The NPAIRS Computational Statistics Framework for Data Analysis in Neuroimaging

Stephen Strother[1][2], Anita Oder[1], Robyn Spring[1][2], and Cheryl Grady[1]

[1] Rotman Research Institute, Baycrest
3560 Bathurst Street, Toronto, ON, Canada ∗sstrother@rotman-baycrest.on.ca
[2] Department of Medical Biophysics, University of Toronto

Abstract. We introduce the role of resampling and prediction (p) metrics for flexible discriminant modeling in neuroimaging, and highlight the importance of combining these with measurements of the reproducibility (r) of extracted brain activation patterns. Using the NPAIRS resampling framework we illustrate the use of (p, r) plots as a function of the size of the principal component subspace (Q) for a penalized discriminant analysis (PDA) to: optimize processing pipelines in functional magnetic resonance imaging (fMRI), and measure the global SNR (gSNR) and dimensionality of fMRI data sets. We show that the gSNRs of typical fMRI data sets cause the optimal Q for a PDA to often lie in a phase transition region between gSNR $\simeq 1$ with large optimal Q versus SNR $\gg 1$ with small optimal Q.

Keywords: prediction, reproducibility, penalized discriminant analysis, fMRI

1 Introduction

Mapping of brain function is a major area of brain imaging. In the 1980s it was dominated by positron emission tomography (PET) and single photon emission tomography (SPECT) but since the discovery of the blood oxygenation level dependent (BOLD) signal in the 1990's, BOLD functional magnetic resonance imaging (fMRI) and related techniques now dominate the brain imaging literature. The early PET-based applications used some machine learning and neural networks techniques for the analysis of functional neuroimages, but most the current fMRI experimental and analysis paradigms are still based on simple univariate general linear models with inferential statistical tests, and in some instances their predictive, machine learning equivalent (e.g., Gaussian Naïve Bayes, Kjems et al. (2002); Pereira et al. (2009)). However, there has been a recent explosion of interest in using related multivariate classification approaches—dubbed "mind reading" by some.

2 Data-driven performance metrics

In brain mapping it is crucial to optimize and evaluate models and to select the most salient features. These tasks must be guided by a performance met-

Y. Lechevallier, G. Saporta (eds.), *Proceedings of COMPSTAT'2010*,
DOI 10.1007/978-3-7908-2604-3_10, © Springer-Verlag Berlin Heidelberg 2010

ric. A variety of possible performance metrics including crossvalidated prediction (p) are briefly reviewed in Afshinpour et al. (in press). Although prediction accuracy alone can be an effective metric for general machine-learning problems, neuroimaging also demands that the spatial pattern (encoded by the predictive model) be reproducible (r) or generalizable between different groups of subjects or different scans of the same subject. The reproducibility of models' estimated parameters when optimizing prediction in such ill-posed data sets (variables $\gg$ observations) is a neglected issue in the field of predictive modeling. In some problems this is unimportant as prediction performance may be the primary result that matters (Schmah et al. 2008). However, in high-dimensional brain mapping problems the reliability of the extracted brain maps and the voxels that influence prediction performance are often the critical outputs of the modeling process that reflects underlying brain processes. One approach is to include a greedy search procedure because this reduces the size of the voxel feature space to the subset relevant for prediction. This may be iteratively driven by prediction metrics using classical machine learning approaches or simply based on a subset of voxels that are detected with a separate voxel-based, general linear model (GLM). Some tradeoffs of such purely prediction-driven analysis approaches are discussed in Pereira et al. (2009). Together with prediction accuracy, reproducibility is an important metric because it provides a data-driven substitute for receiver operator characteristic (ROC) analysis. We also address model performance in real data sets where the true SNR structure is unknown and ROC curves cannot be measured. In particular, we illustrate the use of (p, r) metrics to optimize the pipeline of image pre-processing steps for fMRI data sets before data analysis, e.g., scan-to-scan registration, spatial and temporal filtering, etc. (for a review see Strother (2006)). And we demonstrate the use of (p, r) metrics to optimize subspace selection for a penalized discriminant analysis (PDA) model built on a PCA basis.

3 Nonparametric, activation, influence and reproducibility resampling (NPAIRS)

NPAIRS provides a resampling framework for combining prediction metrics with the reproducibility of the brain-activation patterns, or statistical parametric maps (SPM), as a data-driven substitute for ROCs. However, any measure of similarity between patterns extracted from independent data sets is subject to an unknown bias (Afshinpour et al., in press). To obtain combined prediction and reproducibility values Strother et al. (2002); Kjems et al. (2002) proposed a novel split-half resampling framework dubbed NPAIRS and applied it first to PET and later to fMRI (see Strother et al. (2004); LaConte et al. (2003); Yourganov et al., in press). While NPAIRS may be applied to any analysis model we have focused on LDA built on a regularized PCA basis (i.e., PDA). This allows us to (1) regularize the model by choosing soft (e.g., ridge) or hard thresholds on the PCA eigenspectrum or other basis set (e.g., tensor product splines) (2) maintain the link to covariance decomposition previously used with PET for elucidating network structures, and (3) easily produce robust whole-brain activation maps useful for discovering features of brain function and/or disease.

The basic outline of NPAIRS follows[1]. Consider an fMRI data set $\mathbf{S}$ of v voxels by NT scans for N subjects' data sets of T scans each. The independent observations of N subjects are split into two independent halves $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2]$: training and test sets of size $\frac{N}{2}$. This split-half resampling represents a form of repeated, 2-fold cross-validation that has the benefits of smooth, robust metrics obtained with delete-d jackknife and the 0.632+ bootstrap (Efron and Tibshirani (1993, 1997)). Typically in neuroimaging we have $v \gg NT$, with $v = 10k - 100k$ voxels, and $N = 10$s of subjects and $T = 50 - 100s$ of scans/subject. Consequently $\mathbf{S}$ is large and ill-posed and cannot be directly inverted. Therefore, we proceed with an initial dimensionality reduction step using PCA that also serves as a preliminary denoising process. Further the PCA ensures that we have captured at least the first order voxel interactions that represent the important functional connectivity of underlying brain networks. We can obtain estimates of the PCA basis components needed using a singular value decomposition (SVD) or equivalently from the eigenvalue decomposition (EVD) of the smaller outer-product covariance matrix (which is considerably faster than an SVD). We proceed as follows

1. Given the singular SVD, $\mathbf{S} = \mathbf{ULV}^T$, we compute the EVD, $\mathbf{S}^T\mathbf{S} = \mathbf{VL}^2\mathbf{V}^T$, and proceed with a reduced basis set, $\mathbf{X}^* = \mathbf{U}^{*T}\mathbf{S} = \mathbf{L}^*\mathbf{V}^{*T}$, where we typically retain 30% of the PCA components so that $\mathbf{X}^*$ has size $(0.3NT \times NT)$, assuming $v \gg NT$.

2. Randomly partition $\mathbf{X}^*$ into two independent split-half groups across the subjects to obtain $\mathbf{X}^* = [\mathbf{X}_1, \mathbf{X}_2] = \mathbf{U}^{*T}[\mathbf{S}_1, \mathbf{S}_2]$, where $\mathbf{X}_i$ has size $(0.3NT \times N_iT)$, $N_i = N/2$ for N even, or $N_i = N/2 \pm 0.5$ for N odd.

3. Given the SVD $\mathbf{X}_i = \mathbf{Y}_i\mathbf{L}_i\mathbf{R}_i^T$, we compute second-level EVDs $\mathbf{X}_i^* = \mathbf{Y}_i^{*T}\mathbf{X}_i = \mathbf{L}_i^*\mathbf{R}_i^T$, on $\mathbf{X}_1$ and $\mathbf{X}_2$, and retain Q components from each, so that $\mathbf{X}_i^*$ has size $(Q \times T_i)$ where $T_i = N_iT$. With Q typically $\leq \min(2 - 500, 0.3NT)$ we achieve a large dimensionality (and computational) reduction. For example from Strother et al. (2004) with $N = 16$, $T = 187$ scans and $v = 23{,}389$ brain voxels, $\mathbf{S}$ is $(23{,}389 \times 2992)$, but $\mathbf{X}_i^*$ is only $Q \times 1496$, and for PDA we only calculate $(Q \times Q)$ covariances with $Q \leq 500$.

4. Now apply the prediction model separately to $\mathbf{X}_1^*$ and $\mathbf{X}_2^*$ using a scan-label structure. This label structure may directly reflect the experimental design (i.e., number of experimentally defined conditions or brain states), or it may be chosen to reflect other possibilities, such as agnostic labels that will extract an unknown but common, data-driven temporal-covariance across subjects (e.g., Strother et al. (2004); Kustra and Strother (2001); Kjems et al. (2002); Evans et al. (2010)). For the rest of this paper we focus on Canonical

[1] Software available at `http://code.google.com/p/plsnpairs/`.

Variates Analysis (CVA, Mardia et al. (1979)), which reflects a Gaussian mixture model across classes with the strong regularization constraint that all class covariances are equal and may therefore be estimated using a pooled, within-class covariance estimate; this CVA is equivalent to LDA, although we further regularize by calculating CVA on a subspace of size Q, as in a PDA (Kustra and Strother (2001)). For $g = 1, \ldots, G$ classes, and $k = 1, \ldots, K_g$, with K_g the number of scans in class g, let $\mathbf{x}_{gk}$ represent a column of $\mathbf{X}_i^*$ with Q component features of the kth scan in class g. We calculate,

$$\mathbf{W}_i = \sum\nolimits_{gk}^{GK_g} \left(\mathbf{x}_{gk} - \bar{\mathbf{x}}_g\right)\left(\mathbf{x}_{gk} - \bar{\mathbf{x}}_g\right)^T \tag{1}$$

$$\mathbf{B}_i = K_g \sum\nolimits_{g}^{G} \left(\bar{\mathbf{x}}_g - \bar{\mathbf{x}}\right)\left(\bar{\mathbf{x}}_g - \bar{\mathbf{x}}\right)^T \tag{2}$$

where $\bar{\mathbf{x}}_g = \frac{1}{K_g}\sum_k^{K_g} \mathbf{x}_{gk}$ is the mean of scans in class g, and $\bar{\mathbf{x}} = \frac{1}{T_i}\sum_{kg}^{GK_g} \mathbf{x}_{gk}$ is the mean over all scans in split-half $\mathbf{X}_i^*$. The canonical variates that represent a penalized, generalized likelihood ratio solution of the G-class discriminant problem are obtained by the following EVD:

$$\mathbf{W}_i^{-1}\mathbf{B}_i\mathbf{C}_i = \mathbf{C}_i\mathbf{M}_i \tag{3}$$

where $\mathbf{C}_i$ has $G - 1$ columns of canonical variates, $\mathbf{c}_j$ with dimension Q, normalized such that $\mathbf{C}_i^T\left(\mathbf{W}_i/\left(T_i - G\right)\right)\mathbf{C}_i = \mathbf{I}$, and $\mathbf{M}_i$ is a $(G-1) \times (G-1)$ diagonal matrix containing eigenvalues, m_j. From $\mathbf{C}_i$ we obtain PCA-like, canonical-coordinate time series defined by

$$\mathbf{Z}_i = \mathbf{X}_i^{*T}\mathbf{C}_i \tag{4}$$

where $\mathbf{Z}_i$ has $G-1$ columns of $\mathbf{z}_j$, with time-series dimension T_i, and $\mathbf{z}_j^T\mathbf{z}_h = 0$ where $(j \neq h)$, and $\mathbf{z}_j^T\mathbf{z}_j = (T_i - G)(1 + m_j)$, since $\mathbf{X}_i^*\mathbf{X}_i^{*T} = \mathbf{B}_i + \mathbf{W}_i$. The associated canonical eigenimages are given by

$$\mathbf{E}_i = \mathbf{U}^*\mathbf{Y}_i^*\mathbf{C}_i \tag{5}$$

where $\mathbf{E}_i$ has $G - 1$ columns $\mathbf{e}_j$ with dimension v.

Prediction accuracy is defined as the posterior probability of a test-scan, $\mathbf{s}_{gk(\text{test})}$, being assigned to its true class label, g, given by $p\left(g|\mathbf{s}_{gk(\text{test})}; \theta_{\text{train}}\right)$, where θ_{train} are model parameters calculated in an independent training set. Assume the scans represented by the split-half set, $\mathbf{X}_1^*$, form a training set in which we calculate the PDA model parameters in Eqn. 5. The prediction accuracy for scans in the test set, $\mathbf{X}_2^*$, is given by

$$\begin{aligned}
&p\left(\mathrm{g}_{gk(2)}\middle|\,\mathbf{s}_{\mathrm{gk}(2)};\theta_{(1)}\right) \\
&= \tfrac{1}{a}\exp\left\{-\tfrac{1}{2}\left(\mathbf{s}_{gk(2)} - \bar{\mathbf{s}}_{g(1)}\right)^T \mathbf{U}^*\mathbf{Y}_1^*\mathbf{W}_1^{-1}\mathbf{Y}_1^{*T}\mathbf{U}^{*T}\left(\mathbf{s}_{gk(2)} - \bar{\mathbf{s}}_{g(1)}\right)\right\} p(g_{gk(2)}) \\
&= \tfrac{1}{a'}\exp\left\{-\tfrac{1}{2}\left(\mathbf{s}_{gk(2)} - \bar{\mathbf{s}}_{g(1)}\right)^T \mathbf{E}_1\mathbf{E}_1^T\left(\mathbf{s}_{gk(2)} - \bar{\mathbf{s}}_{g(1)}\right)\right\} p(g_{gk(2)})
\end{aligned}$$

from Eqn. 1 with $\mathbf{C}_1\mathbf{C}_1^T = (T_i - G)\mathbf{W}_1^{-1}$, ($a$ and a' are normalizing constants). In practice we swap training and test sets and average across all scans to obtain the average prediction value for a particular split-half.

Each independent split-half PDA produces a set of canonical eigenimages, $\mathbf{E}_i$, and canonical coordinate time series, $\mathbf{Z}_i$, which can have arbitrary signs and component ordering. To address this before comparing the split-half eigenimages we perform a PDA on the full data set $\mathbf{S}$ from step 1, without splitting, using $2Q$ components from the 2nd-level EVD in steps 3 and 4. This $\mathbf{Z}_\mathbf{S}$ result provides a reference set against which we compare each $\mathbf{Z}_i$ set of canonical-coordinate time series using a Procrustes matching procedure restricted to sign changes and permutations of component order. The operations performed on the $\mathbf{Z}_i$ components are then also performed on the $\mathbf{E}_i$ components to match them across the spit-halfs. For a particular canonical component, the reproducibility of the two split-half eigenimages is defined as the correlation (r) between all pairs of the spatially aligned voxels. This correlation value r is directly related to the available SNR in each extracted pair of split-half SPMs. For transformed eigenimages of mean=0, and length=1, the two eigenvalues are equal to $1 + r$ (signal) and $1 - r$ (noise). Therefore, we define a global SNR metric for each split-half as

$$\mathrm{gSNR} = \sqrt{\left((1+\mathrm{r}) - (1-\mathrm{r})\right)/(1-\mathrm{r})} = \sqrt{2\mathrm{r}/(1-\mathrm{r})} \tag{6}$$

Note that the Procrustes matching procedure is likely to make r positive but that low-reproducibility components will still reflect the distribution of r around 0. From Eqn. 6 we see that r maps the $[0, \infty]$ range of gSNR to $[0, 1]$. In general when the number of unique split-resamplings (i.e., $\frac{1}{2}{}^N C_{N/2}$) is large enough, we perform $\gg 10$ split-halfs and record the average, or median, of the p and r distributions across for a particular choice of Q. This procedure is then repeated as a function of Q to obtain the best (p, r) values possible as a function of Q. We recognize that the resulting p-values are biased upwards as a result of optimizing model parameters (i.e, Q) using only training and validation sets, and then biased downwards, relative to leave-one-out cross-validation, as a result of using split-half resampling. Finally, we obtain a single Z-scored SPM from each split-half pair of eigenimages (i.e., rSPM(z)). In the scatter plot used to calculate r we project all pairs of voxel values onto the principal axis to obtain a consensus rSPM. These projected rSPM values are then scaled by the pooled noise estimate, $(1 - r)$, from the minor axis. As this noise estimate is uncorrelated by construction the resulting rSPM(z) values will be approximately normally distributed; in practice this is a good approximation for brain imaging. Finally, this procedure is robust to heterogeneity across the split objects (e.g., subjects) as more heterogeneous split-half pairs produce smaller r's and larger $(1 - r)$ pooled noise estimates, and thus lower rSPM(z) values than more homogeneous splits. Then we average all rSPM(z)'s to obtain a robust, consensus technique for Z-scoring any prediction model that produces voxel-based parameter estimates.

4 Measuring pipeline performance

Figure 1 plots NPAIRS (p, r) curves for an 11-class CVA of 2992 fMRI scans from 16 subjects performing a static force task (Strother et al. (2004)). The two curves reflect a small change in a single preprocessing step: the number of half cosines used for removal of low-frequency trends in fMRI time series. The points on the curves are the number of PCA components from $1 - Q$ ($Q \in \{10, 25, 75, 100, 150, 200, 300, 500\}$). A full NPAIRS analysis with 50 split-halfs was run for each value of Q. In Figure 1 as the PDA parameterization initially increases with Q, both p and r (i.e., gSNR(r)) initially increase. Then at $Q = 50$, while p continues to slowly increase, r starts to decreases quite rapidly. This appears to be a fundamental feature of predictive modeling in ill-posed neuroimaging data sets. with p typically being optimized at larger values of Q than for optimal r, but both eventually decreasing. This (p, r) tradeoff has also been demonstrated in the context of parameterization of nonlinear hemodynamic models estimated using MCMC, with r replaced by a Kullback-Leibler measure on posterior distributions (Jacobsen et al. (2008)). The (p, r) plot provides a data-driven, ROC-like space where perfect performance is represented by the upper-right-hand corner with perfect prediction (p=1) and infinite gSNR ($r = 1$). For a given set of preprocessing steps and parameters our goal is to move the (p, r) curve closer to $(1, 1)$. As this is a relative change we assume that the p-value bias is approximately constant when measuring (p, r) curves that lie closer to $(1, 1)$. We have been experimenting with using the minimum Euclidian distance from $(1, 1)$ to define an optimal (p, r) tradeoff and a cost function for processing-pipeline optimization. In Fig. 1 if we generate (p, r) curves for each of the 16 subjects and record their mean distance from $(1, 1)$, $\bar{\mathrm{M}}$, then the change, $\triangle\bar{\mathrm{M}}$, across the 16 subjects and their standard deviation may be used to judge improved processing choices. In Fig. 1 we see that on average temporal detrending with a 1.5 cycle cosine will slightly improve (p, r) performance over using a 2.0 cycle cosine.

Zhang et al. (2009) has explored this approach in the context of the same fMRI data set with both a predictive GLM and two-class PDA analysis models (2c-CVA). Table 1 summarizes her greedy search results for the impact of several pipeline processing steps. Slice-timing correction (Step 1) has no significant impact regardless of analysis model. Within-subject motion correction (Step 2) significantly improves performance for 2c-CVA, but not for GLM because of the increased inter-subject heterogeneity. As expected spatial smoothing (Step 3), and high-pass temporal filtering (Step 4) of various sorts, all significantly improve performance, but with quite different subject heterogeneity depending on the analysis model and processing technique.

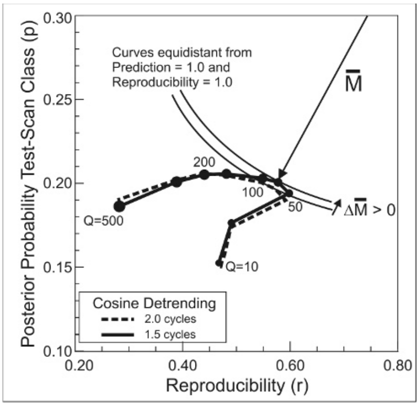
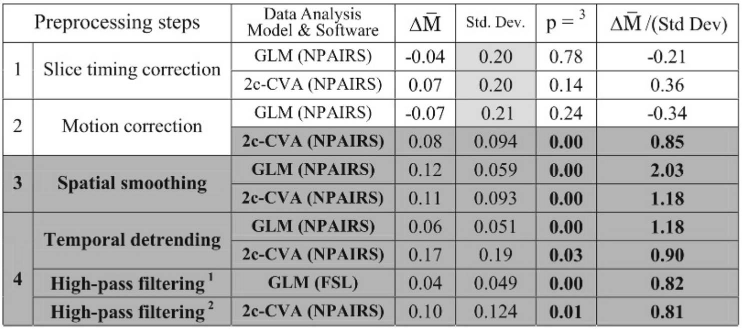


Fig. 1. NPAIRS split-half prediction (p) vs. rSPM(z) reproducibility (r) for an 11-class PDA model as a function of Q, and for a small change in low frequency temporal artefact removal: detrending with 1.5 vs. 2.0 cosine cycles per fMRI run. (Data from Strother et al. (2004)).



Preprocessing steps		Data Analysis Model & Software	$\Delta\bar{M}$	Std. Dev.	p = [3]	$\Delta\bar{M}$/(Std Dev)
1	Slice timing correction	GLM (NPAIRS)	-0.04	0.20	0.78	-0.21
		2c-CVA (NPAIRS)	0.07	0.20	0.14	0.36
2	Motion correction	GLM (NPAIRS)	-0.07	0.21	0.24	-0.34
		2c-CVA (NPAIRS)	0.08	0.094	**0.00**	**0.85**
3	**Spatial smoothing**	**GLM (NPAIRS)**	0.12	0.059	**0.00**	**2.03**
		2c-CVA (NPAIRS)	0.11	0.093	**0.00**	**1.18**
4	**Temporal detrending**	**GLM (NPAIRS)**	0.06	0.051	**0.00**	**1.18**
		2c-CVA (NPAIRS)	0.17	0.19	**0.03**	**0.90**
	High-pass filtering [1]	**GLM (FSL)**	0.04	0.049	**0.00**	**0.82**
	High-pass filtering [2]	**2c-CVA (NPAIRS)**	0.10	0.124	**0.01**	**0.81**

Table 1. Average change in optimal (p, r) curve distance from $(1, 1)$ (e.g., Fig. 1) for turning selected fMRI processing steps on and off across 16 subjects performing a parametric static force task (Zhang et al, (2008, 2009)). High-pass temporal filtering: detrending $\equiv$ removal of cosine cycles/run; **[1]**Sliding window running means. **[2]**Multi-Taper power spectrum. **[3]**Wilcoxon matched-pair per subject rank sum test

5 Measuring dimensionality

We generated 18 separate (p, r) curves from the multi-task, age-dependent data set acquired by Grady et al. (2006). The subjects belonged to three different age groups: young, middle-aged, and old. The experiment consisted of 6 separate task runs per subject of 4 memory encoding tasks (1-4), and 2 recognition tasks (5, 6). During the two recognition tasks, the subjects reported whether or not they recognized the presented stimulus. The BOLD fMRI was measured with a 1.5T MRI scanner. Standard image preprocessing was applied to the data. For each subject, one run was collected for every

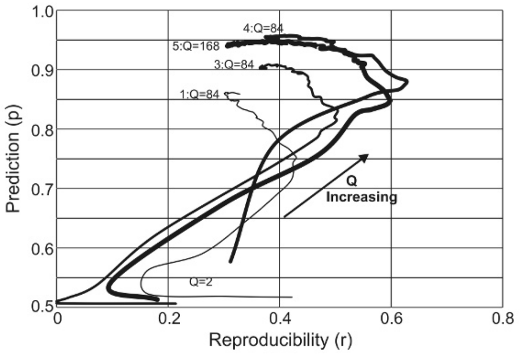


Fig. 2. NPAIRS (p, r) curves for a group of young subjects performing memory tasks: $1, 3, 4, 5$. $< Q$ is the regularizing PCA subspace of a PDA. (see text for details).

task (89 volumes for encoding tasks, and 166 volumes for recognition tasks). Each scan was described 50,308 voxels (for more details see Grady et al.).

Figure 2 shows example (p, r) curves for the 10 young subjects performing Tasks 1, 3, 4 and 5 and analyzed with a 2-class PDA to discriminate task from fixation scans. For each analysis the dimensionality of the 2nd-level PCA subspace on which the PDA was built ranged from $Q = 2$ to $Q = 84$ (Encoding tasks), and $Q = 168$ (Recognition tasks). At the largest values of Q, the PDA started to become unstable due to the large condition number (> 1000) of the within-class matrix $\mathbf{W}$.

The (p, r) curves in Fig. 2 display the same features as those in Fig. 1. For small values of increasing Q, both p and r increase until r is maximized at: Task 1, $Q = 24$; Task 3, $Q = 24$; Task 4, $Q = 12$; Task 5, $Q = 12$. In all cases p continues to rise with increasing Q, but r rapidly decreases as p is maximized at: Task 1, $Q = 76$; Task 3, $Q = 66$; Task 4, $Q = 64$; Task 5, $Q = 108$. We recorded the 18 values of Q that separately maximized r, p, and the Euclidean distance (M) from $(1, 1)$. These 54 values are plotted as a function of gSNR(r) in Figure 3. Here we see that dimensionality for optimum r (circle) and M (cross) values are often very similar, and fall on a curve with a vertical asymptote of gSNR $\simeq 1$ for $q >> 1$, and a horizontal asymptote with $Q \leq 20$ for gSNR ≥ 1.5. The horizontal asymptote with gSNR large enough (e.g., > 1.5) indicates that signal and noise eigenvalues are well separated in the eigenspectra of $\mathbf{X}_i^*$ (NPAIRS step 3), and occur in a relatively compact discrete subspace early in the PCA eigenspectrum. Conversely, the vertical asymptote indicates that as signal eigenvalues merge into the noise spectrum a phase transition occurs requiring large numbers of components from which to extract a discriminant signal, which is now relatively broadly distributed across many components of the PCA eigenspectrum.

This behavior matches recent analytic results from random matrix theory that indicate that such a phase transition occurs and is governed by the

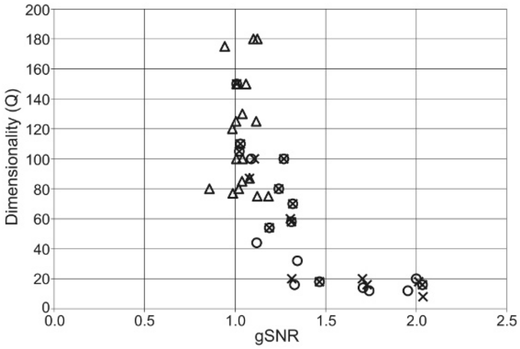


Fig. 3. For 18 NPAIRS (p, r) curves (Fig. 2) the PCA subspace size, Q, is plotted against gSNR(r) for optimal (1) prediction "$\triangle$", (2) reproducibility "O," and (3) Euclidean distance "X" (see Fig. 1). (see text for details)

ratio of variables (i.e., voxels) to observations (i.e., scans) for a particular signal strength. We have recently compared measurement of Q, across simulated and fMRI-data phase transitions with multiple dimensionality estimation approaches proposed in the literature (e.g., optimization of Bayesian evidence, Akaike information criterion, minimum description length, supervised and unsupervised prediction, and Stein's unbiased risk estimator: Yourganov et al., in press). None of the alternate approaches detect the phase transition indicating that they are suboptimal to obtain activation maps with $v >> NT$.

Figure 3 shows that there is a shift of the distribution of Q values for maximum prediction towards higher dimensionality at a gSNR value of approximately 1. This suggests that irrespective of the underlying signal eigenstructure reflected in the possible gSNR (i.e., horizontal asymptote), optimal prediction tends to select a smaller gSNR with a solution typically built from a large number of PCA components. Examination of the associated rSPM(z) for maximum prediction shows that the reduced gSNR is partly a result of a reduced number of signal voxels (e.g., rSPM$(z)> 3$) compared to rSPM(z) for optimal reproducibility. We are exploring the possibly that this reflects the tendency for prediction to select low reliability voxel sets. It remains an unresolved and important issue whether or not optimal prediction based on preliminary voxel-based feature selection or recursive feature selection can detect highly reliable spatial patterns in neuroimaging. Our PDA results suggest that this may not be the case for linear multivariate models.

References

AFSHINPOUR B, HAMID S-Z, GHOLAM-ALI H-Z, GRADY C, and STROTHER SC. (In press). Mutual Information Based Metrics for Evaluation of fMRI Data Processing Approaches. *Hum Brain Mapp.*

EFRON B, and TIBSHIRANI R. (1993). *An Introduction to the Bootstrap.* London, U.K.: Chapman & Hall.

EFRON B, and TIBSHIRANI R. (1997). Improvements on cross-validation: The .632+ bootstrap method. *J. Amer. Statist. Assoc. 92*, 548–560.

EVANS JW, TODD RM, TAYLOR MJ, and STROTHER SC. (2010). Group specific optimisation of fMRI processing steps for child and adult data. *NeuroImage, 50*, 479–490.

GRADY C, SPRINGER M, HONGWANISHKUL D, MCINTOSH A, and WINOCUR G. (2006). Age-Related Changes in Brain Activity across the Adult Lifespan: A Failure of Inhibition? *J Cogn Neurosci, 18*, 227–241.

JACOBSEN D.J., HANSEN L.K. and MADSEN K.H. (2008): Bayesian model comparison in nonlinear BOLD fMRI hemodynamics. *Neural Comput, 20, 738–55.*

KJEMS U., HANSEN L.K., ANDERSON J., FRUTIGER S., MULEY S., SIDTIS J., ROTTENBERG D. and STROTHER S.C. (2002): The quantitative evaluation of functional neuroimaging experiments: mutual information learning curves. *NeuroImage, 15, 772–86.*

KUSTRA R. and STROTHER S.C. (2001). Penalized discriminant analysis of [15O]-water PET brain images with prediction error selection of smoothness and regularization hyperparameters. *IEEE Trans Med Imaging, 20, 376–87.*

LACONTE S., ANDERSON J., MULEY S., ASHE J., FRUTIGER S., REHM K., HANSEN L.K., YACOUB E., HU X., ROTTENBERG D. and STROTHER S. (2003): The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. *NeuroImage, 18, 10–27.*

MARDIA K.V., KENT J.T. and BIBBY J.M. (1979): *Multivariate Analysis.* San Diego: Academic Press.

PEREIRA F., MITCHELL T., and BOTVINICK M. (2009): Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage, 45, 199–209.*

SCHMAH T., HINTON G., ZEMEL R.S., SMALL S.L., and STROTHER S.C. (2008): Generative versus discriminative training of RBMs for classification of fMRI images. *Neural Information Processing Systems.* Vancouver, Canada. p 1409–1416.

STROTHER S.C. (2006): Evaluating fMRI preprocessing pipelines. *IEEE Eng Med Biol Mag, 25, 27–41.*

STROTHER S.C., ANDERSON J., HANSEN L.K., KJEMS U., KUSTRA R., SIDTIS J., FRUTIGER S., MULEY S., LACONTE S. and ROTTENBERG D. (2002): The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage, 15, 747–71.*

STROTHER S.C., LA CONTE S., KAI HANSEN L., ANDERSON J., ZHANG J., PULAPURA S. and ROTTENBERG D. (2004): Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. *NeuroImage, 23, 196–207.*

YOURGANOV G., XU C., LUKIC A., GRADY C., SMALL S., WERNICK M. and STROTHER S.C. (In press): Dimensionality estimation for optimal detection of functional networks in BOLD fMRI data. *NeuroImage.*

ZHANG J., ANDERSON J.R., LIANG L., PULAPURA S.K., GATEWOOD L., ROTTENBERG D.A. and STROTHER S.C. (2009): Evaluation and optimization of fMRI single-subject processing pipelines with NPAIRS and second-level CVA. *Magn Reson Imaging, 27, 264–78.*