The Statistician and the Scientist

A statistician and a scientist are going to be executed, and the executioner asks each for their last request.

When asked, the statistician says that he'd like to give one last lecture on his theory of statistics.

When the scientist is asked, he says, "I'd like to be shot first!"

Rob Tibshirani @ S-Plus Users conference Oct., 1999.

http://www-stat.stanford.edu/~tibs

Plurality, Quality Metrics and Consensus in the Analysis of fMRI Data Sets

Stephen C. Strother[1,2,4], Stephen LaConte[4,6], Jon R. Anderson[2], Suraj A. Muley[2,3], James Ashe[2,3], Sujit Pulapura[5], Essa Yacob[1,6], Xiaoping Hu[1,6], Lars Kai Hansen[7], David Rottenberg[1,2,3]

[1]Radiology and [2]Neurology Departments., University of Minnesota
[3]Neurology Department, VA Medical Center, Minneapolis
[4]Biomedical Engineering, University of Minnesota
[5]Computer and Electrical Engineering,
[6]Center for Magnetic Resonance Research, University of Minnesota
[7]Institute for Mathematical Modeling, Danish Technical University

Principal Funding Sources: NIH Human Brain Project, P20-MH57180.

Outline

- Plurality and the functional neuroimaging data analysis problem.
- Plurality, Sample Size, and Bias-Variance.
- Simulation of univariate versus multivariate plurality.
- Dealing with plurality in a machine/ statistical learning framework.
- Quality metrics in the NPAIRS resampling framework with Canonical Variates Analysis.
- Multi-dimensional parametric static force results from a 16-subject BOLD fMRI data set.
  - Preprocessing and quality metrics with CVA
  - CVA versus GLM
- Consensus as a solution to Plurality.

The Data Analysis Problem

- We collect a set of functional neuroimaging scans (high-dimensional multivariate image vectors) with an unknown spatio-temporal structure.
- Each scan is acquired under one of a finite set of experimental design conditions, or ***brain states***, that may be defined in a ***design matrix***.
- ***PROBLEM:*** How should we determine the spatio-temporal structure that "best" describes the variation among these experimental brain states?

Plurality is a Pervasive Feature of the Functional Neuroimaging Literature

"Functional magnetic resonance imaging (fMRI) is currently studied through use of many different signal processing strategies; see, e.g., Petersson *et al.* (1999a,b) and Lange *et al.* (1999). Some strategies emphasize spatial aspects, others focus on temporal aspects; some strategies are formulated as hypothesis tests others are exploratory in nature. Most analysis schemes involve a notion of activation strength. …"

FROM: Hansen LK, Nielsen FA, Strother SC, Lange N. Consensus Inference in Neuroimaging. Neuroimage 13:1212-1218, 2001.

Petersson, K.M., Nichols, T.E., Poline, J.B., Holmes, A.P. 1999a. Statistical limitations in functional neuroimaging. I. Non-inferential methods and statistical models. *Phil. Trans. Royal Soc. - Series B: Biological Sciences* **354**:1239-60

Petersson, K.M., Nichols, T.E., Poline, J.B., Holmes, A.P. 1999b. Statistical limitations in functional neuroimaging. II. Signal detection and statistical inference. *Phil. Trans. Royal Soc. - Series B: Biological Sciences* **354**:1261-81

Lange, N., Strother, S.C., Anderson, J.R., Nielsen, F.A., Holmes, A., Kolenda, T., Savoy, R., Hansen, L.K. 1999. Plurality and resemblance in fMRI data analysis. *Neuroimage* **10**:282-303.

Some Philosophy

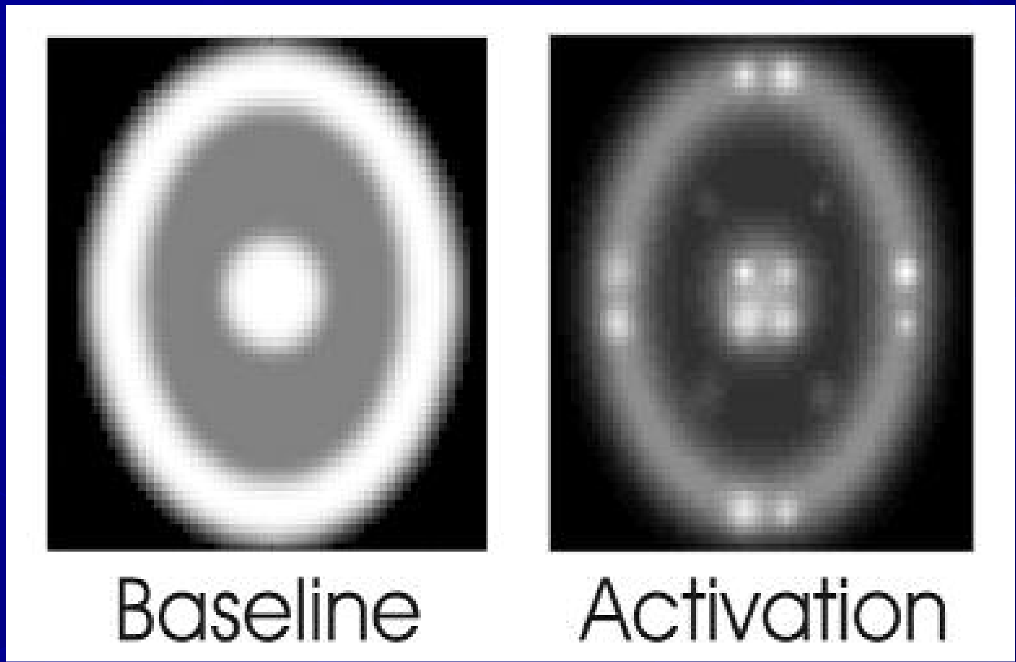
Data-analysis philosophy that leads to a pluralistic viewpoint:

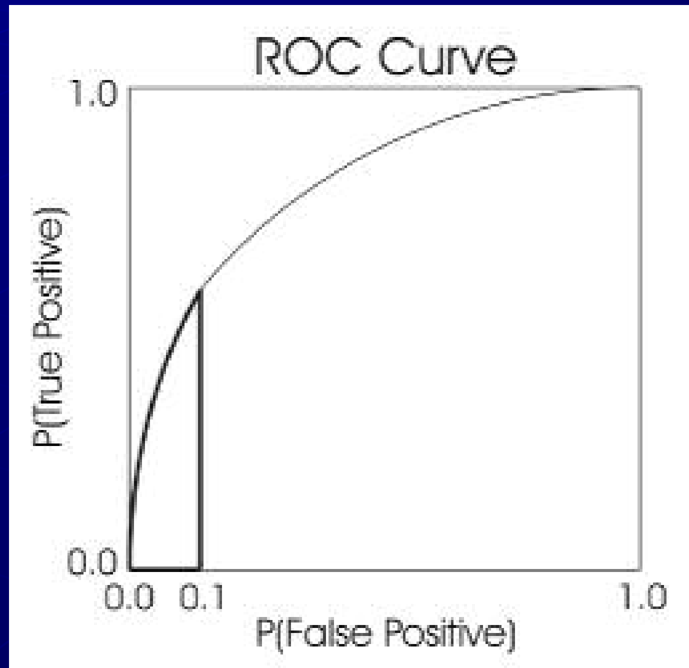
- "All models (data analysis approaches) are wrong, but some are useful!"
  - "All models are wrong." G.E. Box (1976) quoted by Marks Nester in, "An applied statistician's creed," Applied Statistics, 45(4):401-410, 1996.
- "I believe in ignorance based methods because humans have a lot of ignorance and we should play to our strong suit."
  - Eric Lander, Whitehead Institute, M.I.T.

Plurality, Sample Size and the Bias-Variance Problem

- Traditional statistical parameter estimation (e.g., Maximum Likelihood, ML) focuses on asymptotically unbiased, minimum-variance estimates.
- We have no idea what large ≈ asymptotic means in real, finite, functional neuroimaging data sets.
- Compared to ML-based estimators there are better signal detectors that are asymptotically-biased but have smaller parameter variance in finite samples.
- In all real, finite data sets there is a bias-variance tradeoff to be considered across a plurality of models.

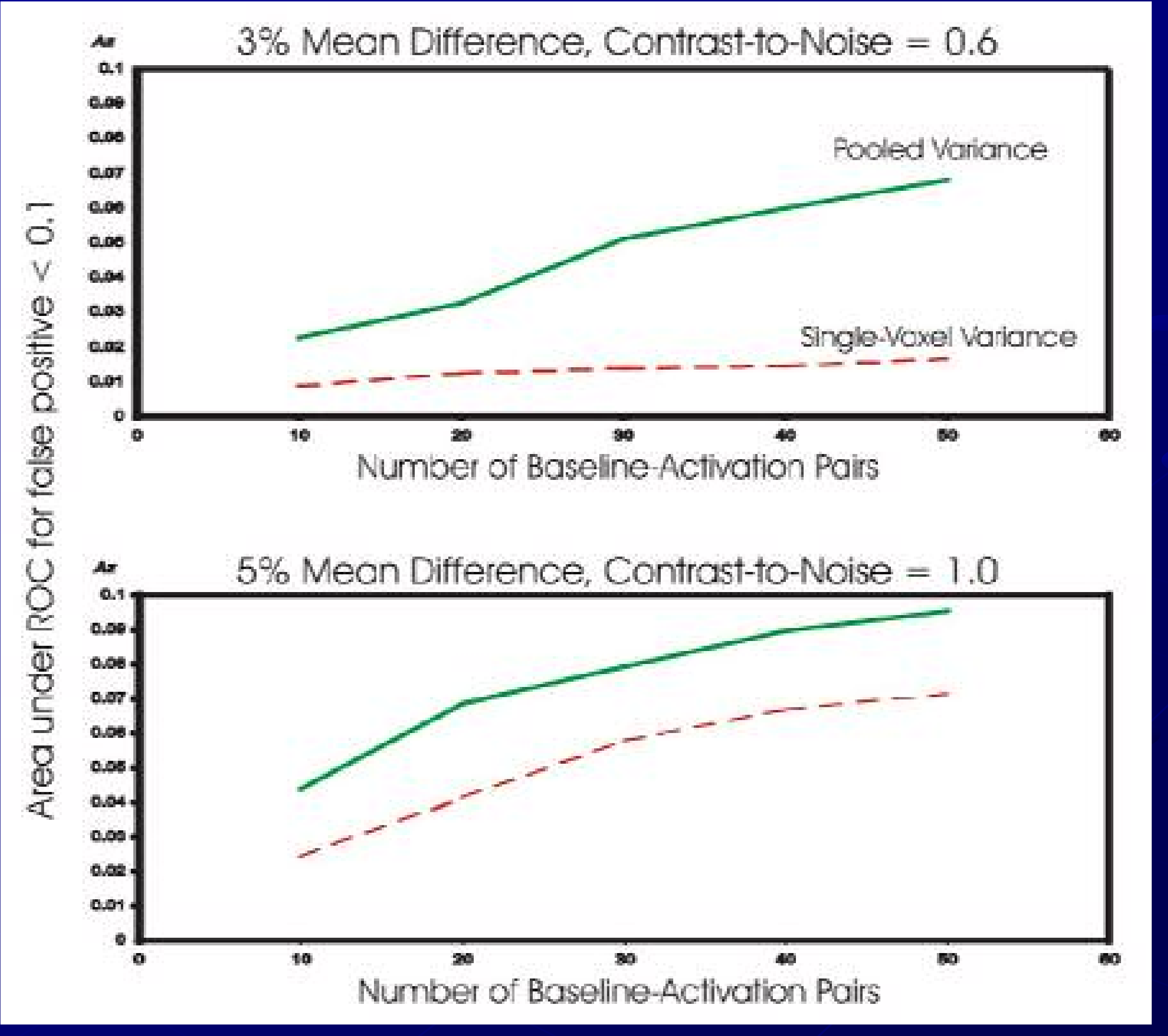
Simulation of Simple Bias-Variance Tradeoffs



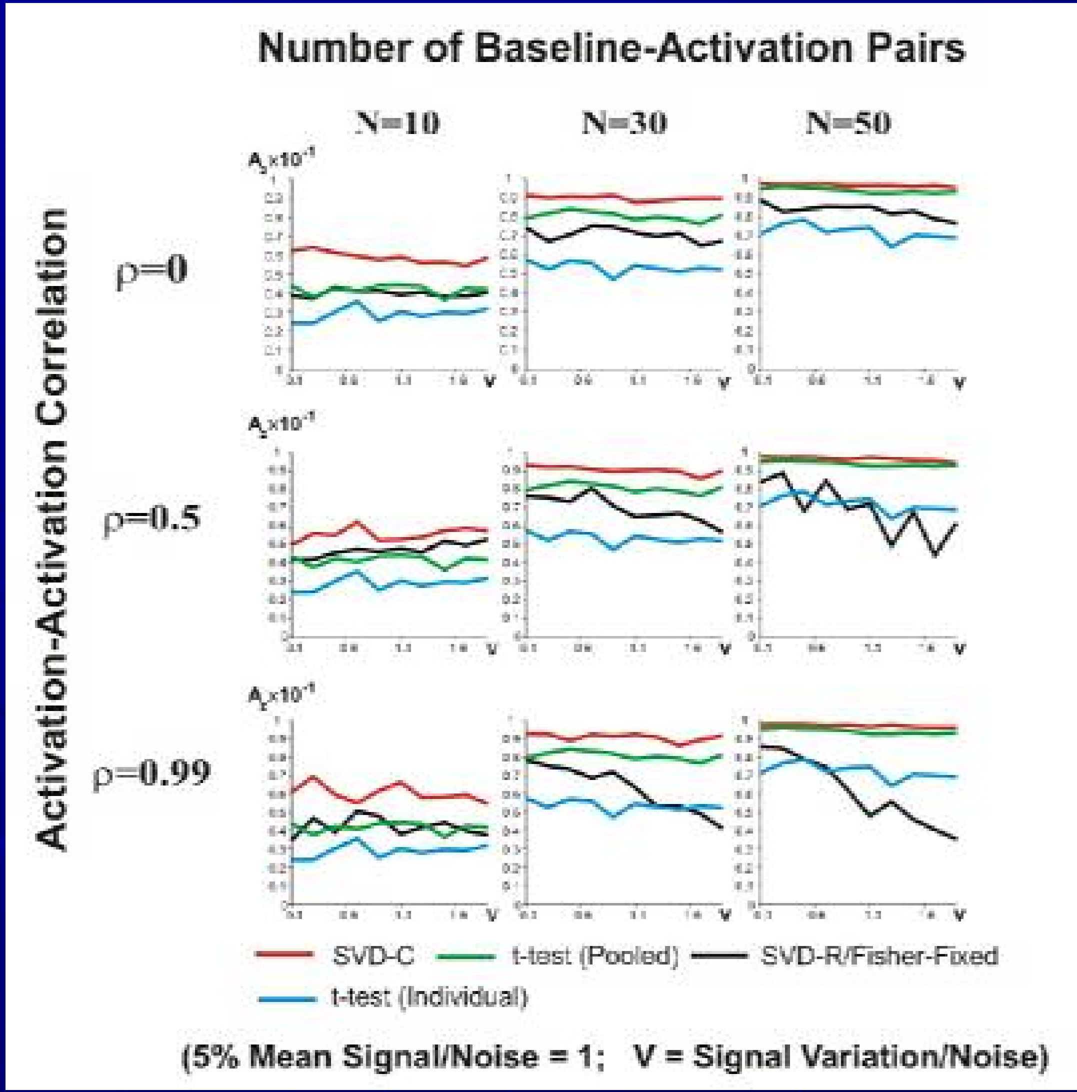


FROM: Lukic AS, Wernick MN, Strother SC. "An evaluation of methods for detecting brain activations from PET or fMRI images." Artificial Intelligence in Medicine, 25:69-88, 2002.

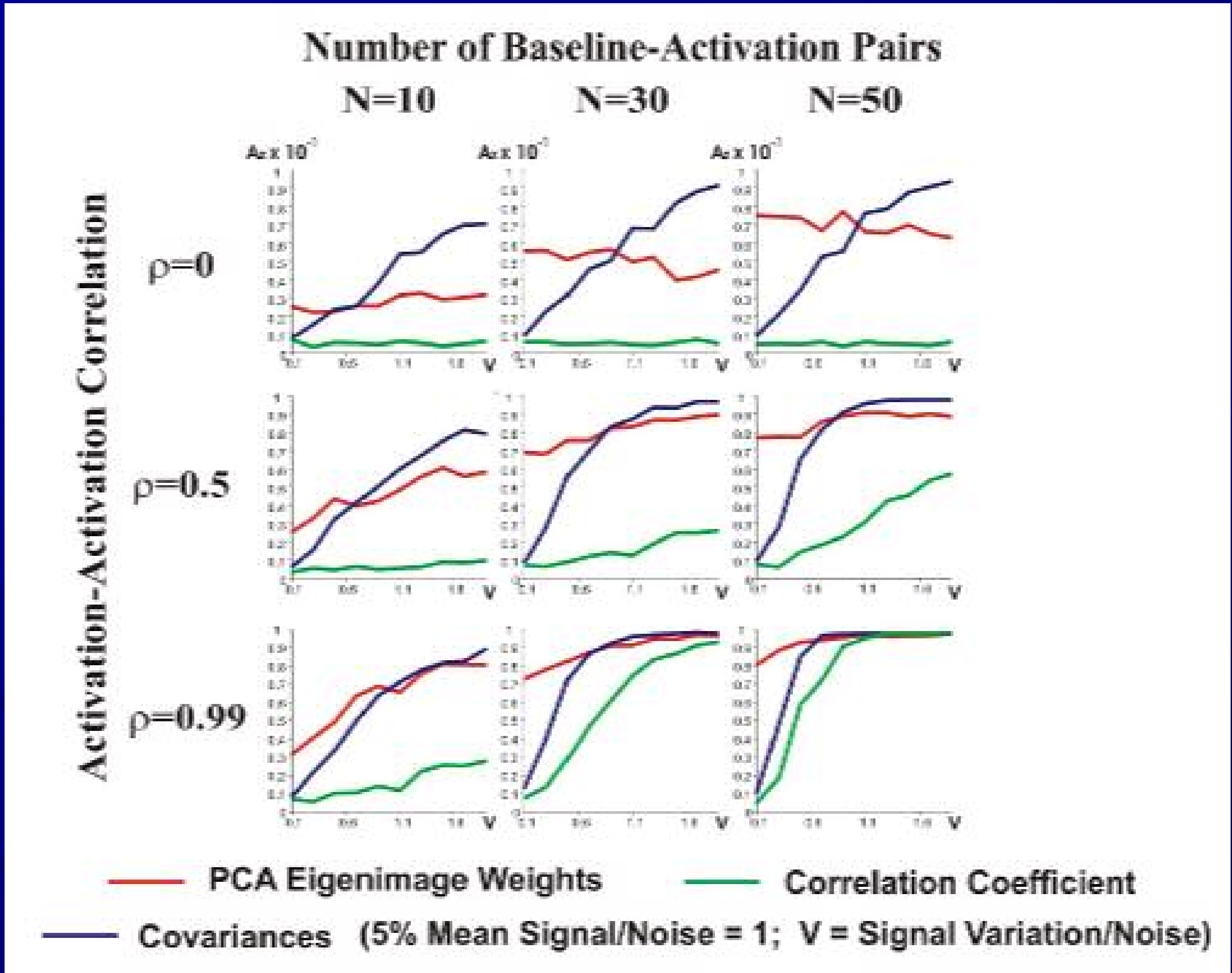
Bias-Variance of Mean-Difference Detection



Bias-Variance of Mean-Difference Detection



Bias-Variance of Covariation Detection

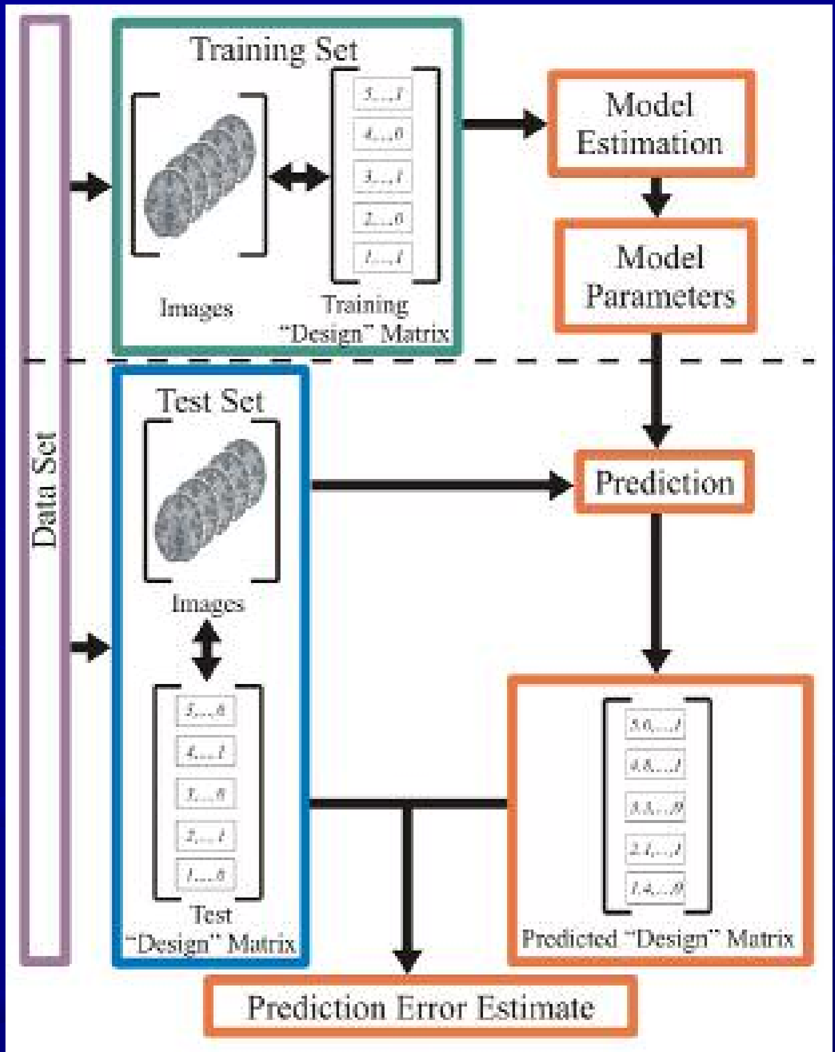


The Plurality of Preprocessing Meta-Models: Neuroscientific Bias, Metrics and Data Consistency

- Consider meta-models that include all experimental and methodological choices in the fMRI data analysis chain.
- We have only a limited understanding of the relative importance of the choices that generate this plurality of meta-models.
- The generation of a "neuroscientifically plausible" result is typically used to justify the meta-model choices made, encouraging a systematic bias towards prevailing neuroscientific expectations.
- This issue may be addressed by using prediction and reproducibility metrics in a resampling framework to test internal data consistency, independent of neuroscientific expectations.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning theory. Springer-Verlag, New York, 2001

Prediction/Crossvalidation reSampling



Morch N, Hansen LK, Strother SC, Svarer C, Rottenberg DA, Lautrup B, Savoy R, Paulson OB. *Nonlinear versus linear models in functional neuroimaging: Learning curves and generalization crossover*. In: Duncan J, Gindi G, eds: Lecture Notes in Computer Science 1230: Information Processing in Medical Imaging. Springer-Verlag, 1997, 259-270.

Hansen LK, Larsen J, Nielsen FA, Strother SC, Rostrup E, Savoy R, Lange N, Sidtis J, Svarer C, Paulson OB. *Generalizable patterns in neuroimaging: How many principal components?.* Neuroimage, 9:534-544, 1999.

Kustra R, Strother SC. *Penalized discriminant analysis of [15O]water PET brain images with prediction error selection of smoothing and regularization hyperparameters.* IEEE Trans Med Img 20:376-387, 2001.

The NPAIRS Framework

- Uses *"split-half"* resampling to provide:
  - measurements of prediction (generalization) error and SPM pattern reproducibility (reliability);
  - uncorrelated signal and noise SPMs from any data analysis model;
  - a reproducible SPM (rSPM) on a common statistical Z-score scale;
  - provides an empirical random effects correction;
  - a measure of individual observation influence.

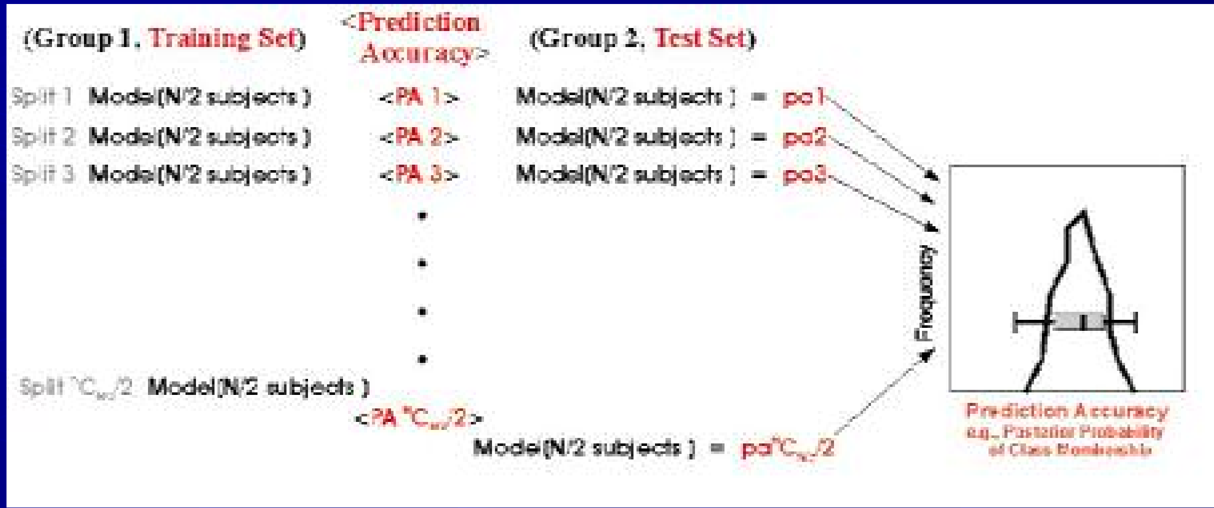
N Nonparametric
P Prediction
A Activation
I Influence
R Reproducibility
S reSampling

Strother SC, Anderson J, Hansen LK, Kjems U, Kustra R, Siditis J, Frutiger S, Muley S, LaConte S, Rottenberg D. 2002. The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. *Neuroimage* 15:747-771.

NPAIRS: Split-half reSampling

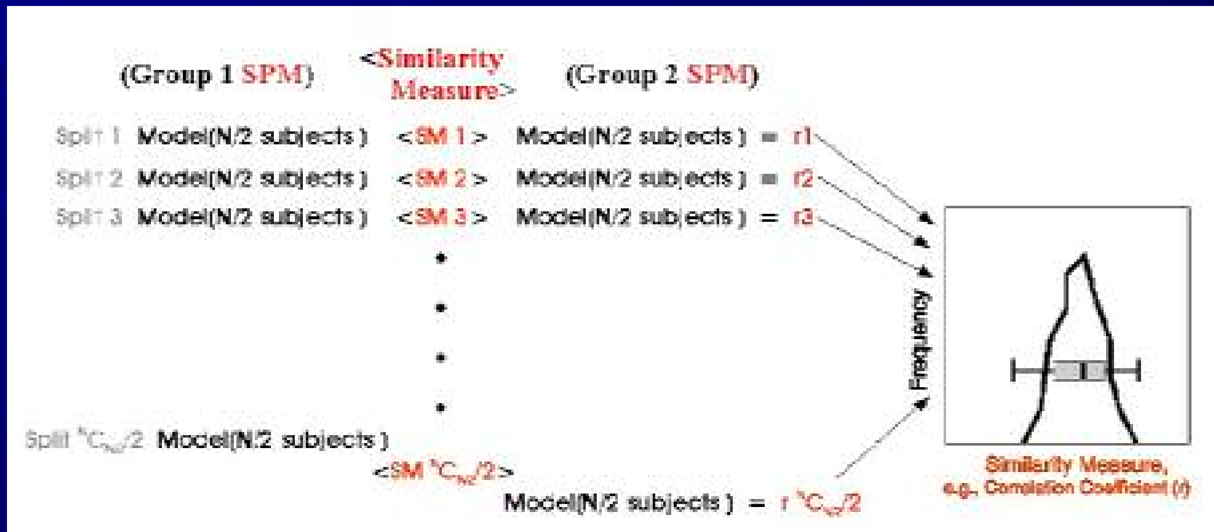
Prediction of

Design Matrix

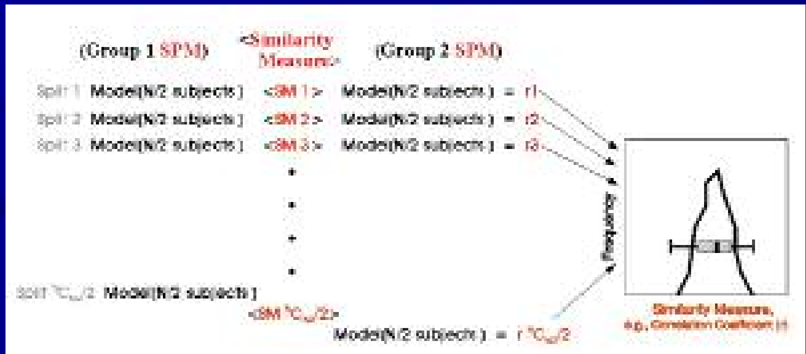


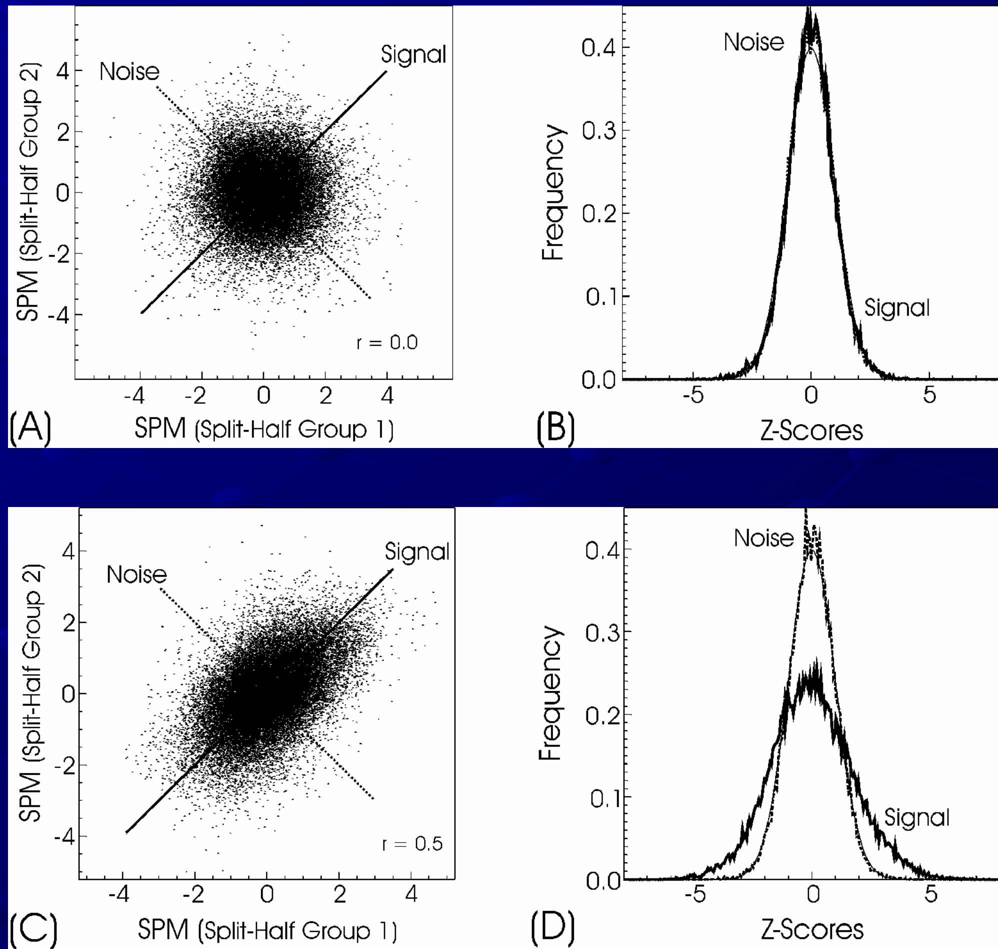
Reproducibility

of SPMs



NPAIRS: Split-half reSampling to Obtain Activation-Pattern Reproducibility Metrics





© S. C. Strother, 2002

A Flexible Multivariate-Analysis Approach for NPAIRS

- Canonical Variates Analysis:
  - Closely related to Linear Discriminant Analysis, Canonical Correlation Analysis and Partial Least Squares;
  - Maximizes the multivariate signal-to-noise ratio of (Between-Class)/(Pooled Within-Class) covariance;
  - Provides an approximate correction for random subject effects;
  - Efficiently detects mean AND spatial interaction signals;
  - Easily vary model complexity with:
    - experimental state-driven or exploratory data-driven class structures;
    - regularization of different types/numbers of basis functions.

Mardia, K.V., Kent, J.T., Bibby, J.M. *Multivariate analysis.* Academic Press, 1979

Kustra R, Strother SC. Penalized discriminant analysis of [15O]water PET brain images with prediction error selection of smoothing and regularization hyperparameters. IEEE Trans Med Img 20:376-387, 2001.

NPAIRS: Reproducibility Metrics in Functional Neuroimaging Studies

- Strother SC, Lange N, Anderson JR, Schaper KA, Rehm K, Hansen LK, Rottenberg DA. Activation pattern reproducibility: Measuring the effects of group size and data analysis models. Hum Brain Mapp, 5:312-316, 1997.
- Frutiger S, Strother SC, Anderson JR, Sidtis JJ, Arnold JB, Rottenberg DA. Multivariate predictive relationship between kinematic and functional activation patterns in a PET study of visuomotor learning. Neuroimage 12:515-527, 2000.
- Muley SA, Strother SC, Ashe J, Frutiger SA, Anderson JR, Sidtis JJ, Rottenberg DA. Effects of changes in experimental design on PET studies of isometric force. Neuroimage 13:185-195, 2001.
- Shaw M, Strother SC, McFarlane AC, Morris P, Anderson J, Clark CR, Egan GF. Abnormal functional connectivity in post-traumatic stress disorder. Neuroimage 15:661-674, 2002.
- Tegeler C, Strother SC, Anderson JR, Kim S-G. Reproducibility of BOLD-based functional MRI obtained at 4T. Hum Brain Mapp, 7:267-283, 1999.
- LaConte S, Anderson J, Muley S, Frutiger S, Hansen LK, Yacoub E, Xiaoping H, Rottenberg D, Ashe J, Strother SC. Evaluating preprocessing choices in single-subject BOLD-fMRI studies using data-driven performance metrics. Neuroimage **18:**10-23, 2003

Predicting the Brain State with CVA

$$p(c^{(j)} \mid \mathbf{x}_{te}^{(j)};\theta_{tr}) = \frac{1}{K}\exp\left\{-\frac{1}{2}\left\| \mathbf{L}_{tr}^{T}\left(\mathbf{U}_{tr}^{*}\right)^{T}\left(\mathbf{x}_{te}^{(j)} - \overline{\mathbf{x}}_{tr}^{(\mathbf{c}^{(\bullet)})}\right)\right\|^{2}\right\} p(c^{(j)})$$

- Identifies the regions needed to explain systematic variations between scans by linearly combining with a new scan to predict the experimental state of the brain, i.e. the class of the new test scan.
  - The probability of predicting the class, c, of a new scan $p(c^{(j)} \mid \mathbf{x}_{te}^{(j)};\theta_{tr})$
  - Is a weighted, multivariate Gaussian distribution $\frac{1}{K}\exp\left\{-\frac{1}{2}\| \quad \|^{2}\right\} p(c^{(j)})$
  - Dependent on the Euclidean distance $\| \quad \|^{2}$
  - Between the training class mean and the new scan $\left(\mathbf{x}_{te}^{(j)} - \overline{\mathbf{x}}_{tr}^{(\mathbf{c}^{(\bullet)})}\right)$
  - Projected onto a set of non-orthogonal canonical eigenimages $\mathbf{L}_{tr}^{T}\left(\mathbf{U}_{tr}^{*}\right)^{T}$
  - With flexibly chosen type and number of basis functions $\mathbf{U}_{tr}^{*}$

Static-Force, BOLD fMRI Data Set

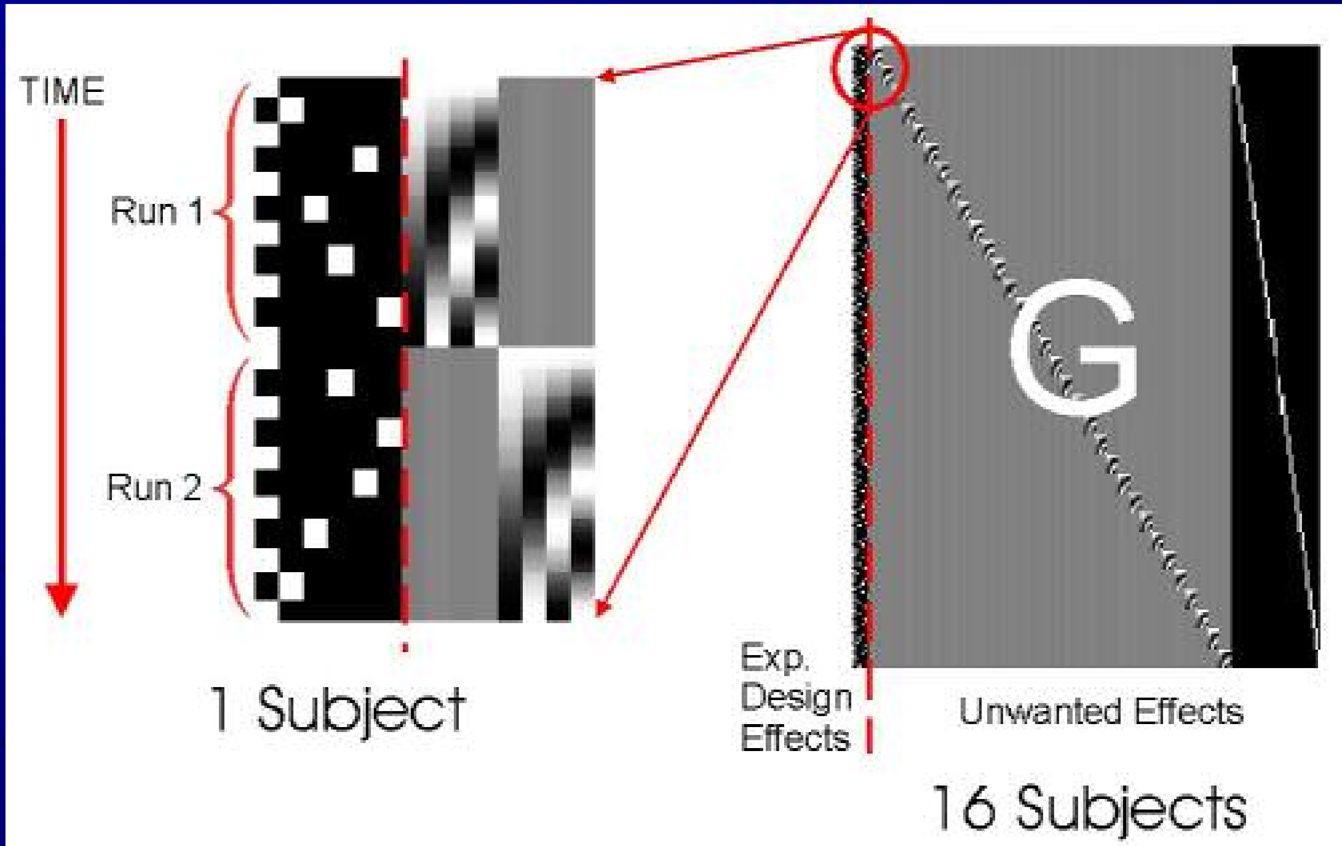
- Sixteen subjects with 2 runs/subject
- ***Acquisition:***
  - Whole-brain, interleaved 1.5T BOLD-EPI;
  - 30 slices = 1 whole-brain scan;
  - 1 oblique slice = 3.44 x 3.44 x 5 mm^3;
  - TR/TE = 4000 ms/70 ms
- ***Experimental Design:***
  - Parametric static isometric force (sf);
  - Run: [5 x (b1, ... , b11, sf#1, ... , sf#11), b1, ... , b11] = 121 scans;
  - sf1=200g, sf2=400g, sf3=600g, sf4=800g, sf5=1000g.
- ***Analyzed with PCA and Penalized CVA:***
  - Single-subject 2-class and 22-class analyses;
  - 16-subject, 11-class group analysis;
  - Dropped initial non-equilibrium and state-transition scans.

Preprocessing for Static Force

- **All runs/subject(s) passed initial quality control:**
  - movement (AIR 3) < 1 voxel;
  - no artifacts in functional or structural scans;
  - no obvious outliers in PCA of centered data matrix.
- **Within-Subject Alignment:**
  - None;
  - Across runs using AIR 3.08 to 1st scan of run one.
- **Temporal Detrending using GLM Cosine Basis (SPM):**
  - None;
  - 0.5, 2.0 cosines/run.
- **Spatial Smoothing with 2D Gaussian:**
  - None;
  - FWHM = 1, 1.5, 2, 3, 4, 6, 8 pixels (3.44 mm)
  - FWHM = 1.5 voxels = 0.52 mm; FWHM = 6 voxels = 21 mm.

GLM Design Matrix

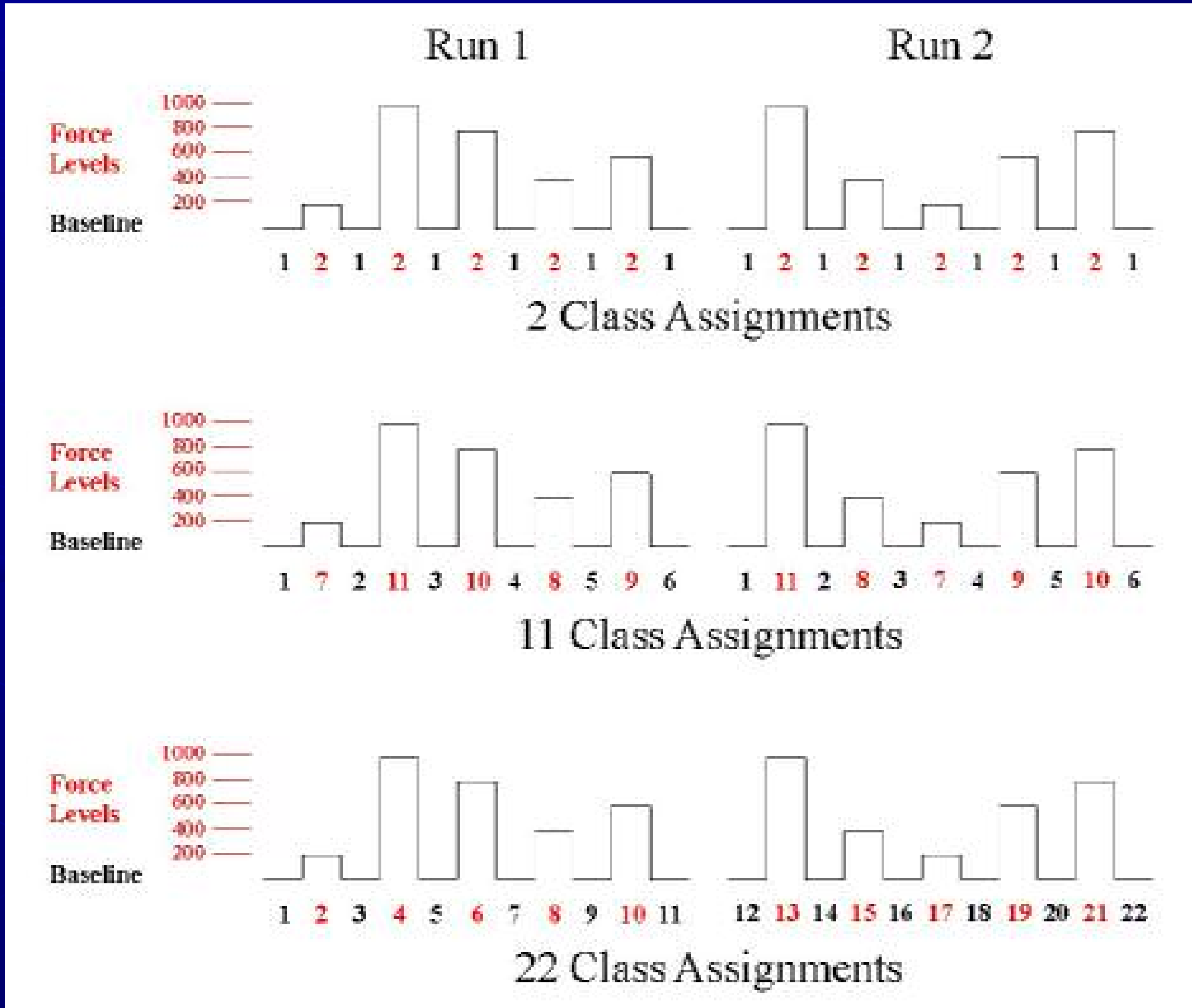
$$Y_{[\text{Subject(time) x Voxels}]} = G_{[\text{Subject(time) x Effects}]} \times B_{[\text{Effects x Voxels}]} + \text{error}$$



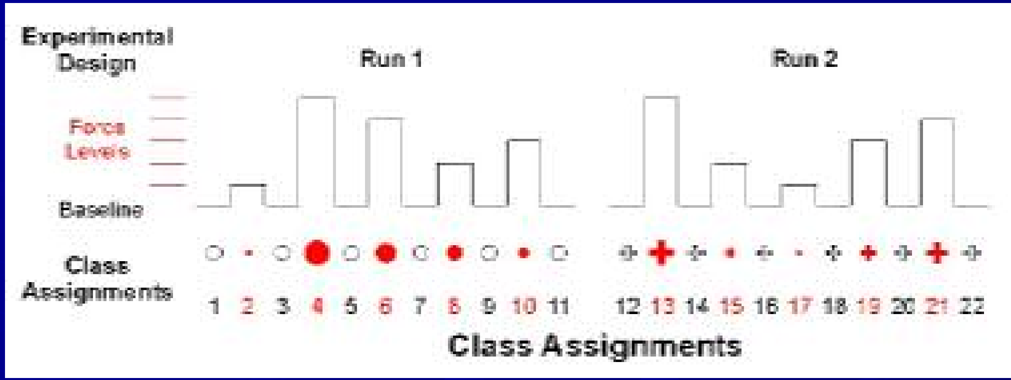
Static Force Class Assignments

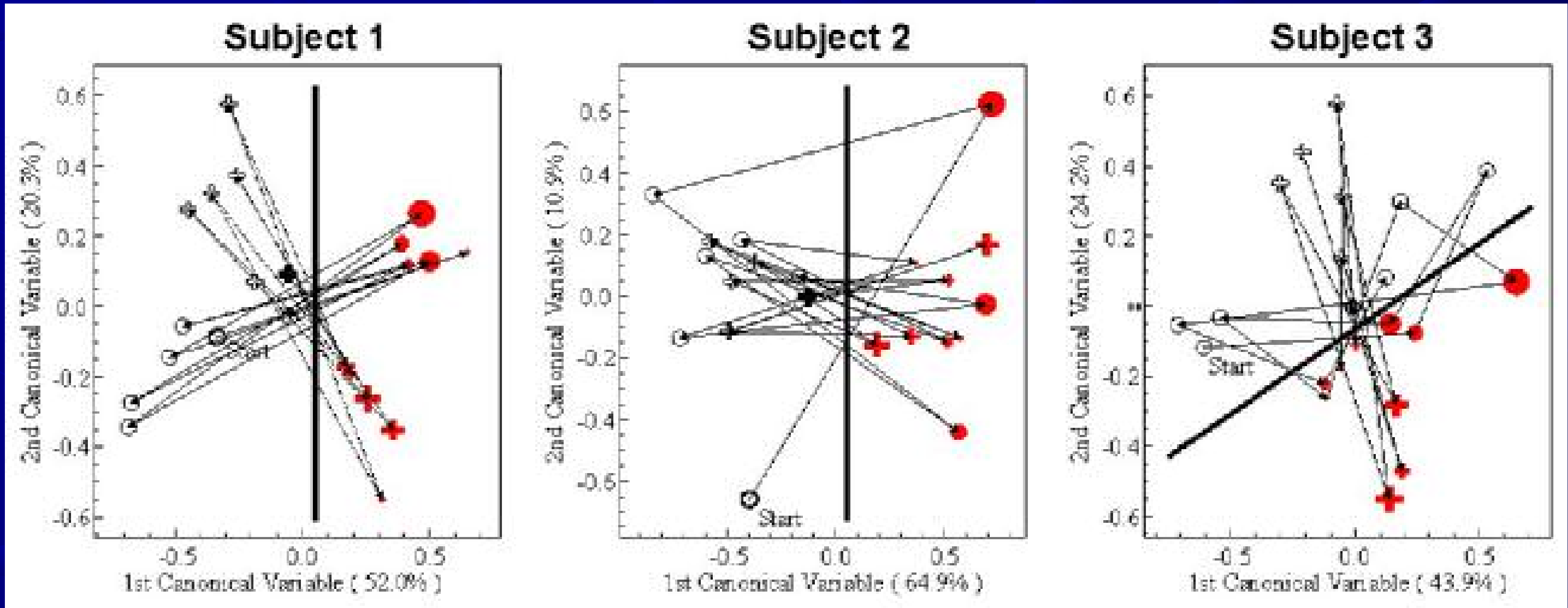
Experimental State-Driven Classes

Exploratory Data-Driven Classes



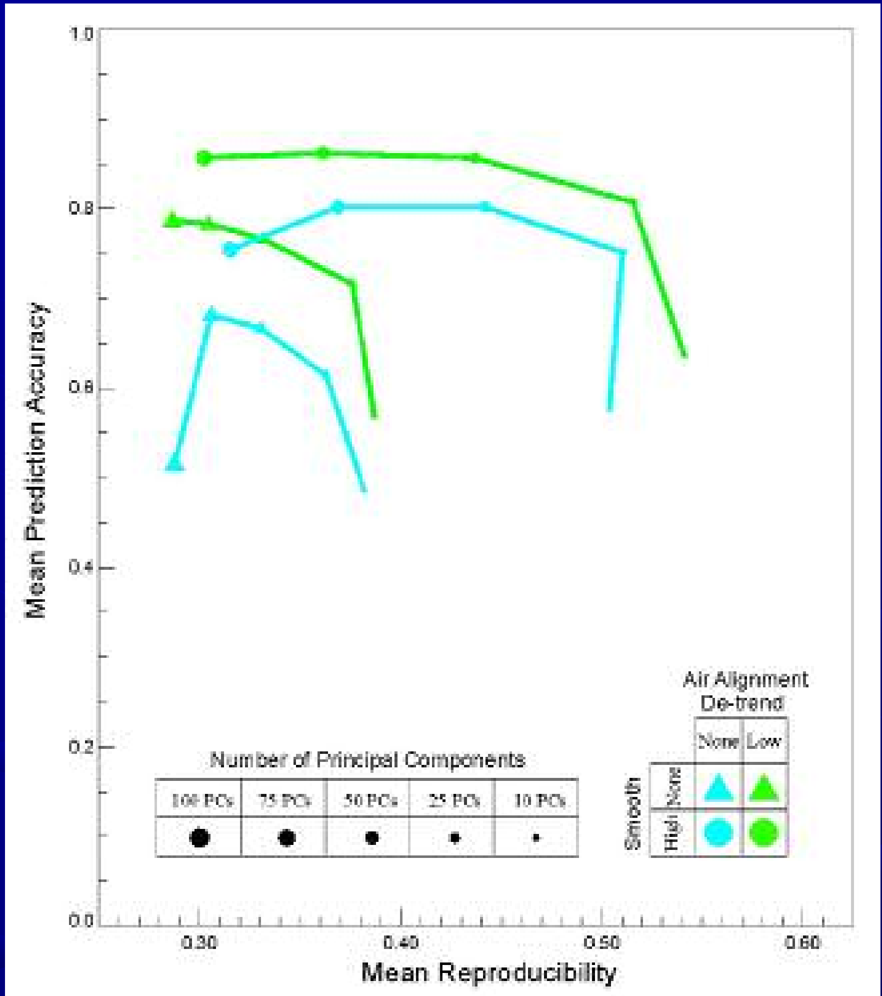
Single-Subject, 22-Class Exploratory CVA





Prediction vs. Reproducibility

(2-Class CVA/Subject)

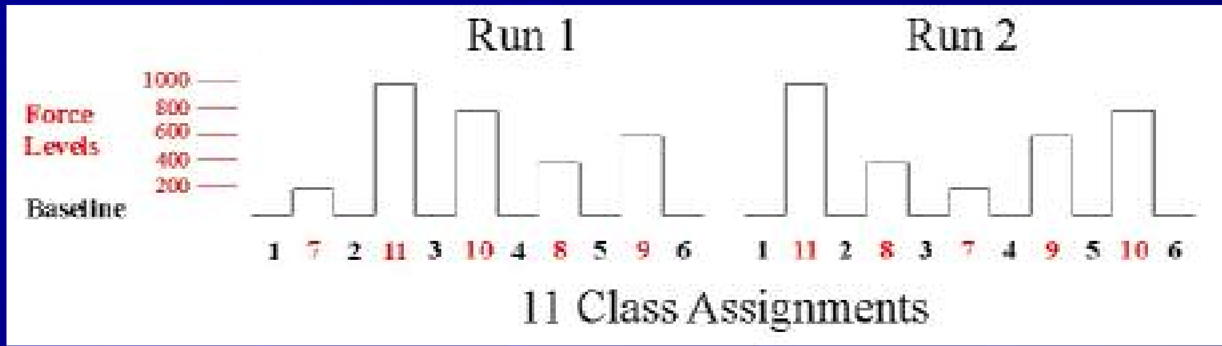


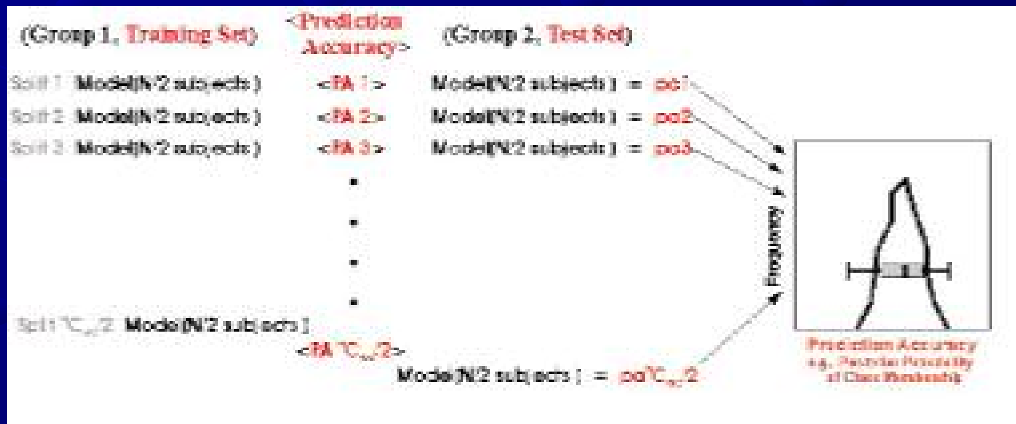
- **A Bias-Variance Tradeoff.**
  As model complexity increases (i.e., #PCs 10 →100), prediction of design matrix's class labels improves and reproducibility
  (i.e., activation SNR) decreases.

- **Optimizing Performance.**
  Like an ROC plot there is a single point, (1, 1), on this prediction vs. reproducibility plot with the best performance; at this location the model has perfectly predicted the design matrix while extracting an SNR.

LaConte S, et. al. Evaluating preprocessing choices in single-subject BOLD-fMRI studies using data-driven performance metrics. *Neuroimage* **18:**10-23, 2003

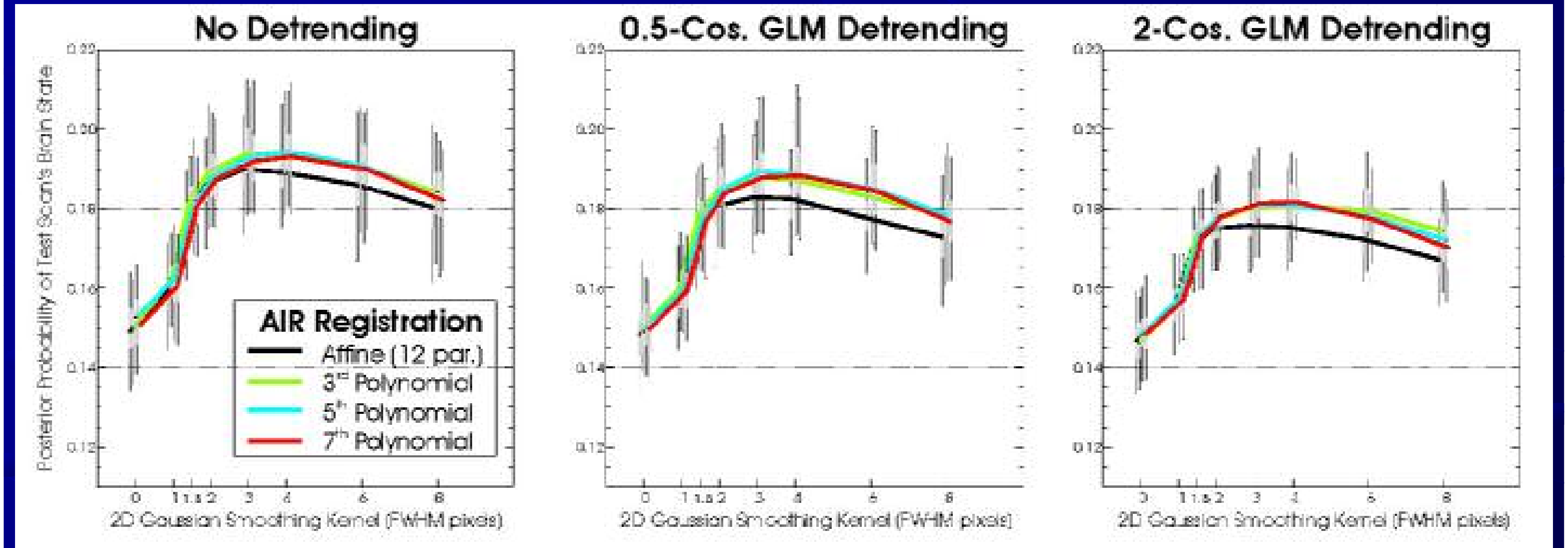
Static Force: Prediction Accuracy as a Function of Preprocessing





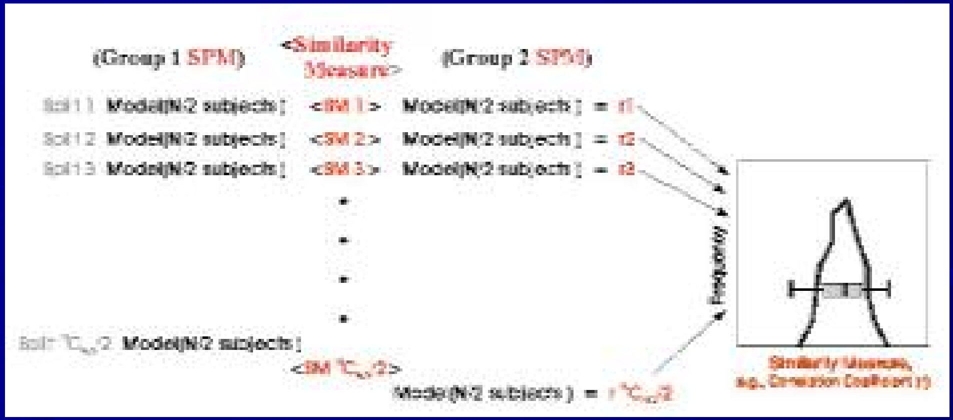
Strother SC, LaConte S, Anderson J, Muley S, Pulapura S, Ashe J, Yacoub E, Hu X, Rottenberg D. Detecting Large-Scale Brain Networks with BOLD fMRI: Visuo-motor and Sensory-motor Interactions in a Static Force Task. [abstract]. 8th Int. Conf. on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. CDROM-NeuroImage, Vol. 16, No. 2.

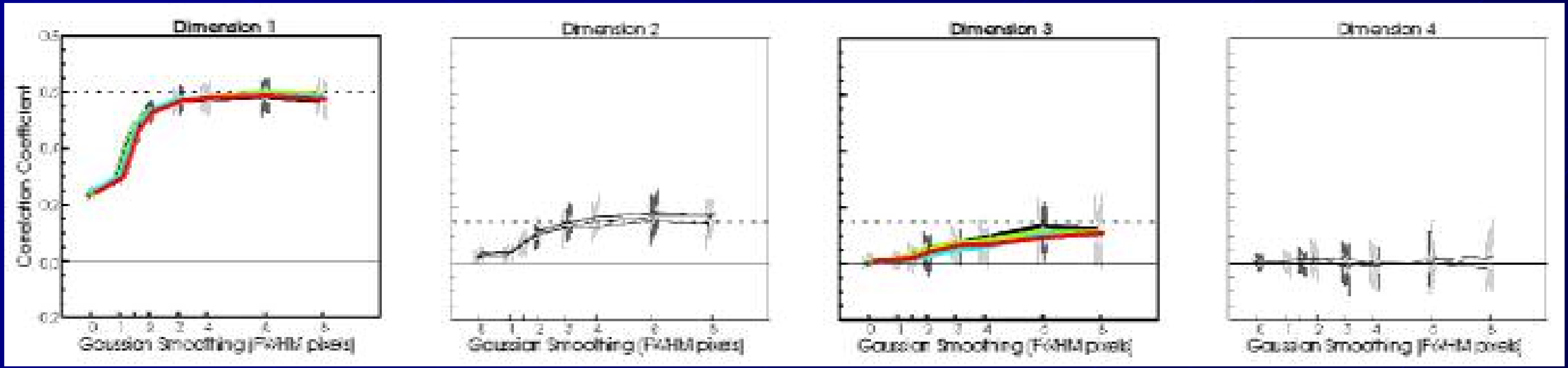
Static Force: Prediction Accuracy



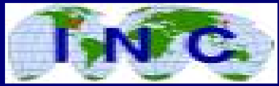
$$p\left(c^{(j)} \middle| \mathbf{x}_{te}^{(j)}; \theta_{tr}\right) = \frac{1}{K} \exp\left\{ -\frac{1}{2} \left\| \mathbf{L}_{tr}^{T} \left(\mathbf{U}_{tr}^{*}\right)^{T} \left(\mathbf{x}_{te}^{(j)} - \overline{\mathbf{x}}_{tr}^{(\mathbf{c}^{(\bullet)})}\right) \right\|^{2} \right\} p\left(c^{(j)}\right)$$

Static Force: Reproducibility & Dimensionality

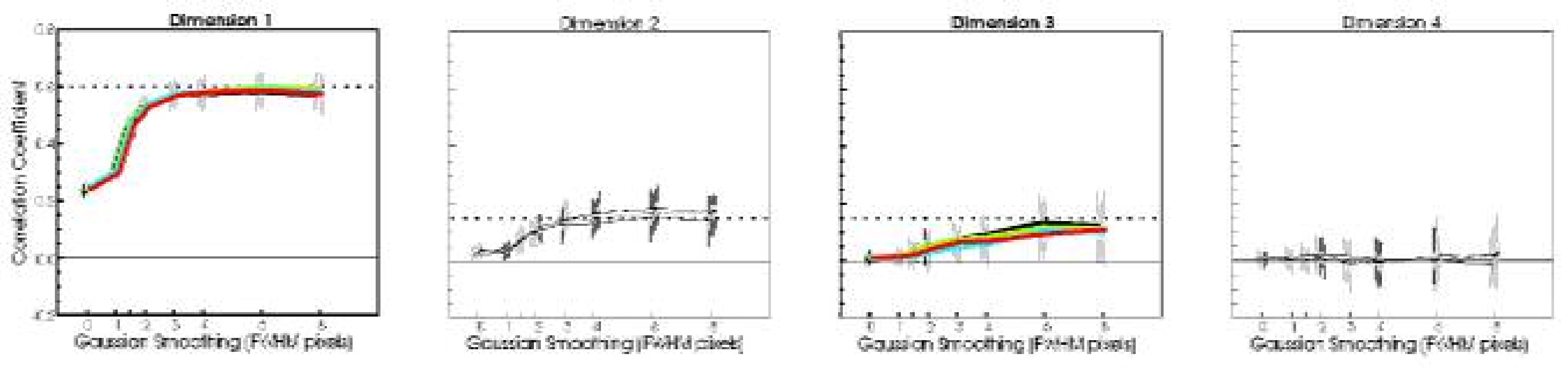


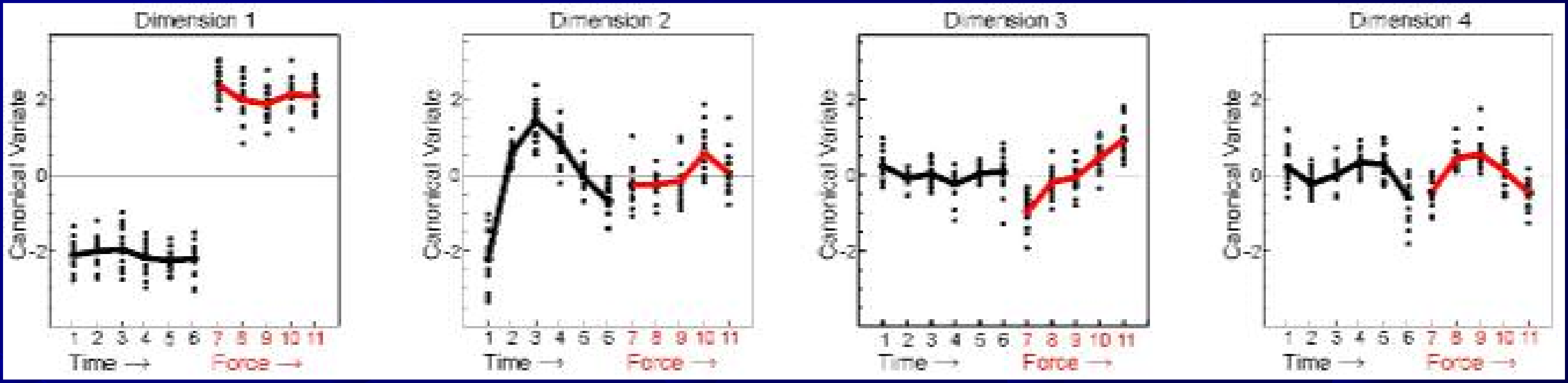


0.5 Cos Detrending

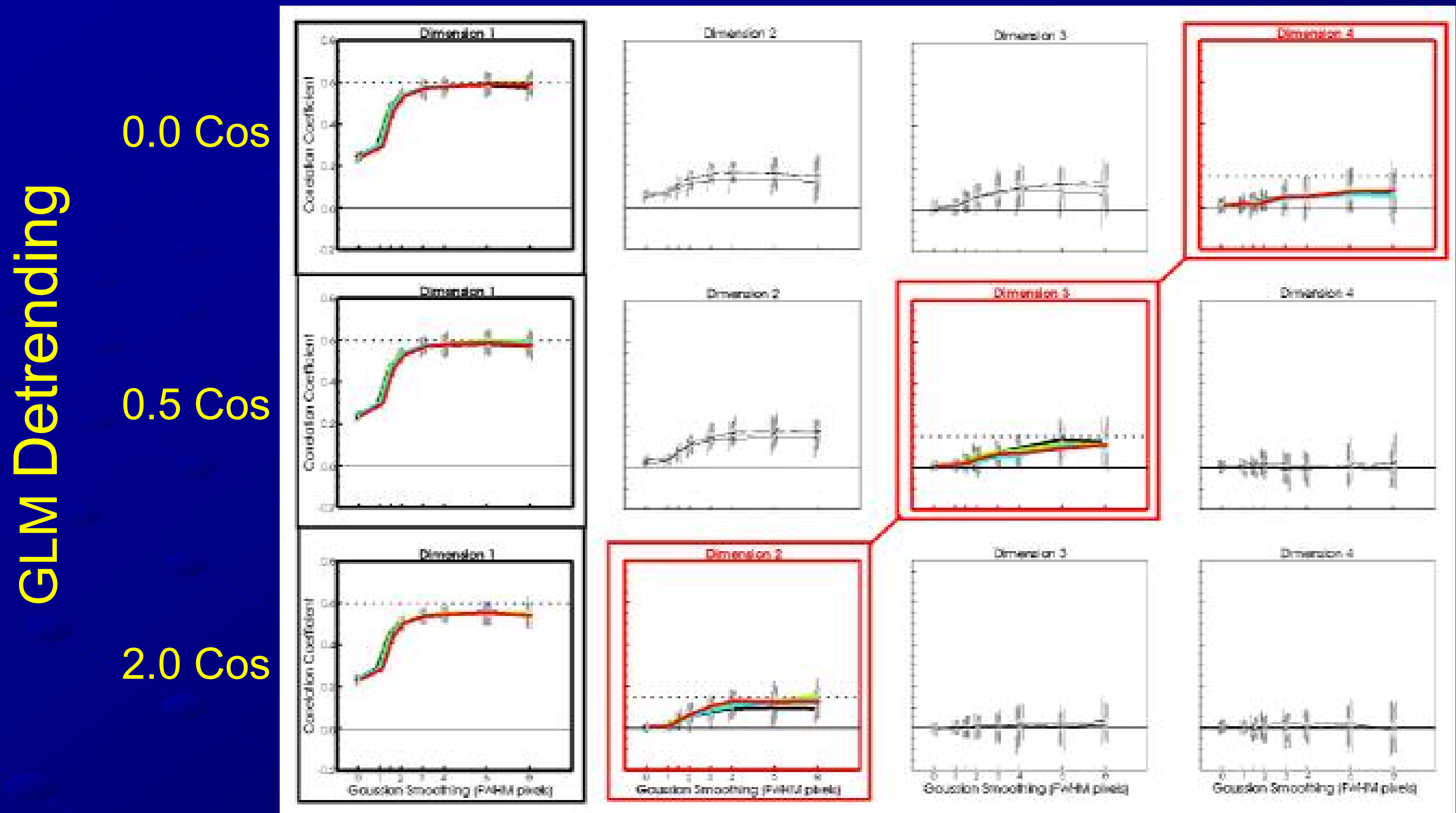


Static Force: Reproducibility & Time Course





Static Force: Reproducibility (SNR)



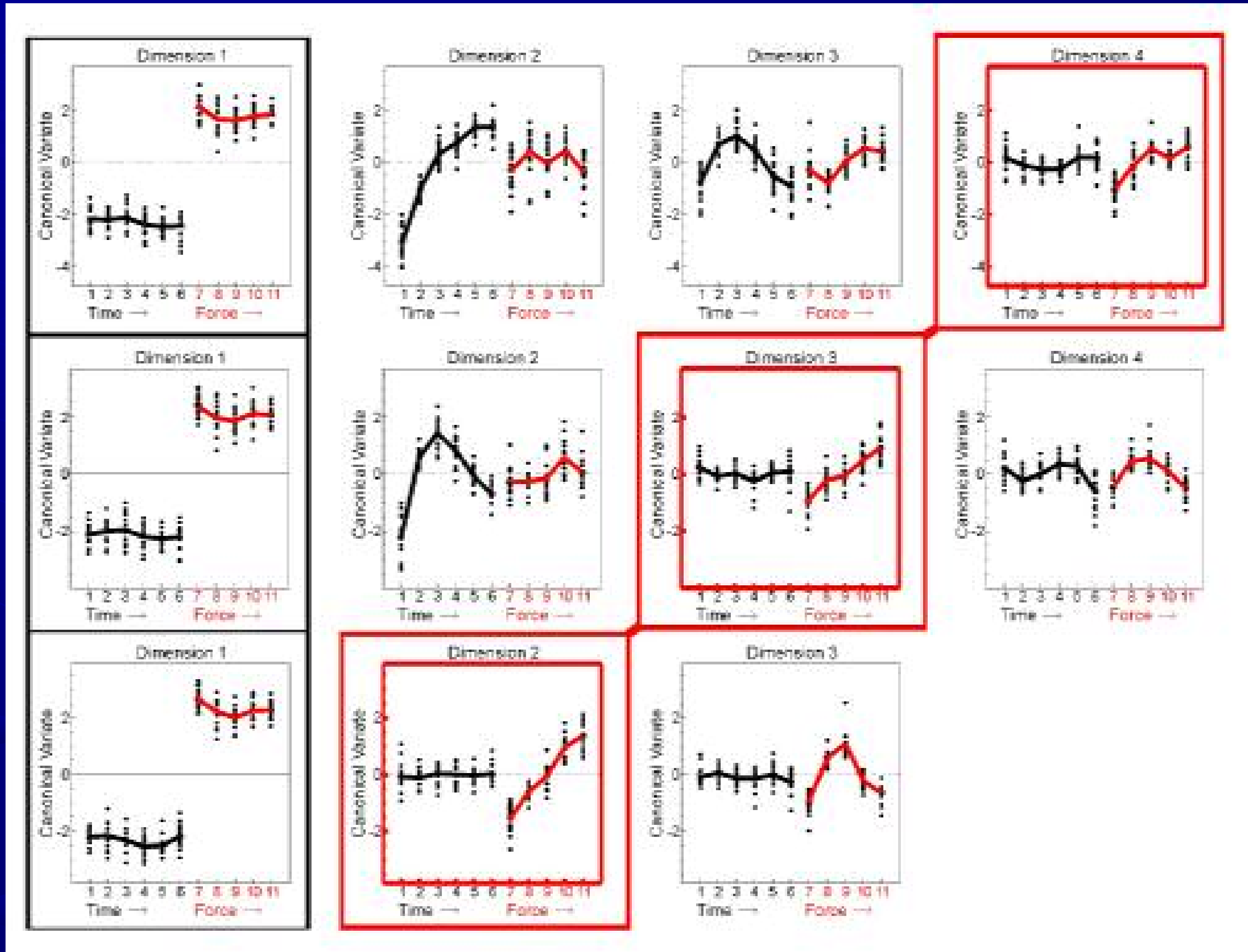
Static Force: 11-Class CVA

GLM Detrending

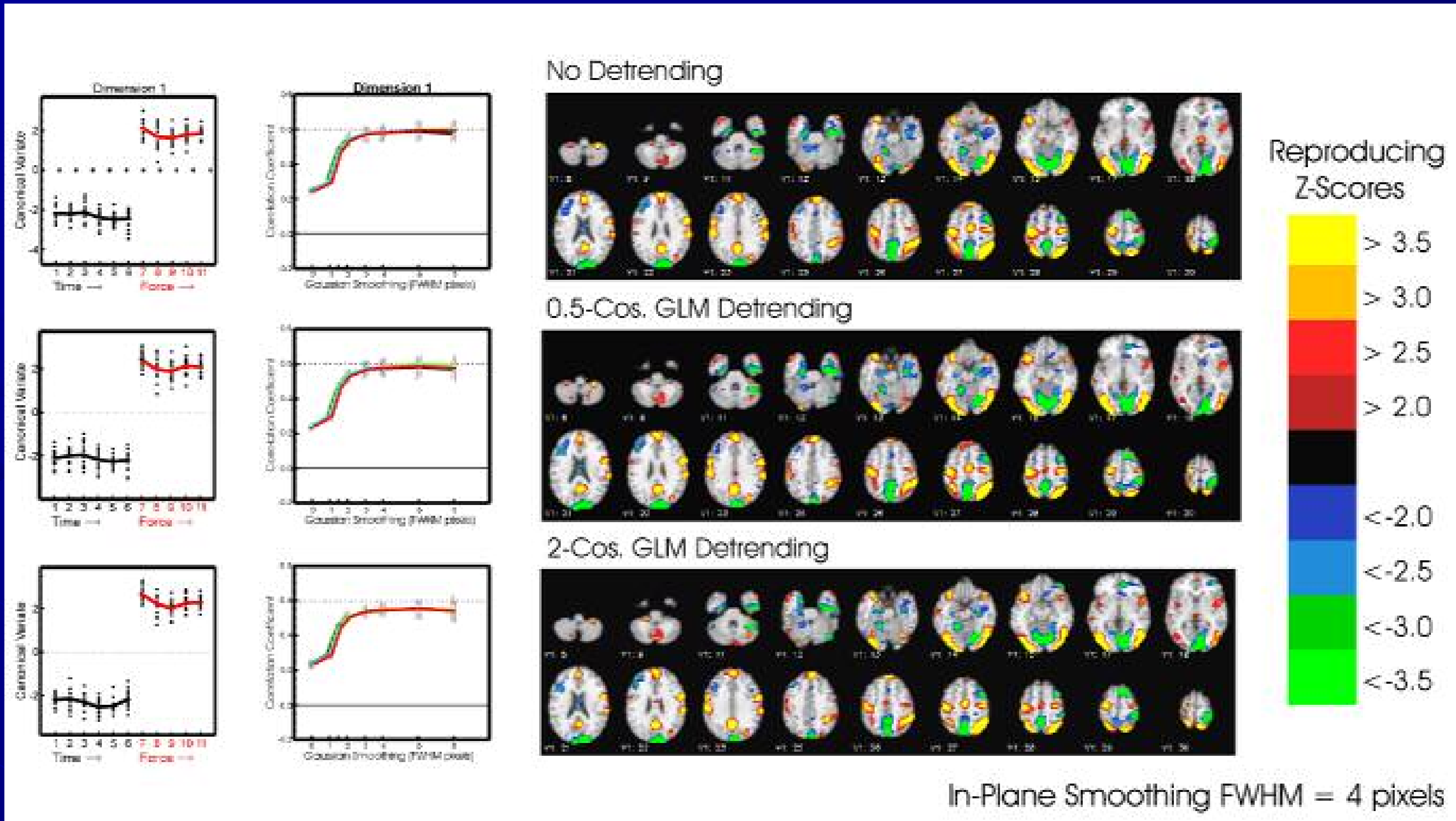
0.0 Cos

0.5 Cos

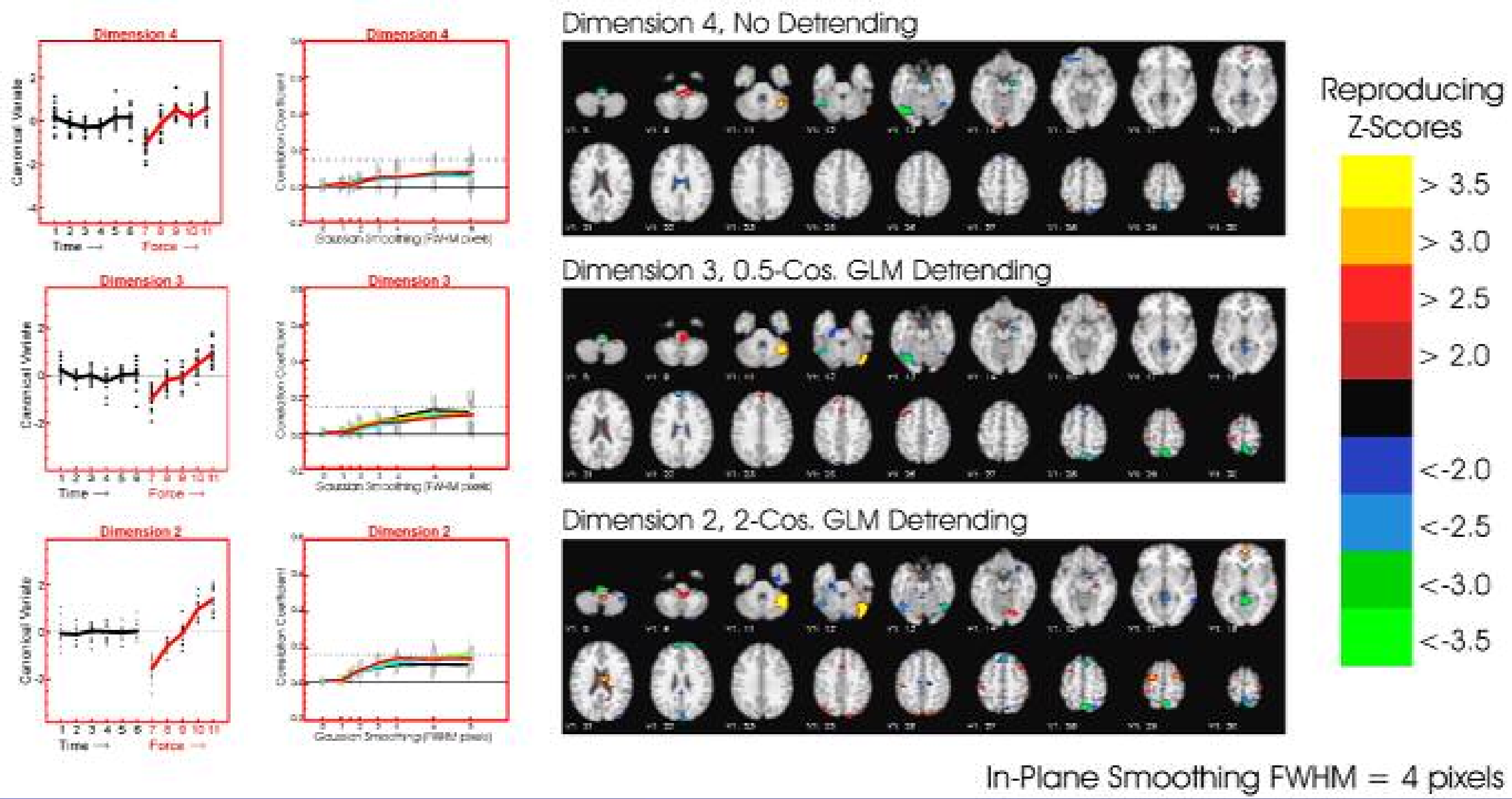
2.0 Cos



ON-OFF Static-Force Response

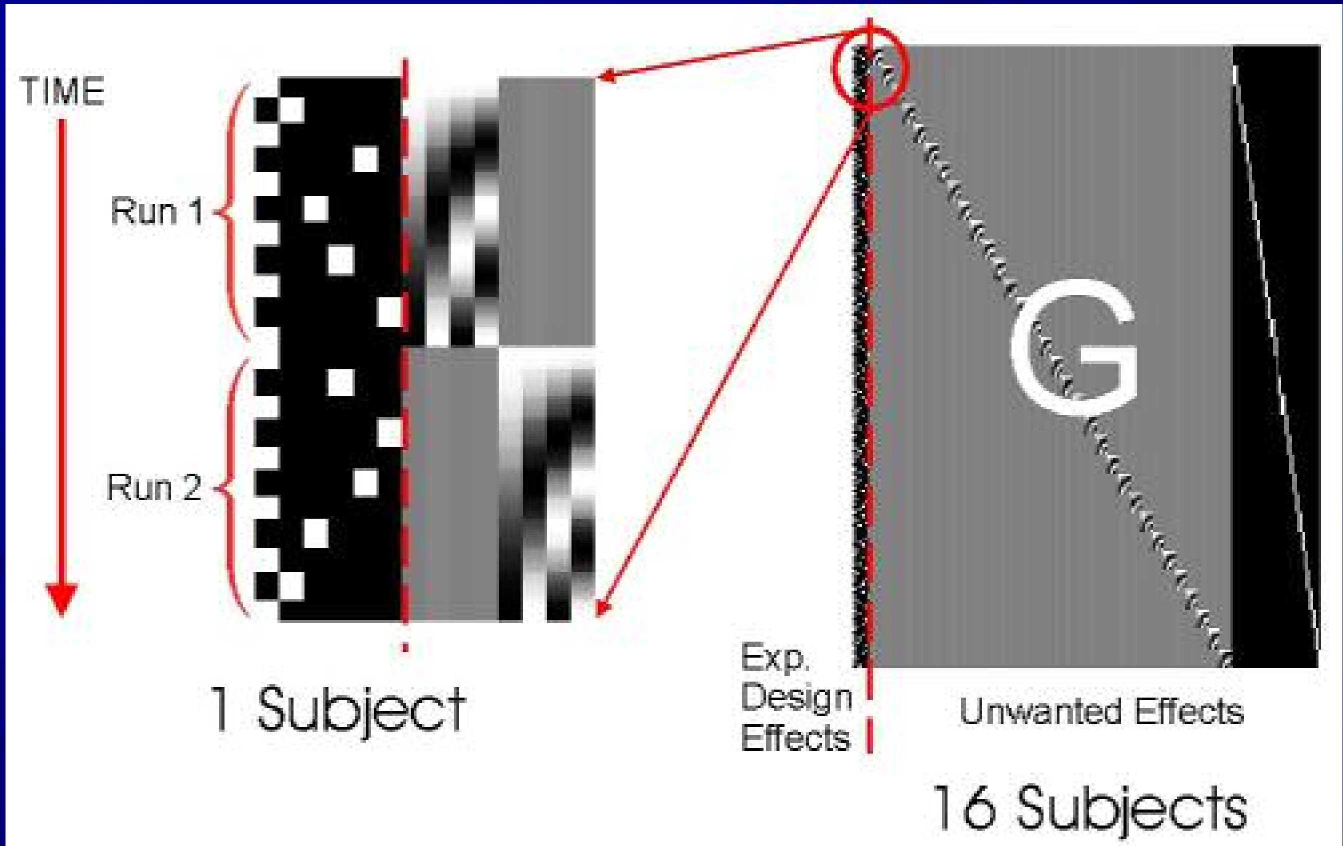


Parametric Static-Force Response



GLM Design Matrix

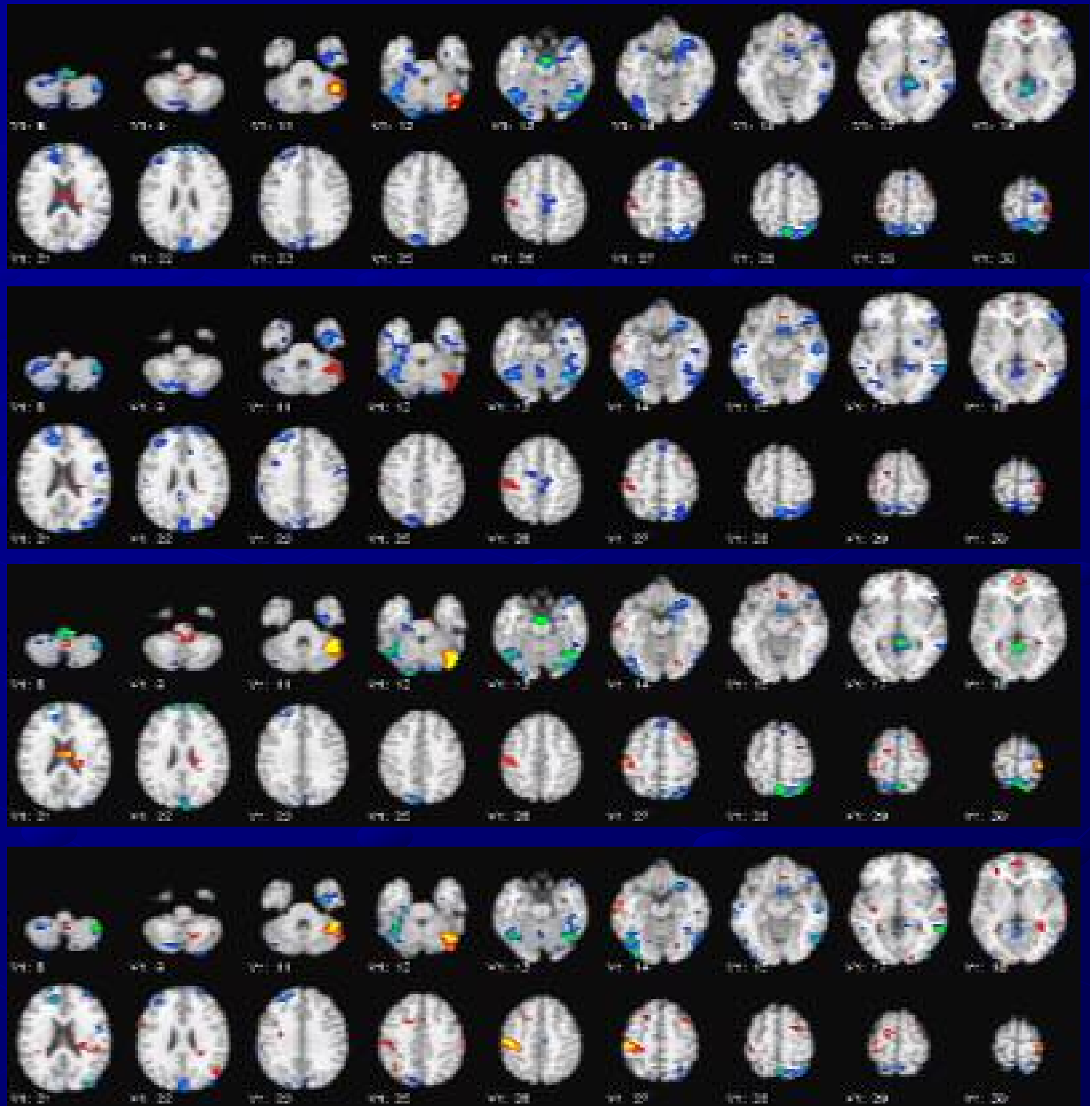
$$Y_{[\text{Subject(time)} \times \text{Voxels}]} = G_{[\text{Subject(time)} \times \text{Effects}]} \times B_{[\text{Effects} \times \text{Voxels}]} + \text{error}$$



GLM, Z-Scored, Reproducing Activation Patterns for Linear Static Force

Unnormalized

Normalized



β

t-stat.

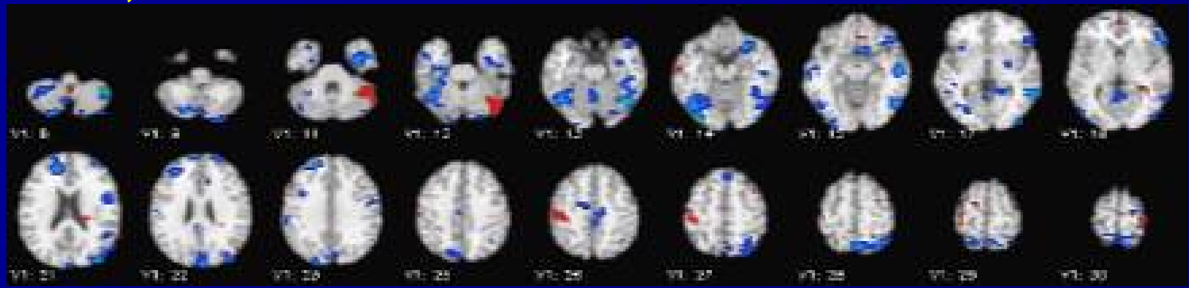
β

t-stat.

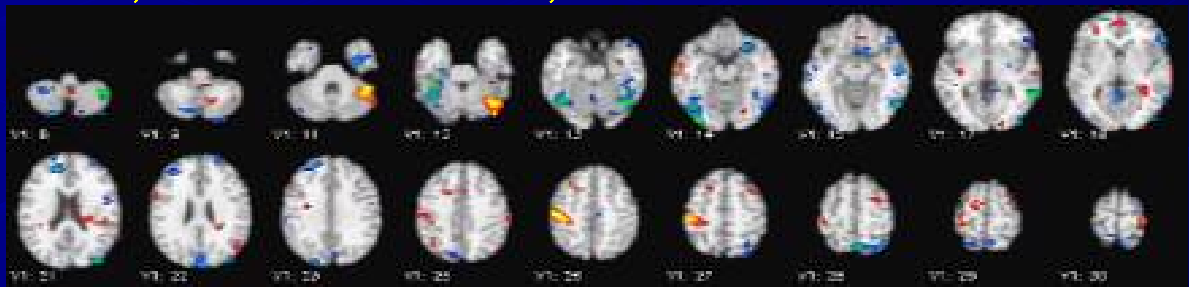
Activation Pattern Plurality: Linear Static Force

(AIR7, In-plane FWHM=4pixels, Detrend=4cos)

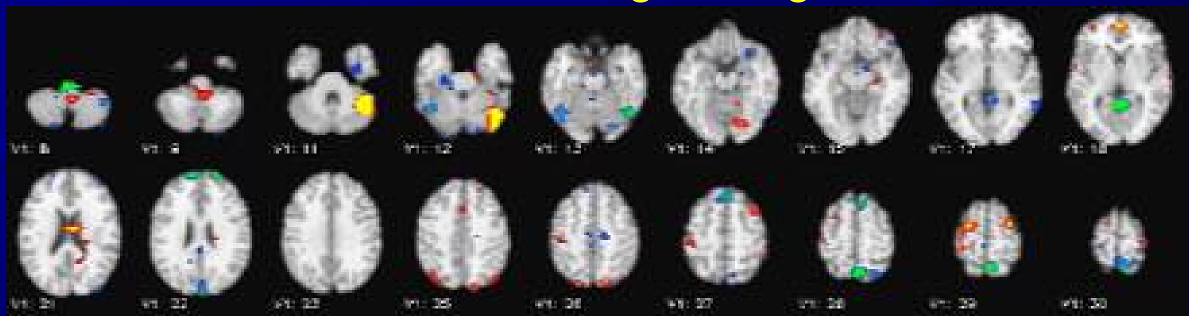
GLM, Z-Scored t-stat

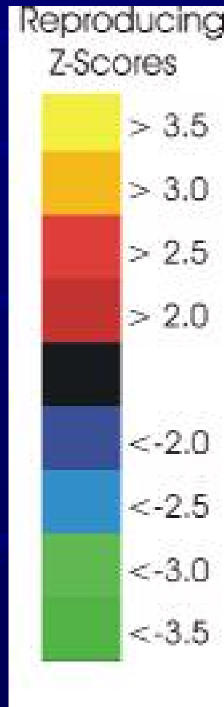


GLM, Normalized Scans, Z-Scored t-stat



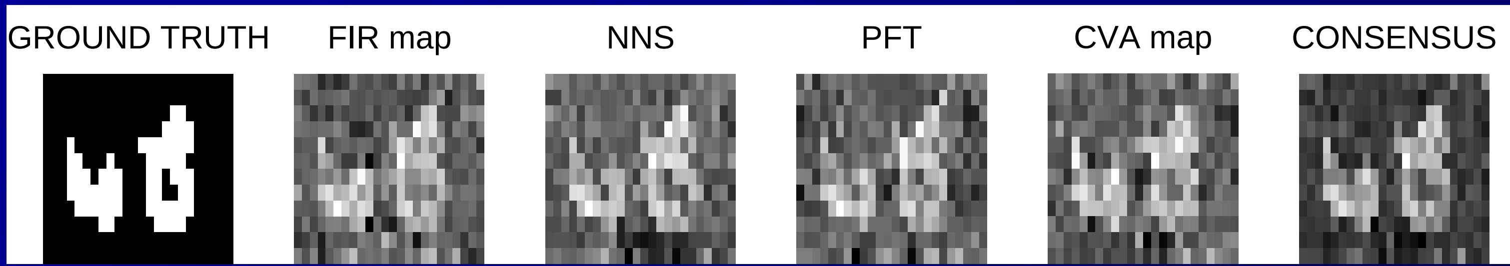
CVA, Z-Scored Canonical Eigenimage



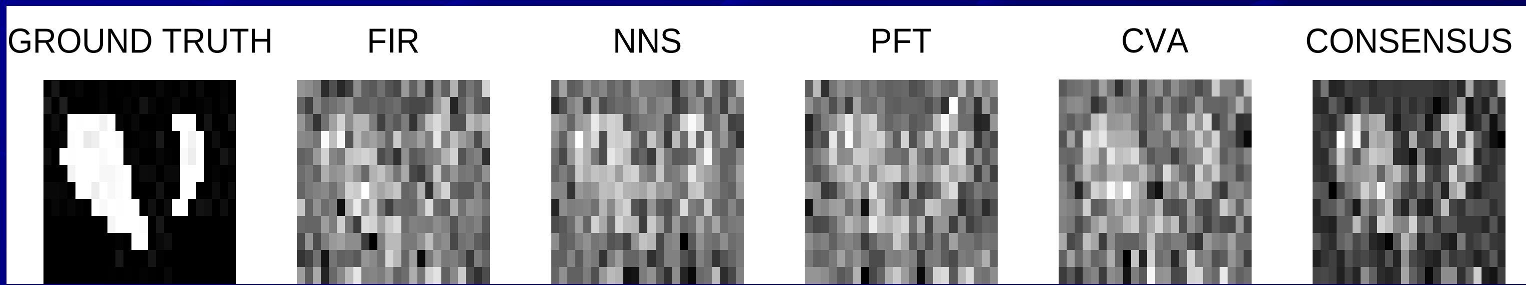


Consensus Through Plurality

Simple Signal



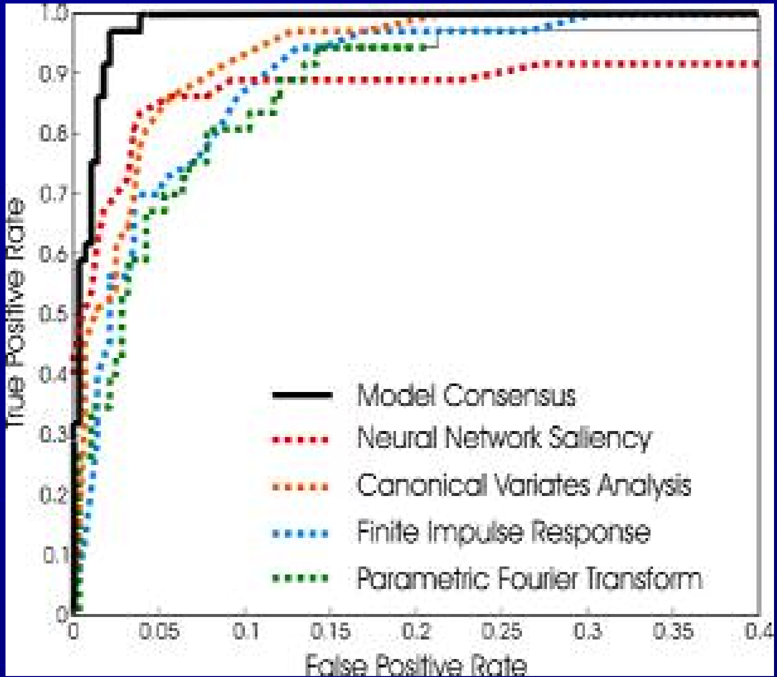
Complex Signal



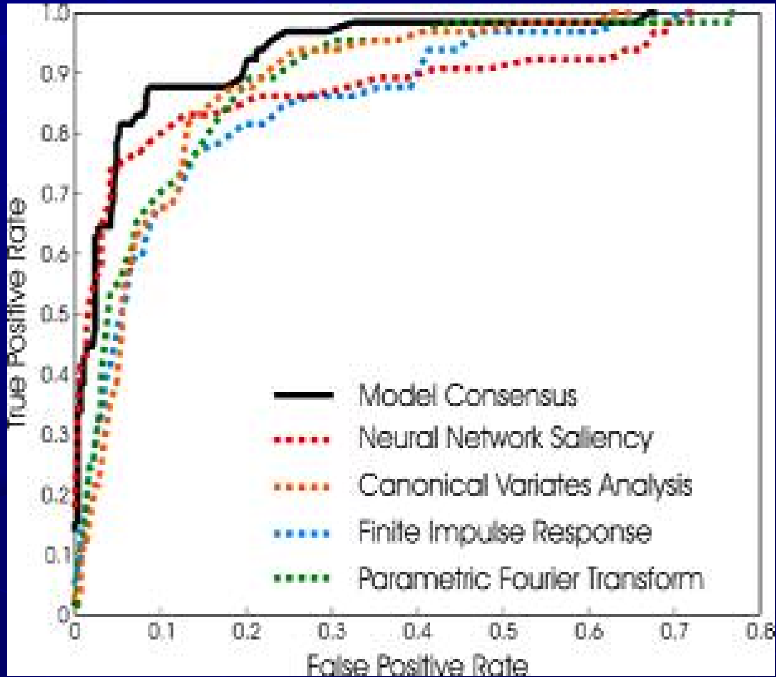
FROM: Hansen LK, Nielsen FA, Strother SC, Lange N. Consensus Inference in Neuroimaging. Neuroimage 13:1212-1218, 2001

Consensus ROC Results

Simple Signal



Complex Signal



Some Conclusions

- Living with plurality is tightly coupled to understanding the bias-variance tradeoffs within and across meta-models as a function of preprocessing and sample size.
- Standard GLM techniques may be outperformed as signal detectors by using biased estimators with a smaller parameter variance for a given sample size.
- Statistical learning theory provides quality metrics for testing data consistency within and across meta-models, independent of neuroscientific expectations, e.g., NPAIRS.
- In the absence of sufficient knowledge to choose an optimal meta-model (i.e., ignorance) a consensus average of multiple models may outperform any of the individual models as a signal detector.

Conclusions: NPAIRS Results

- Spatial smoothing (and voxel size), and temporal detrending are critical choices for optimizing the analysis of fMRI studies;
- The dimensionality of the response and the reduction of spurious temporal interactions is a function of spatial smoothing and temporal detrending;
- There are several distinct spatial-smoothing ranges:
  - 0 ~ 12 mm with rapidly increasing prediction and activation signal-to-noise (SNR) driven by individual subject responses;
  - ≥ 12 mm with decreasing prediction and flat to decreasing activation SNR;
- Results are somewhat dependent on the anatomical accuracy of between-subject registration depending on temporal detrending;
- Univariate (GLM) and multivariate (CVA) appear to sample quite different regions of fMRI model space and need to be reconciled;
- NPAIRS prediction and reproducibility metrics provide a systematic framework for studying and optimizing activation signal structure and associated bias-variance tradeoffs.