel-to-voxel correspondence cannot be attained, and the impact of anatomical variability is tentatively reduced by smoothing (Frackowiak et al., 2003). This problem holds for any statistical test, including those associated with multivariate procedures. In the absence of ground truth, choosing the best procedure to analyze the data is a challenging problem. Practitioners as well as methodologists tend to prefer models that maximize the sensitivity of a test under a given control for false detections. The level of sensitivity conditional to this control is indeed informative on the usefulness of a model.

* Corresponding authors at: CEA, DSV, I²BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France.
E-mail addresses: benoit.da_mota@inria.fr (B. Da Mota), bertrand.thirion@inria.fr (B. Thirion).
[1] These authors contributed equally to this work.
[2] URL: http://www.imagen-europe.com

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.11.012

Classic statistical tests for neuroimaging

The reference approach in neuroimaging is to fit and test a model at each voxel (univariate voxelwise method), but the large number of tests performed yields a multiple comparison problem. The statistical significance of the voxel intensity test can be corrected with various statistical procedures. First, Bonferroni correction consists in adjusting the significance threshold by dividing it by the number of tests performed. This approach is known to be conservative, especially when non-independent tests are involved, which is the case of neighboring voxels in neuroimaging. Another approach consists in a permutation test to perform a family-wise correction of the p-values (Nichols and Holmes, 2002). Although computationally costly, this method has been shown to yield more sensitive results than studies involving Bonferroni-corrected experiments (Petersson et al., 1999). A good compromise between computation cost and sensitivity can be found in analytic corrections based on Random Field Theory (RFT), in which the smoothness of the images is estimated (Worsley et al., 1992). However, this approach requires both high threshold and data smoothness to be really effective (Hayasaka et al., 2004).

Another widely used method is a test on cluster size, which aims to detect spatially extended effects (Friston et al., 1993; Poline and Mazoyer, 1993; Roland et al., 1993). The statistical significance of the size of an activation cluster can be obtained with theoretical corrections based on the RFT (Hayasaka et al., 2004; Worsley et al., 1996b) or with a permutation test (Holmes et al., 1996; Nichols and Holmes, 2002). Cluster-size tests tend to be more sensitive than voxel-intensity tests, especially when the signal is spatially extended (Friston et al., 1996; Moorhead et al., 2005; Poline et al., 1997), at the expense of a strong statistical control on all the voxels within such clusters. This approach however suffers from several drawbacks. First, such a procedure is intrinsically unstable and its result depends strongly on an arbitrary cluster-forming threshold (Friston et al., 1996). The threshold-free cluster enhancement (TFCE) addresses this issue, by avoiding the choice of an explicit, fixed threshold (Salimi-Khorshidi et al., 2011; Smith and Nichols, 2009) but leads to other arbitrary choices: the TFCE statistic mixes cluster-extent and cluster-intensity measures in proportions that can be defined by the user. More generally, tests that combine cluster size and voxel intensity have been proposed (Hayasaka and Nichols, 2004; Poline et al., 1997). Second, the correlation between neighboring voxels varies across brain images, which makes detection difficult where the local smoothness is low. Combining permutations and RFT to adjust for spatially-varying smoothness leads to more sensitive procedures (Hayasaka et al., 2004; Salimi-Khorshidi et al., 2011). A more complete discussion of the limitations and comparisons of these techniques can be found in (Moorhead et al., 2005; Petersson et al., 1999).

Spatial models for group analysis in neuroimaging

Spatial models try to overcome the lack of correspondence between individual images at the voxel level. The most straightforward and widely used technique consists of smoothing the data to increase the overlap between subject-specific activated regions (Worsley et al., 1996a). In the literature, several approaches propose more elaborate techniques to model the noise in neuroimaging, like Markov Random Fields (Ou et al., 2010), wavelet decomposition (Van de Ville et al., 2004), spatial decomposition or topographic methods (Flandin and Penny, 2007; Friston and Penny, 2003) and anatomically informed models (Keller et al., 2009). These techniques are not widely used probably because they are computationally costly and not always well-suited for analysis of a group of subjects. A popular approach consists of working with subject-specific Regions of Interest (ROIs), that can be defined in a way that accommodates inter-subject variability (Nieto-Castanon et al., 2003). The main limitation of such an approach (Bohland et al., 2009) is that there is no widely accepted standard for partitioning the brain, especially for the neocortex. Data-driven parcellation was proposed by Thirion et al. (2006) to overcome this limitation: they improve the sensitivity of random effect analysis by considering parcels defined at the group level.

Neuroimaging–genetic studies

While most studies investigate the difference of activity between groups or the level of activity within a population, neuroimaging studies are often concerned by testing the effect of exogeneous variables on imaging target variables, and there is increasing interest in the joint study of neuroimaging and genetics to improve understanding of both normal and pathological variability of the brain organization. Single nucleotide polymorphisms (SNPs) are the most common genetic variants used in such studies: They are numerous and represent approximately 90% of the genetic between-subject variability (Collins et al., 1998). Voxel intensity and cluster size methods have been used for genome-wide association studies (GWAS) (Stein et al., 2010), but the multiple comparison problem does not permit finding significant results, despite efforts to estimate the effective number of tests (Gao et al., 2010) or by running computationally expensive, but accurate permutation tests (Da Mota et al., 2012). Recently, important efforts have been done to design more sophisticated multivariate methods (Floch et al., 2012; Kohannim et al., 2011; Vounou et al., 2010), the results of which are more difficult to interpret; another alternative is to work at the gene level instead of SNPs (Ge et al., 2012; Hibar et al., 2011).

The randomized parcellation approach

The parcellation model (Thirion et al., 2006) has several advantages: (*i*) it is a simple and easily interpretable method, (*ii*) by reducing the number of descriptors, it reduces the multiple comparisons problem, and (*iii*) the choice of the parcellation algorithm can lead to parcels adapted to the local smoothness. But parcellations, when considered as spatial functions, highly depend on the data used to construct them and the choice of the number of parcels. In general, a parcellation defined in a given context might not be a good descriptor in a slightly different context, or may generalize poorly to new subjects. This implies a lack of reproducibility of the results across subgroups, as illustrated later in Fig. 7. The weakness of this approach is the large impact of a parcellation scheme that cannot be optimized easily for the sake of statistical inference; it may thus fail to detect effects in poorly segmented regions. We propose to solve this issue by using several randomized parcellations (Bühlmann et al., 2012; Varoquaux et al., 2012) generated using resampling methods (bootstrap) and average the corresponding statistical decisions. Replacing an estimator such as parcel-level inference by means of bootstrap estimates is known to *stabilize* it; a fortunate consequence is that the *reproducibility* of the results (across subgroups of subjects) is improved. Formally, this can be understood as handling the parcellation as a hidden variable that needs to be integrated out in order to obtain the posterior distribution of statistical values. The final decision is taken with regard to the stability of the detection of a voxel (Alexander and Lange, 2011; Meinshausen and Bühlmann, 2010) across parcellations, compared to the null hypothesis distribution obtained by a permutation test.

A multivariate problem: the detection of outliers

The benefits of the randomized parcellation approach can also be observed in multivariate analysis procedures, such as predictive modeling (Varoquaux et al., 2012) or outlier detection. In this work, we focus on the latter: neuroimaging datasets often contain atypical observations; such *outliers* can result from acquisition-related issues (Hutton et al., 2002), bad image processing (Wu et al., 1997), or they can merely be extreme examples of the high variability observed in the population.