Alternative validation measure

We construct precision-recall curves by reporting the proportion of true positives in the detections (precision = 1 − False Discovery Rate, FDR) for different levels of recovery of the ground truth (the recall is the proportion of true positives among detections). The curves have to be read vertically: at a fixed level of precision, the best method is the one with the highest recall (ie. less false negatives). In practice, it is standard to choose a FDR at 5%. Precision-recall analysis constitute an alternative to ROC analysis that is better suited to unbalanced classes.

Synthetic data

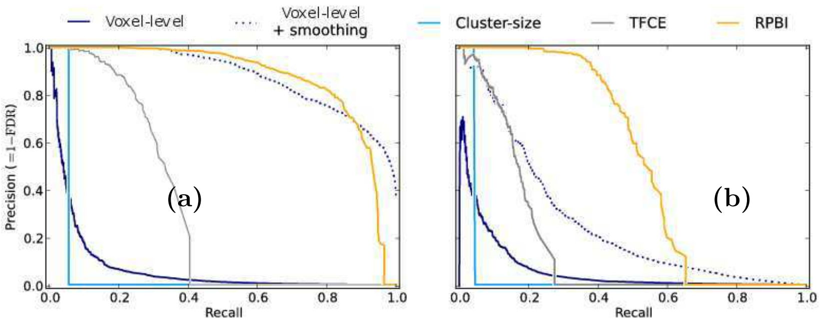

Figure 1: **Simulated data.** Precision-recall curves for various analysis methods across 10 random subsamples containing 20 subjects. SNR = 2 and noise spatial smoothness: **(a)** $\sigma_{\text{noise}} = 0$, **(b)** $\sigma_{\text{noise}} = 1$. The curves are obtained by thresholding the statistics brain maps at various levels, yielding as many points on the curves. RPBI outperforms other methods.

Real data

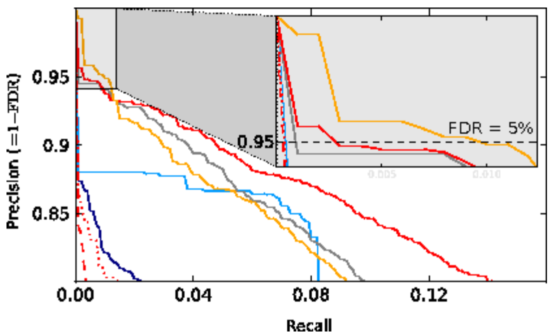

Figure 2: **Real fMRI data.** Evaluation of the performances for various analysis methods across 10 random subsamples containing 20 subjects, on a *[angry faces - control]* fMRI contrast from the *faces* protocol. ROC curves built with a pseudo ground truth where 5% of the most active voxels across 1430 subjects are kept.

Geometric parcellations

We run experiment on real data with parcellations coming from a geometric parcellation approach. We built parcellations of 1000 parcels with the K-means clustering algorithm [1] (using random initializations) on the 3D coordinates of a brain mask voxels. Geometric parcellations yield more regular parcels than those obtained by performing a Wards clustering algorithm on simulated and real data. Geometric parcellations lose the anisotropic effect of Wards parcellations. In practical terms, they do not give really good results, as compared to RPBI with Wards clustering.

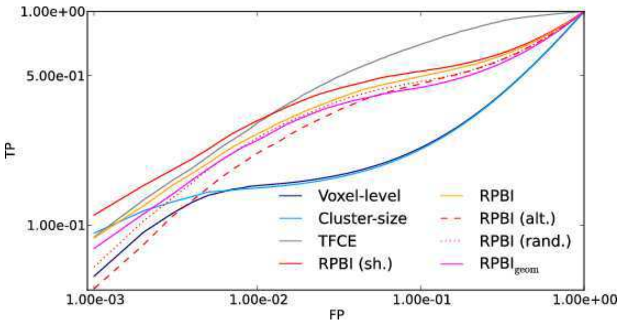

Figure 3: **Real fMRI data – RPBI with geometric parcellations.** Evaluation of the performances for various analysis methods across 10 random subsamples containing 20 subjects, on a *[angry faces - control]* fMRI contrast from the *faces* protocol. ROC curves built with a pseudo ground truth where 5% of the most active voxels across 1430 subjects are kept. RPBI with geometric parcellations has poorer performance than RPBI with Ward's clustering parcellations.

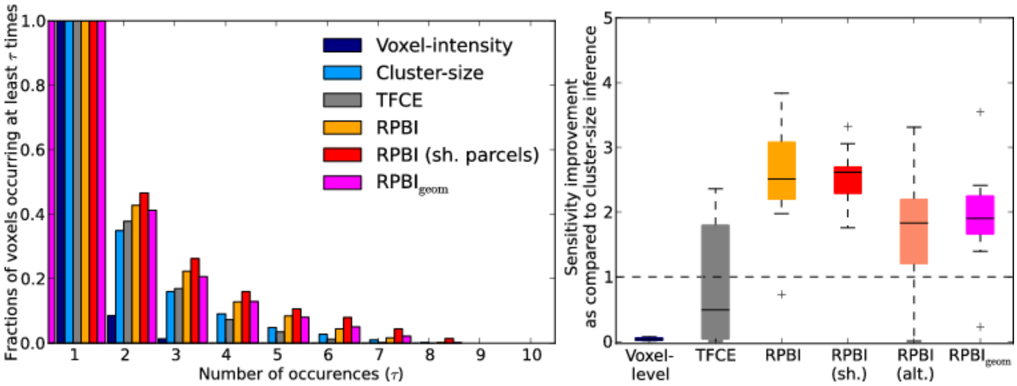

Figure 4: **(a)** Sensitivity improvement relative to cluster-size under control of the specificity at 5% FWER. **(b)** Inverse cumulative histograms of the relative number of voxels that were reported as significant several times through the 10 subsamples ($P < 0.05$ FWER corrected), on a *[angry faces - control]* fMRI contrast from the *faces* protocol. RPBI with geometric parcellations yields poor sensitivity and poor reproducibility.

References

[1] MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, California, USA. p. 14.