

Principal Component Regression predicts functional responses across individuals

Bertrand Thirion, Gaël Varoquaux, Olivier Grisel, Cyril Poupon, Philippe Pinel

▶ **To cite this version:**

Bertrand Thirion, Gaël Varoquaux, Olivier Grisel, Cyril Poupon, Philippe Pinel. Principal Component Regression predicts functional responses across individuals. MICCAI, Sep 2014, Boston, United States. hal-01015173

HAL Id: hal-01015173
https://inria.hal.science/hal-01015173

Submitted on 25 Jun 2014

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.

Principal Component Regression predicts functional responses across individuals

Bertrand Thirion, Gaël Varoquaux, Olivier Grisel, Cyril Poupon, and Philippe Pinel

Parietal team, INRIA Saclay-le-de-France, 91128 Palaiseau, France
CEA, DSV, I^{2}BM, Neurospin, 91191 Gif sur Yvette, France
http://team.inria.fr/parietal/

Abstract. Inter-subject variability is a major hurdle for neuroimaging group-level inference, as it creates complex image patterns that are not captured by standard analysis models and jeopardizes the sensitivity of statistical procedures. A solution to this problem is to model random subjects effects by using the redundant information conveyed by multiple imaging contrasts. In this paper, we introduce a novel analysis framework, where we estimate the amount of variance that is fit by a random effects subspace learned on other images; we show that a principal component regression estimator outperforms other regression models and that it fits a significant proportion (10% to 25%) of the between-subject variability. This proves for the first time that the accumulation of contrasts in each individual can provide the basis for more sensitive neuroimaging group analyzes.

Keywords: fMRI; principal components regression; random effects

1 Introduction

Functional brain imaging is a tool of choice to understand the functional specialization of brain areas and decipher the mechanisms or the impact of various brain diseases. It outputs *contrast maps* that quantify metabolic responses of the brain gray matter in various experimental situations; more precisely, a *contrast* is a certain combination of the activation evoked by various experimental conditions that reveals brain responses specific to a cognitive process; the accumulation of such contrasts can then be used to derive subject-specific or population-level functional cartography [1]. The underlying inference is based on a mass-univariate framework that detects which brain voxels show in average a positive response. Between-subjects variability enters the model as a random effect: larger variability entails lower evidence that an effect is present, hence less sensitivity [2].

While classical analyzes ignore the structure underlying these random effects, there is a growing evidence that some of it can be characterized or even predicted by several factors, such as *i)* the presence of an average effect in the population (high variability coincides spatially with larger mean effect) [3] *ii)* anatomical

differences, such as the cortical ends of anatomical connections, which has been illustrated by a prediction of *faces versus scenes* contrast in the fusiform gyrus [4] and *iii)* individual functional connectivity between brain regions, which has been shown to provide a relevant prior for group analyzes in [5]. Modeling such random effects is important in order to facilitate future correlative studies that compare brain features to genetics and behavior.

In this work, we consider the prediction of functional contrasts by other contrasts observed independently in the same subjects, assuming that the between-subject variability can be captured through a latent subspace structure. This approach is reminiscent of the popular *modality propagation* analysis paradigm (among others, see [6]), in which a target image is reconstructed from a reference one by machine learning tools; the main difference is that functional neuroimaging data have a much lower signal-to-noise ratio. However, a large number of contrast images can potentially be used to capture the structure of the random effects space. Altogether, this implies that the same procedures are not optimal for *random effects characterization* and *modality prediction*.

The contribution of this paper is two-fold: *i)* we propose a formalization of the random effects subspace framework and derive effective methods to estimate such models on real data; in particular, principal component regression (PCR [7]) comes naturally as a tool of choice in this problem; *ii)* we benchmark several estimators to test the effectiveness of PCR in that setting as well as three spatial frameworks in the inference process: a *local* framework based on each spatial unit of the domain, a *regional* model that uses information from neighboring regions, and a *global* framework that estimates the random effects subspace on an initial decomposition of the images into principal components.

2 The random effects subspace model

Let $(\mathbf{X}^s), s \in \{1, .., n_{subjects}\}$ be $(n_{contrasts}, n_{voxels})$ matrices that represent a set of imaging contrasts observed in a given group of subjects. $\mathbf{X}^s$ is typically analyzed as the sum of a fixed effect $\bar{\mathbf{X}}$ across individuals, random effects that represent the between subject variability and observation noise. The simplest approach to extract the latent structure of these random effects is a factorization that represents $\mathbf{X}^s - \bar{\mathbf{X}}$ as a combination of subject-specific latent factors $\boldsymbol{\zeta}^s = (\boldsymbol{\zeta}_k)^s, k = 1..K$ of shape (n_{voxels}), weighted by voxel-specific coefficients $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_k), k = 1..K$ that are common to to the population:

$$\mathbf{X}^s = \bar{\mathbf{X}} + \sum_{k=1}^{K} \boldsymbol{\zeta}_k^s \boldsymbol{\alpha}_k + \boldsymbol{\varepsilon}, \ \forall s \in \{1, .., n_{subjects}\} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the observation noise, K is the number of latent components, and the product $\boldsymbol{\zeta}_k^s \boldsymbol{\alpha}_k$ is taken element-wise. Note that both $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}^s$ can vary spatially. The identification of model (1) yields a series of questions: *i)* what fraction of the variance is fit by the latent components versus observation noise? *ii)* to which extent is the amount of explained variance correlated to the fixed effects? *iii)* Are

the latent components shared across brain regions or specific to a given location? These questions can be handled by predictive models that quantify the amount of residual variance $var(\mathbf{X_t}|\mathbf{X_{ref}})$, where $\mathbf{X_t}$, $\mathbf{X_{ref}}$ correspond respectively to a test and a set of reference contrasts acquired independently in the same subjects. The corresponding model is the following one:

$$\mathbf{X_t}^s = \bar{\mathbf{X}}_\mathbf{t} + \sum_{k=1}^{K} (\boldsymbol{\zeta}_\mathbf{ref})_k^s \boldsymbol{\alpha}_{\mathbf{t}k} + \boldsymbol{\varepsilon},\ \forall s \in \{1, .., n_{subjects}\} \tag{2}$$

and it goodness of fit is can be measured by $\left\| \sum_{s=1}^{n_{subjects}} \mathbf{X_t}^s - \left(\bar{\mathbf{X}}_\mathbf{t} + \sum_{k=1}^{K} (\boldsymbol{\zeta}_\mathbf{ref})_k^s \boldsymbol{\alpha}_{\mathbf{t}k} \right) \right\|^2$.
In this work, we explore several solutions to estimate the latent factors $\boldsymbol{\zeta}_\mathbf{ref}$:

- Principal component regression (PCR), that defines $\boldsymbol{\zeta}_\mathbf{ref}$ by a PCA of $\mathbf{X_{ref}}$ and estimates the loadings ($\boldsymbol{\alpha}_\mathbf{t}$) by simple linear regression.
- Ridge regression considers $\boldsymbol{\zeta}_\mathbf{ref}^s = \mathbf{X_{ref}}^s - \bar{\mathbf{X}}_\mathbf{ref}$, but applies a uniform shrinkage on the loadings ($\boldsymbol{\alpha}_\mathbf{t}$).
- Lasso regression, that performs a non-linear, sparsity-inducing shrinkage on the loadings ($\boldsymbol{\alpha}_\mathbf{t}$), given $\boldsymbol{\zeta}_\mathbf{ref}^s = \mathbf{X_{ref}}^s - \bar{\mathbf{X}}_\mathbf{ref}$.
- Extremely Randomized Trees (Extra Trees) [8], that also take $\boldsymbol{\zeta}_\mathbf{ref}^s = \mathbf{X_{ref}}^s - \bar{\mathbf{X}}_\mathbf{ref}$, yet learn ($\boldsymbol{\alpha}_\mathbf{t}$) by tree-based regression as in random forests.

Next, to address the question whether the latent components are local to some image region, shared locally, or across brain regions, we perform the analysis at three different spatial scales:

- The *local* scale, in which the model (2) is estimated independently at each brain location considered.
- The *regional* scale, in which $\boldsymbol{\zeta}_\mathbf{ref}$ and ($\boldsymbol{\alpha}_\mathbf{t}$) are estimated in the neighborhood of a given image location.
- The *global* scale, in which the model is estimated on a global model of the data. In this work, we use a PCA of the whole dataset, written as a $(n_{voxels}, n_{subjects} \times n_{contrasts})$ matrix; we used 400 components.

Comparing the predictive scores of these approaches captures a bias/variance trade-off –local estimators being less biased but more variable– but also indicates how redundant the information is, as redundancy is used only by global or regional methods. The learners used in this paper are the scikit-learn implementation [9] v0.14, used with default parameters. The code can be found at `https://github.com/bthirion/fMRI_PCR`.

3 Experiments

As our aim is model comparison, we directly turn to an actual MRI dataset to assess how well test contrasts are predicted by the above estimators.
Data. A dataset consisting of four functional MRI sessions acquired under different experimental protocols and an anatomical image, obtained in 79 subjects,

was used. Standard preprocessing were carried out with SPM8 software and its default parameters; functional images were warped in the MNI152 coordinate space and resampled at 3mm resolution. A population mask of the gray matter voxels was obtained by averaging individual gray matter probability maps obtained from SPM8, resampling this average at the fMRI data resolution and thresholding it, which yields about 57,000 voxels. Contrasts were obtained using a standard linear model, based on the specified task timing information,canonical hemodynamic response function, high-pass filtering and auto-regressive noise model. The estimation of the model parameters was carried out using the Nipy software.

Contrasts. The four functional protocols were the following: *i)* a *generic localizer* protocol, that assesses basic functions, such as finger tapping, listening and reading, attentional tasks; *ii)* a *social* protocol in which the subjects were exposed to false beliefs, to objects moving with or without a putative intention, speech and non-speech *iii)* a so-called *parietal mapping* protocol, during which the subjects had to perform visual saccades, grasping, judge the orientation of an object; *iv)* an *emotional* protocol, during which the subjects had to judge the confidence inspired by or the gender of a face or of an expression, together with a control tasks on the same stimuli. Five to six main contrasts were retained from each experiment to form a set of functional contrasts of interest. The reference set consisted of the contrasts *i), ii) and iii)* or *i), ii) and iv)*, while the test set was the set of contrasts of the left-out acquisition.

Evaluation metric. Based on the approach described in section 2, we use the following criterion: $\rho = 1 - \frac{\|\mathbf{X_t} - (\hat{\mathbf{X}}_\mathbf{t}|\mathbf{X_{ref}})\|^2}{\|\mathbf{X_t} - \bar{\mathbf{X}}_\mathbf{t}\|^2}$, where the norm is taken over the whole image, computed over a cross-validation loop on subjects. ρ can readily be interpreted as the explained variance ratio of the model (the higher the better). In all experiments presented here, the test contrast has been observed in another session than the reference contrasts, in order to enforce independence of the samples. In our experiments, we used a 20-fold shuffle-split procedure, where 90% of the subjects are used for training and 10% are used for testing.

Parameter setting. All the free parameters of the methods (amount of penalty for the ridge and lasso models, number of principal components for PCR) were set by nested cross-validation, using 3-fold cross-validation on the training set. Finally, the data presented here are based upon an initial parcellation of the dataset into $Q = 10,000$ super-voxels based on Ward clustering instead of the $57,000$ voxels of the reference cortical mask. Voxel-specific contrasts were averaged to parcel-specific values. Note that this procedure is simply meant to be a compression of the data and has no influence on the model nor its interpretation. We checked that using finer or coarser parcellations does not alter the experimental results.

4 Results

Method comparison. A boxplot of the ρ score computed over 11 contrast is presented in Fig. 1. Regarding estimator comparison, this shows that, for all spatial scales under consideration (local, regional, global), the following hierarchy

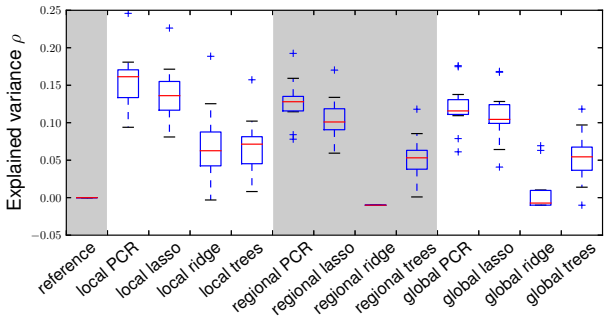


Fig. 1. Comparison of the methods for prediction of test contrasts. The boxplot represents the average rate of explained variance through different methods for eleven different functional contrasts.

holds: PCR outperforms Lasso, that outperforms Extra-Trees regression and the ridge estimator. These differences significant using pairwise Wilcoxon tests (PCR >lasso, lasso > ridge, lasso > extra trees, $p < .005$, in all spatial configurations; Paired Student tests yield much lower p-values). Note that we considered other variants of ensemble estimators, such as Gradient Boosted Trees and Random Forests, but these were systematically outperformed by the Extra Trees estimator. Second, local estimators significantly ($p < 0.005$) outperform regional and global estimators: this means that there is not enough information shared across regions to improve the predictive accuracy of the model, while the addition of new regressors degrades its predictive accuracy, through a classical curse of dimensionality phenomenon.

Well-predicted brain regions are those that show large fixed effects. To better understand the mechanisms of across-contrasts random effect prediction, an important question is whether the prediction accuracy is uniform across brain regions, or whether it coincides with regions that show a strong mean effect. In Fig. 3, we show for a contrast displaying the activation evoked by *deciphering the intention from face expression*, the map of the (across subjects) mean statistical response together with a map of the prediction accuracy obtained with local PCR. This shows that high values are obtained in the same occipital, parietal and pre-frontal cortical regions. This effect holds for all other contrasts: the bottom part of Fig. 3 shows the joint density of ρ versus the average (across contrasts and subjects) contrast absolute value (to discount sign effects); this shows that the regions with high predictability are also those with larger mean activation strength. The 10% of the regions showing the largest average responses and those showing highest predictability (across contrasts and subjects) are displayed in the same figure (right side): these are the occipital, motor cortex and supplementary motor area in both cases; note however that the orbito-frontal cortex shows high activation in average and relatively lower predictability, while

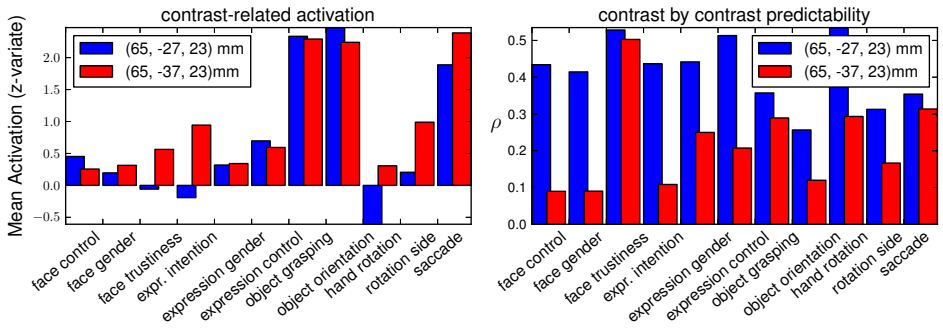


Fig. 2. Predictability reveals other features than average statistical sensitivity. Two close regions of interest have been selected, in the right supra-marginal gyrus (x=65mm, y=-27 or -37mm, z=23mm). While both regions approximately display similar activation strength across 11 test conditions (left), the individual responses in the anterior one is much more predictable than the responses from the posterior region (right).

the posterior superior temporal gyrus is highly predictable although the mean activation is weaker there.

ROI-based analysis. To better illustrate potential differences between predictability and average activation, we display the average functional signal for two nearby ROIs of the supra-marginal gyrus ($x, y, z = 65, -27, 23$ and $65, -37, 23$ mm respectively): both show the same amount of activation across eleven contrasts, but the most anterior one is much more predictable than those of the posterior one. A straightforward explanation is that the anterior region is much more consistently defined in the MNI space as the border of the post-central gyrus, while the posterior one is a region that does not get a consistent definition, and is more variable anatomically and functionally [10, 11].

5 Discussion

This paper presents the first successful attempt to predict *individual* functional responses based on previously observed data from the same subjects, evaluated across all brain regions. Here the prediction is based on the observation of the same contrasts in other subjects and mimics the realistic incomplete dataset situations where some general contrasts have been observed in some subjects while others are missing. The accuracy of the prediction is relatively high: 10% to 25 % of the variance can be fit using the local PCR model. This confirms a posteriori the existence of a *latent subspace* that spans the variability of individual responses. Moreover, the best performing technique, PCR, is precisely the one that builds on the low-dimensional latent space model of the variability.

The potential superiority of PCR against ridge has already been described [7] and in the present case, it confirms that dimension reduction through PCA of a set of reference contrasts is better at separating random effects from noise than

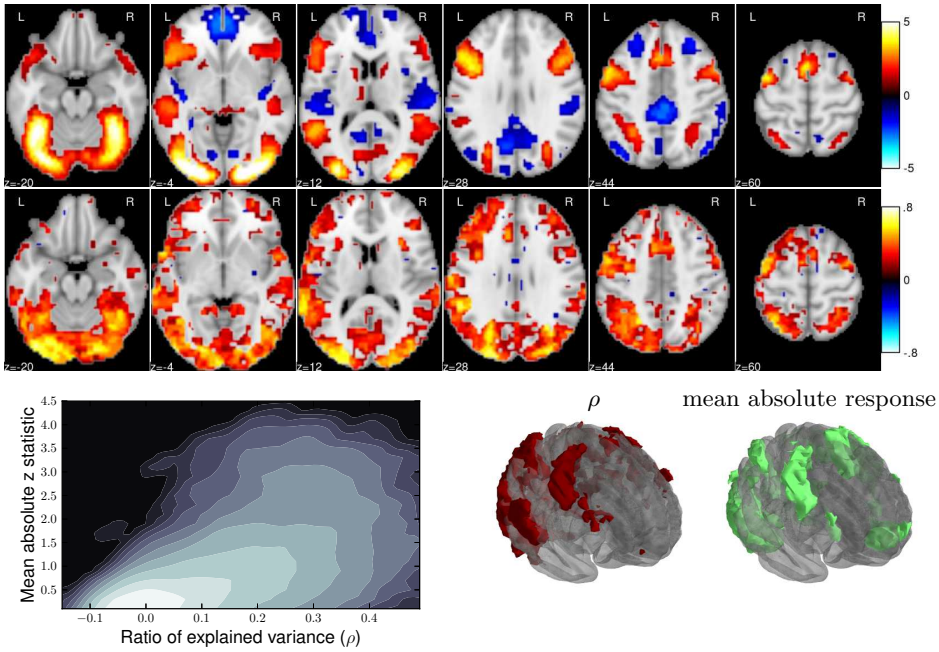


Fig. 3. Coincidence of regions that show high activation values and those that are actually predictable across individuals. (Top) Cross-subject average activation level for a contrast related to deciphering expression intention. (Middle) Rate of prediction across the same regions for the same contrast. (Bottom) left: joint histogram of the mean statistical significance and ratio of explained variance per voxel, across experimental conditions; middle: 3D view of the most predictable regions across conditions; right: 3D view of the regions that display the strongest average responses across conditions.

the uniform shrinkage of ridge estimators. Ridge estimators completely fail in the situation where more regressors are used, e.g. in the so-called *regional* model. Interestingly, lasso estimators have an intermediate behavior, showing that the ℓ_1 shrinkage provides a relevant proxy for subspace estimation, although it performs less well than mere dimension reduction. Unlike other settings, tree-based methods do not perform well here, probably because the sample/dimensionality ratio is not high enough. Local models outperform regional or more global models; however, we conjecture that randomized or multiscale models that average the predictions across spatial models could perform better.

The regions that show high predictability are mostly those that show high average activation across subjects and they tend to coincide with the visual-sensorimotor-auditory network described in [12], with the addition of the attentional network. Most importantly, the fact that predictability is correlated with average activation means that traditional mass-univariate random effects models used in neuroimaging studies are suboptimal, as they handle the random effects as an error term, while they actually represent more complex structures

(mismatch in the coregistration, physiological variability) and may be indicative of relevant anatomical and functional differences. A closer analysis reveals some dissociations: some regions show more variability for a given amount of average activation. This is likely because individual anatomical variability has been captured by the model, an hypothesis suggested in [11] that had not been confirmed on a functional basis so far. Individual response prediction can thus complement the traditional view of brain mapping and will pave the way toward more sensitive functional neuroimaging group analyzes.

Acknwoledgements The authors acknowledge support from the ANR grant BrainPedia ANR- 2010-JCJC-1408-01 and the Human Brain Project. They are very thankful to D. lebihan, the PI of the *archi* study, for sharing the data.

References

1. Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.J., Kanwisher, N., Golland, P.: Search for patterns of functional specificity in the brain: a nonparametric hierarchical bayesian model for group fMRI data. Neuroimage **59**(2) (Jan 2012) 1348–1368
2. Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Evans, A.C.: A general statistical analysis for fMRI data. Neuroimage **15**(1) (Jan 2002) 1–15
3. Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.B.: Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. Neuroimage **35**(1) (Mar 2007) 105–120
4. Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D.E., Saxe, R.R.: Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. Nat Neurosci **15**(2) (Feb 2012) 321–327
5. Ng, B., Abugharbieh, R., Varoquaux, G., Poline, J.B., Thirion, B.: Connectivity-informed fMRI activation detection. In: Fichtinger G., Martel A., Peters T. (eds). MICCAI 2011, Part II. LNCS, vol. 6892, pp. 285–292. Springer, Heidelberg (2011)
6. Ye, D.H., Zikic, D., Glocker, B., Criminisi, A., Konukoglu, E.: Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds). MICCAI 2013, Part I. LNCS, vol. 8149, pp. 606–613. Springer, Heidelberg (2013)
7. Dhillon, P.S., Foster, D.P., Kakade, S.M., Ungar, L.H.: A risk comparison of ordinary least squares vs ridge regression. Journal of Machine Learning Research **14** (2013) 1505–1511
8. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Machine Learning **63**(1) (2006) 3–42
9. Pedregosa, F., Varoquaux et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12** (2011) 2825–2830
10. Plaze, M., Paillère-Martinot et al.: "where do auditory hallucinations come from?"– a brain morphometry study of schizophrenia patients with inner or outer space hallucinations. Schizophr Bull **37**(1) (Jan 2011) 212–221
11. Essen, D.C.V., Glasser, M.F., Dierker, D.L., Harwell, J., Coalson, T.: Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. Cereb Cortex **22**(10) (Oct 2012) 2241–2262
12. Mesmoudi, S., Perlbarg, V., Rudrauf, D., Messe, A., Pinsard, B., Hasboun, D., Cioli, C., Marrelec, G., Toro, R., Benali, H., Burnod, Y.: Resting state networks' corticotopy: the dual intertwined rings architecture. PLoS One **8**(7) (2013) e67444