# Basic introduction to multivariate neuroimaging analysis – for nerds and novices

Christian Habeck

Cognitive Neuroscience Division, Department of Neurology,
Columbia University

**MCOS, February 23, 2024**

# Thanks to MCWG

- https://molecularconnectivity.com/

**Igor Yakushev**

- **Arianna Sala, Sylvia Caminiti, Simon Eickhoff, Vince Calhoun, Daniela Perani, Mattia Veronese, Alessandra Bertoldo, Vesna Sossi, Xin Di, Joana Pereira, Sharna Jamadar**

- **Michael Rullman, Daniel Talmasov, Mary Catanese, Chunmeng Tang, Matej Perovnik, Gabriel Gonzalez-Escamilla , Tommaso Volpi**

# Thanks to stateside colleagues

Buddhist mantra of cognitive neuroscience:

**“There is no I in fMRI (or PET)”**

**Co-PI:** Yaakov Stern

**Co-Investigators:**
Yian Gu
Seonjoo Lee
James Moeller

Non-Columbia:
Yunglin Gazes
Ray Razlighi
Teal Eich
Gene Alexander
David Raichlen

**Post Docs:**
Georgette Argiris
Sharon Sanz Simon
Angeliki Tsapanou
Hengda He
Eleanna Varangis

**Assistance:**
Reshma Babukutty
Caleb Haynes
Mihir Kurkure
Oriana Myers
Joey Hopkins
Charlie Leighton

Support: National Institute on Aging

# Opinionated intro

- Not too long ago, a strange “culture war” was going on with multivariate analysis pitted against univariate analysis, at least in fMRI analytics
  - Univariate proponents would claim that multivariate analysis had no false-positive controls
  - Replication and permutation tests were not highly prized
  - Multivariate results were seen as hard to interpret

- A Yale Biomed-Engineering professor expressed incredulity to me on hearing a talk of mine in 2008: “Multivariate vs. univariate dichotomy is like a war of apples against oranges” – his apt commentary

- At the same, multivariate analysis had been performed in resting FDG-PET already since 1984 (Horwitz, Rappoport, Moeller, Strother, McIntosh, Eidelberg and many more)

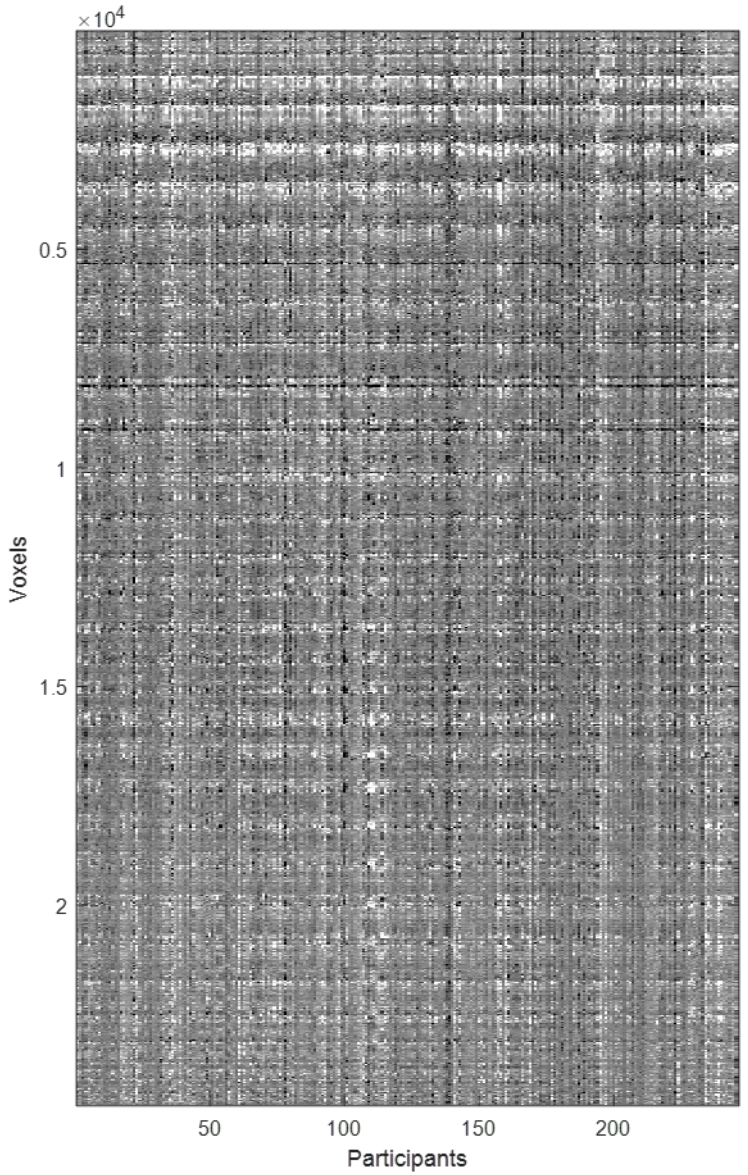
- **Opinionated take: Today, multivariate analysis and highly derivative frameworks have become very common, sometimes apparently more for intellectual ownership than scientific illumination**

➔ **Molecular connectivity strives for informed use of connectivity approaches, following the tradition of the successful work in PET, while avoiding anti-Occam’s razor tendencies**

# Cross-sectional sample data Y

[Y]= number of voxels x number of participants

V x N



# Typical univariate analysis

[Y]= number of voxels X number of participants

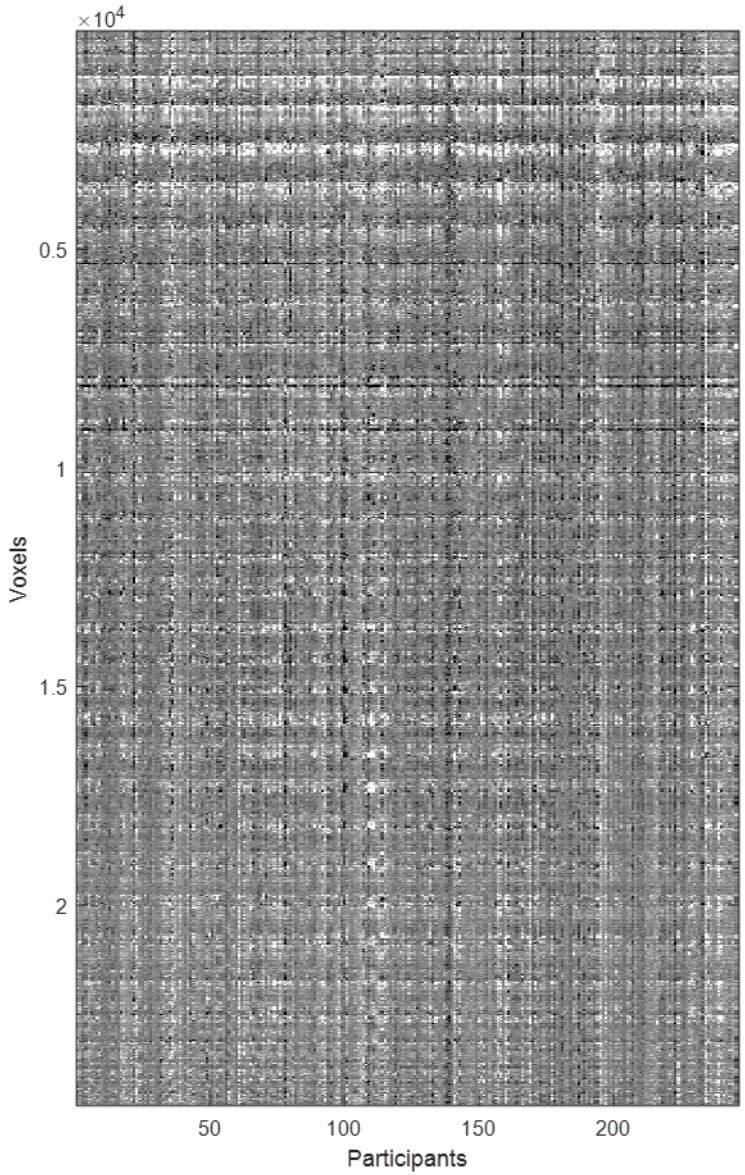
Perform linear models to assess the influence of covariates in voxel activation

- Linear regression is run within voxel (=row wise), across participants

$$\mathbf{Y}(i) = \mathbf{X}\,\boldsymbol{\beta}(i) + \boldsymbol{\varepsilon}$$

i=1…V voxel location

**X** = ‘Design matrix’
**β =** parametric maps, dependent on voxel location



# Simple dimensional thoughts

[Y]= number of voxels X number of participants

[Y]= number of voxels x number of participants

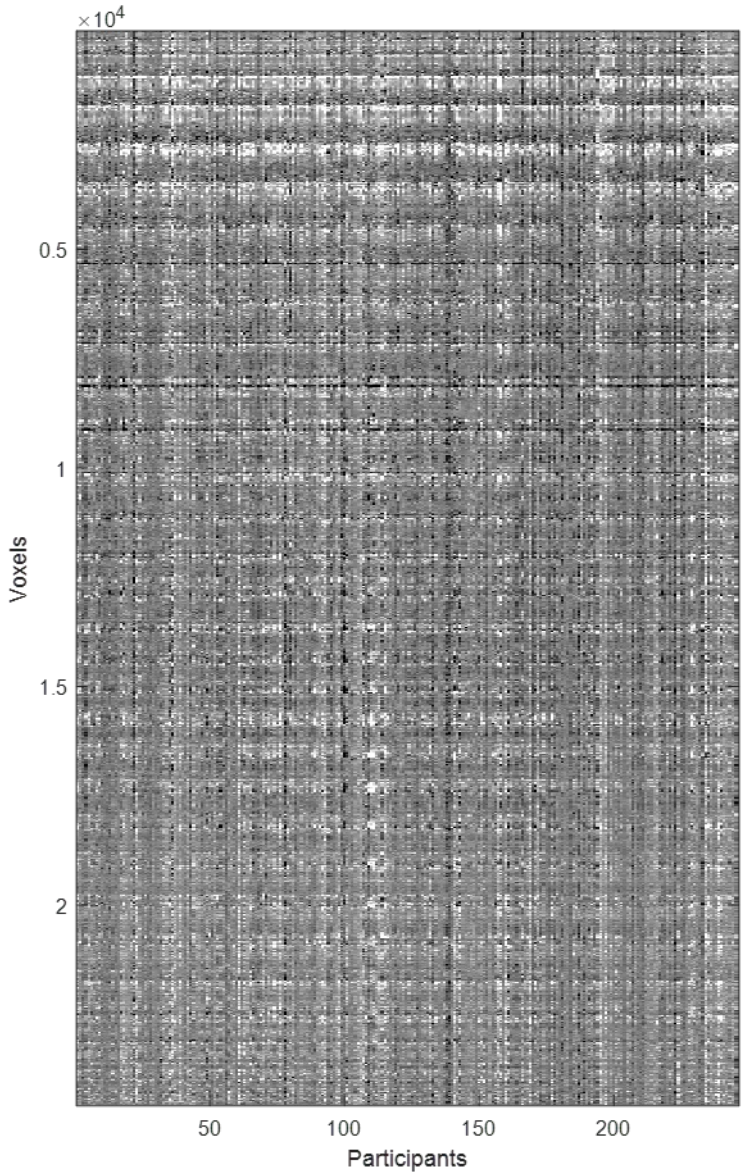
V x N

Neuroimaging, like genetics, and other fields with increasing data richness often has (several orders) more variables than observations

**V > N**

➔ **Rank (Y) = N** this implies that the maximum number of independent sources in the data is N

➔ There ***must*** be correlation between the voxels ***by necessity,*** prediction of an outcome with $F : R^V$ ➔ outcome has to consider this redundancy



# Multivariate analysis

[Y]= number of voxels X number of participants

Multivariate Decompositions - generic form from matrix form of OLS:

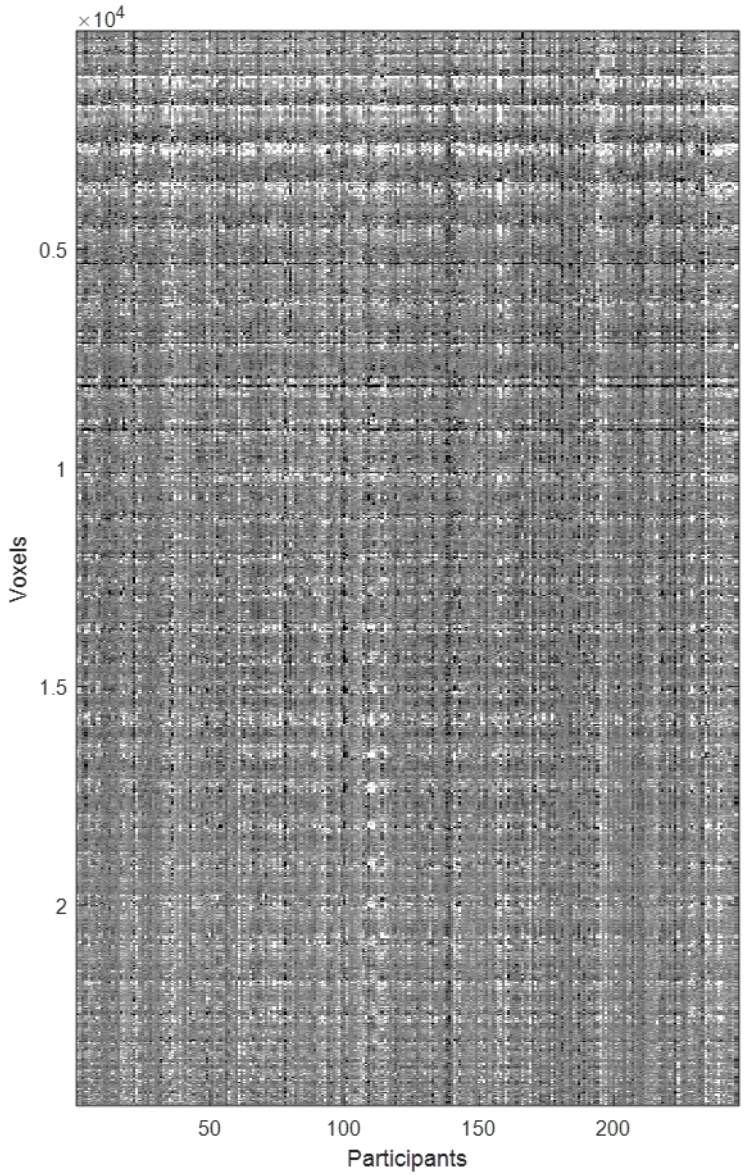
$$\mathbf{Y=V\,W'} + \varepsilon$$

**V** = 'component matrix', columns are components

[**V**] = number of voxels **x** number of components

**W** ='score matrix' or 'mixing matrix', columns are subject expression vectors

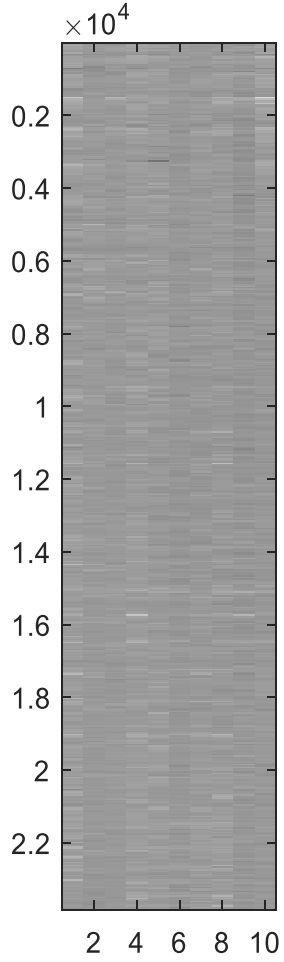
[**W**]= number of subjects **x** number of components



# Generic form of multivariate analysis

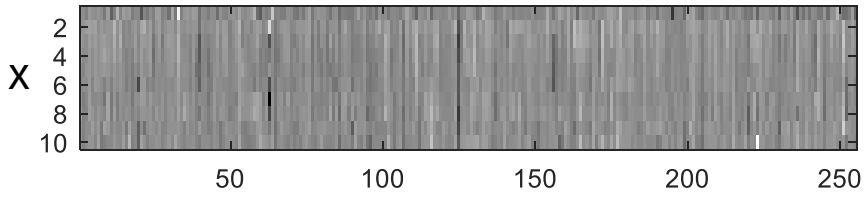
$$Y = V W' + \varepsilon$$

[V]=voxels x 10




Assume 10 components, C=10

W'= 10 x N




**Possible choices**

- **Principal Components Analysis (PCA): Columns of V and W are mutually orthogonal**
- (Spatial) Independent Components Analysis (ICA): Columns of **V** are mutually statistically independent to all orders
- Non-Negative Matrix factorization (NNMF): **V,W** >0
- Many more: sparsity of loadings, VariMax, ProMax, etc.
- ***Make your own?***

Don't confuse mathematical convenience of decomposition principles (=sparsity, orthogonality, independence) with scientific meaning! ➔ Need outside validation

# PCA didactics - Principal Component Analysis (Karl Pearson 1901)

## "Eigen equation" – what is that?

$$\mathbf{COV}\ v(i) = \mathbf{Y}\,\mathbf{Y'}\ \mathbf{v}(i) = \boldsymbol{\lambda}(i)\ \mathbf{v}(i)$$

**Eigen vectors just get scaled when multiplied by covariance matrix**

$$\mathbf{YY'\ V} = \mathbf{V}\,\boldsymbol{\Lambda}$$

Eigen equation in voxel x voxel space

$$\mathbf{Y'\ YY'\ V} = \mathbf{Y'\ V}\,\boldsymbol{\Lambda} \Longrightarrow \mathbf{Y'Y\ W} = \mathbf{W}\,\boldsymbol{\Lambda}$$

$\mathbf{W} = \mathbf{Y'\ V}$

Eigen equation in subject x subject space

**W=** subject eigen vectors

**V =** voxel eigen vectors = brain images

**Λ =** diagonal Eigen value matrix

$$\mathbf{Y} = \mathbf{V}\ \mathrm{sqrt}(\boldsymbol{\Lambda})\ \mathbf{W'}\ (=\mathrm{SVD})$$

"singular value decomposition"

# PCA didactics

$$Y = V(:,1)\,\lambda(1)\,W(:,1)' + V(:,2)\,\lambda(2)\,W(:,2)' + V(:,3)\,\lambda(3)\,W(:,3)' + \ldots$$

$$\mathbf{Y = V}\ \mathrm{sqrt}(\mathbf{\Lambda})\ \mathbf{W'}\ (=\mathrm{SVD}) \qquad \text{with}\ \ \mathbf{W \rightarrow W}\ \mathrm{sqrt}(\mathbf{\Lambda}): \qquad \mathbf{Y = V\,W'}$$

"singular value decomposition"

☝ Decomposition into orthogonal components in descending variance order without inherent stochastic variability

Objective: truncate at the right place to capture signal, and split off noise

$$\mathbf{Y} = \mathbf{V}(:,1{:}C)\ \mathbf{W}(:,1{:}C)' + \boldsymbol{\varepsilon} \qquad \text{for C components}$$

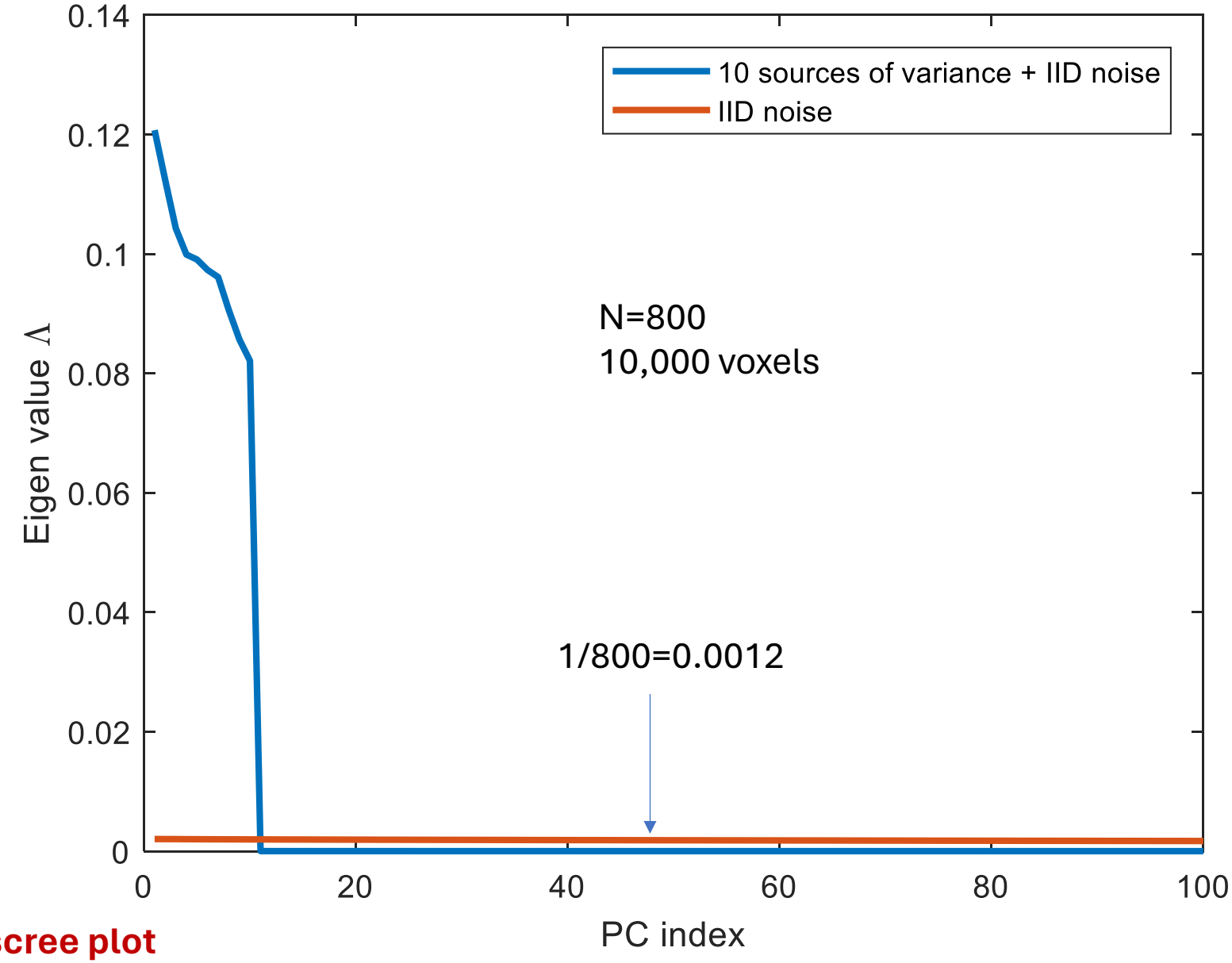
# Scree plot for PCA: compute Eigen values λ (=VAF) from whole PC-spectrum

PCA **(Y)** ➔ **(W,V,Λ)**

➔**Normalize that λs sum to 1**

➔ The flatter the scree plot, the more independent sources in the data

➔ Pure voxel noise gives approximate 1/N line



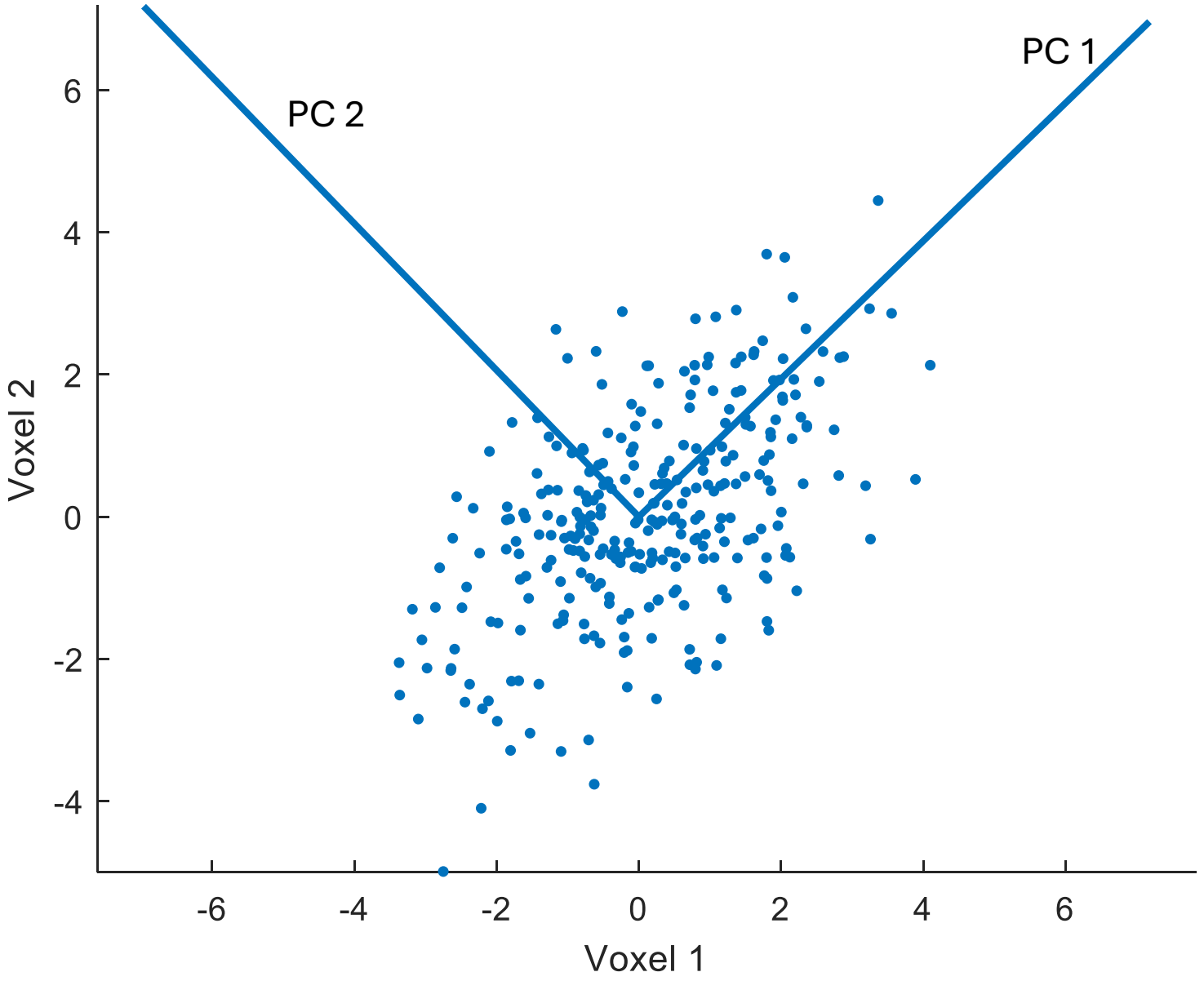

**Simplest visual heuristic: "kink" in scree plot**
**Others: AIC, BIC, MDL, CP-Mallows, N-fold-CV**

# PCA didactics II

Toy data sets with only 2 voxels

V =

| | |
|---|---|
| 0.7178 | 0.6962 |
| 0.6962 | -0.7178 |



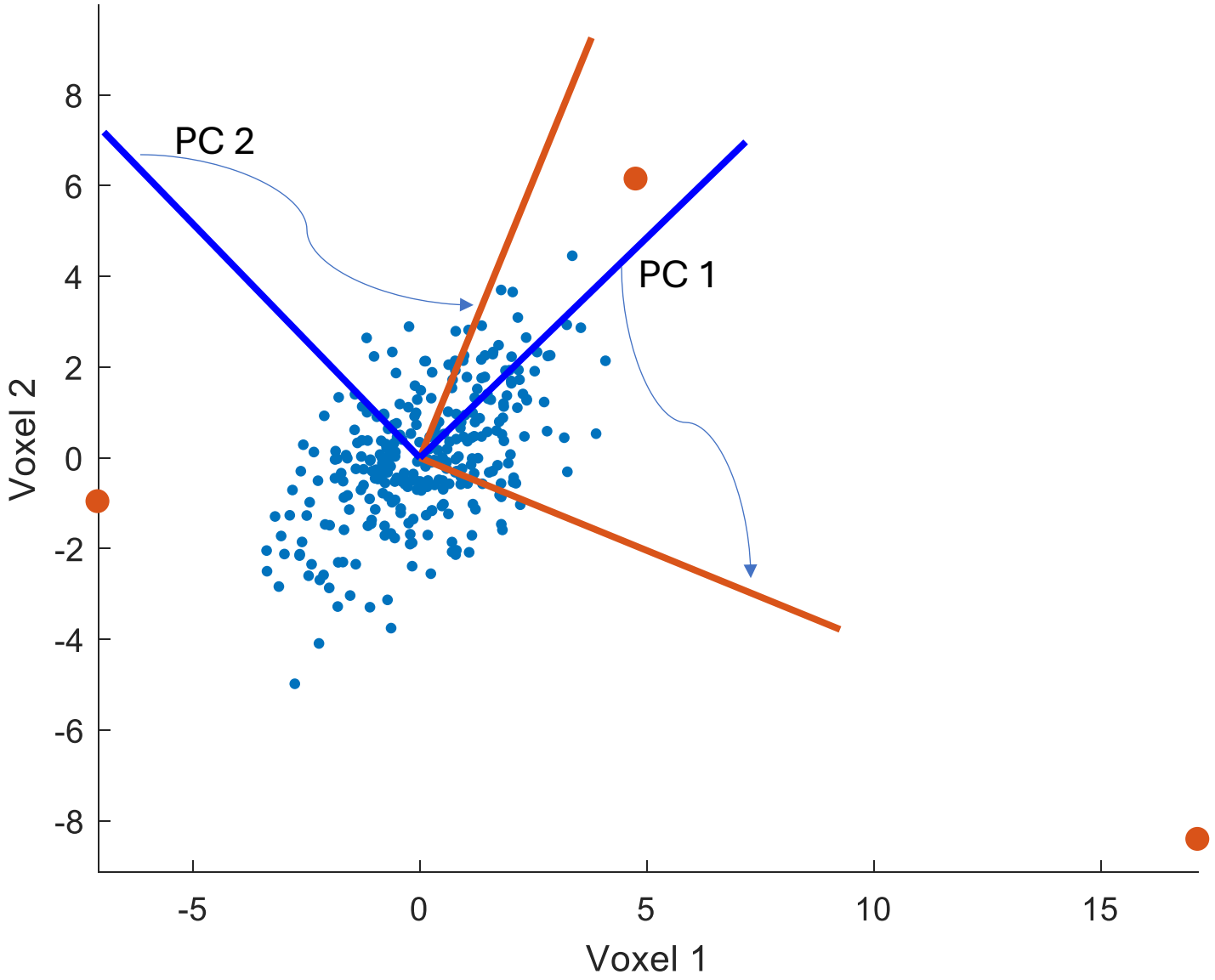
# PCA didactics III

**What happens with prior data reduction?**
**MLM, CCA, PLS, etc. …**

Data get reduced with pre-transformation

V –FULL =
0.7178 0.6962
0.6962 -0.7178

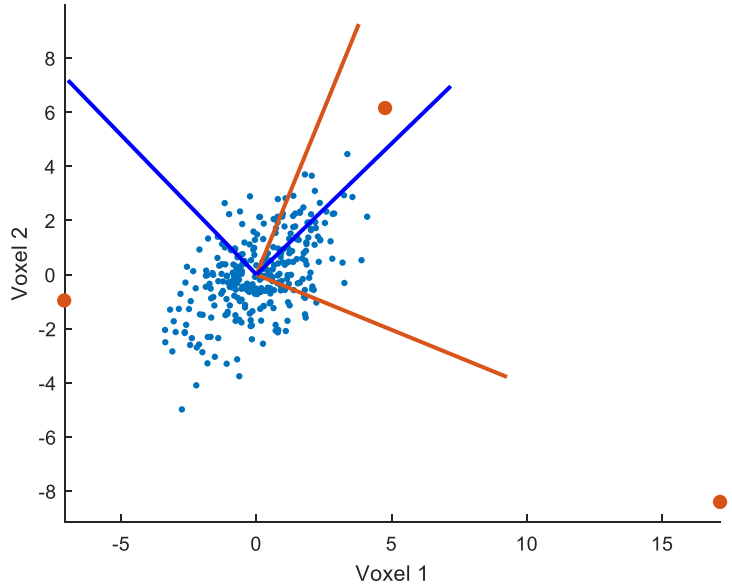
V – REDUCED =
0.9257 0.3782
-0.3782 0.9257



# PCA didactics III

**What happens with prior data reduction?**
**MLM, CCA, PLS, etc. …**

Data get reduced with pre-transformation




V –FULL =

| | |
|---|---|
| 0.7178 | 0.6962 |
| 0.6962 | -0.7178 |

V – REDUCED =

| | |
|---|---|
| 0.9257 | 0.3782 |
| -0.3782 | 0.9257 |

**Christian's opinion:**

**Don't reduce data prior to PCA with CCA or PLS, the results appear less arbitrary and more rigorous, and seemingly forgo the need for subset selection - but you are potentially throwing out baby with the bath water and might lose important signal!**

**(Cf. Simpson's paradox)**

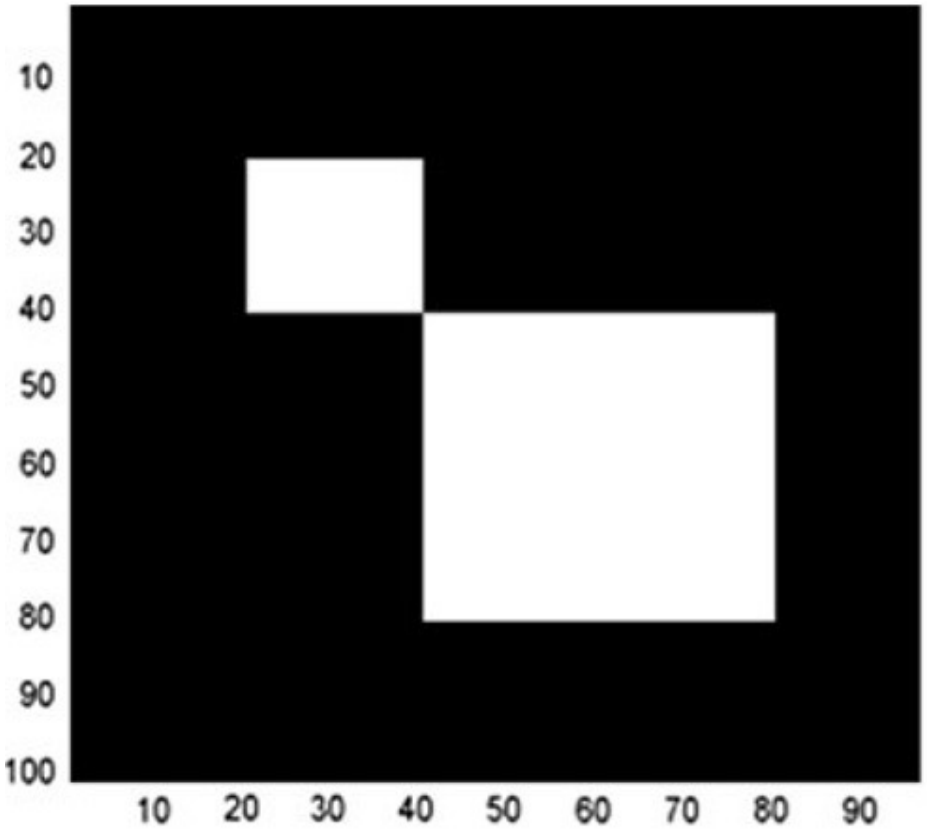
# Brief excursion to didactic article with 2-d example

Cell Biochem Biophys. 2010 Nov;58(2):53-67. doi: 10.1007/s12013-010-9093-0.

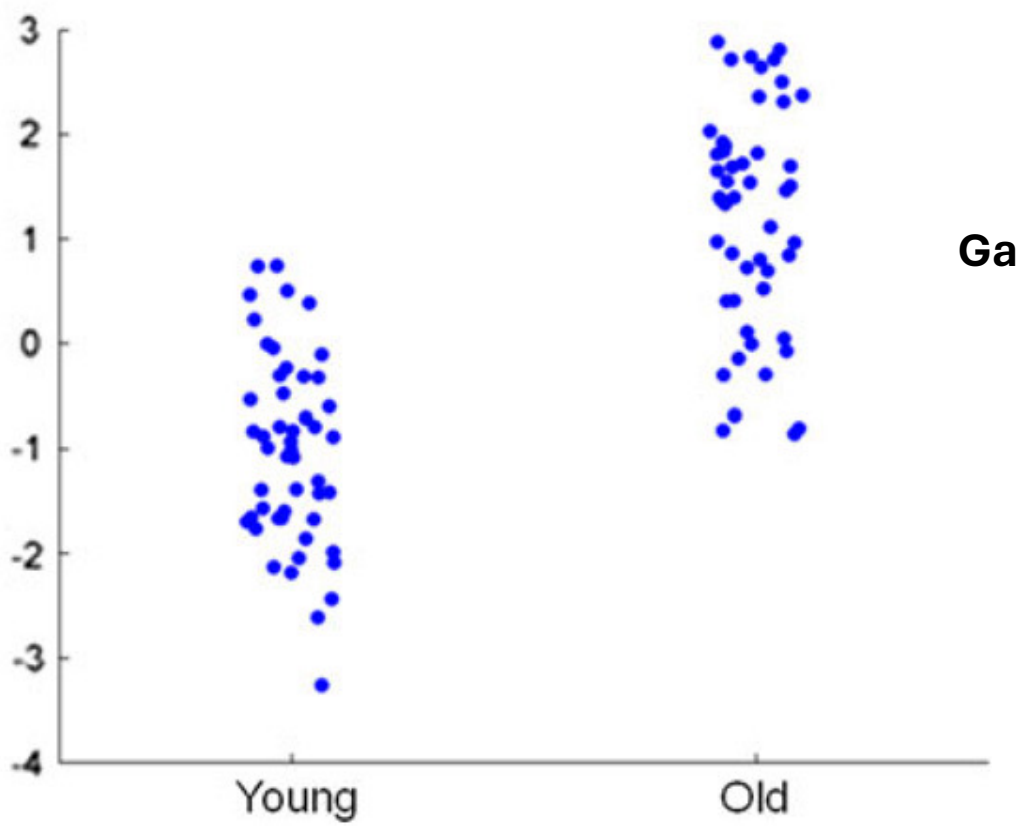
**Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's disease.**

Habeck C[1], Stern Y; Alzheimer's Disease Neuroimaging Initiative.

$$Y = v\,w' + \varepsilon$$



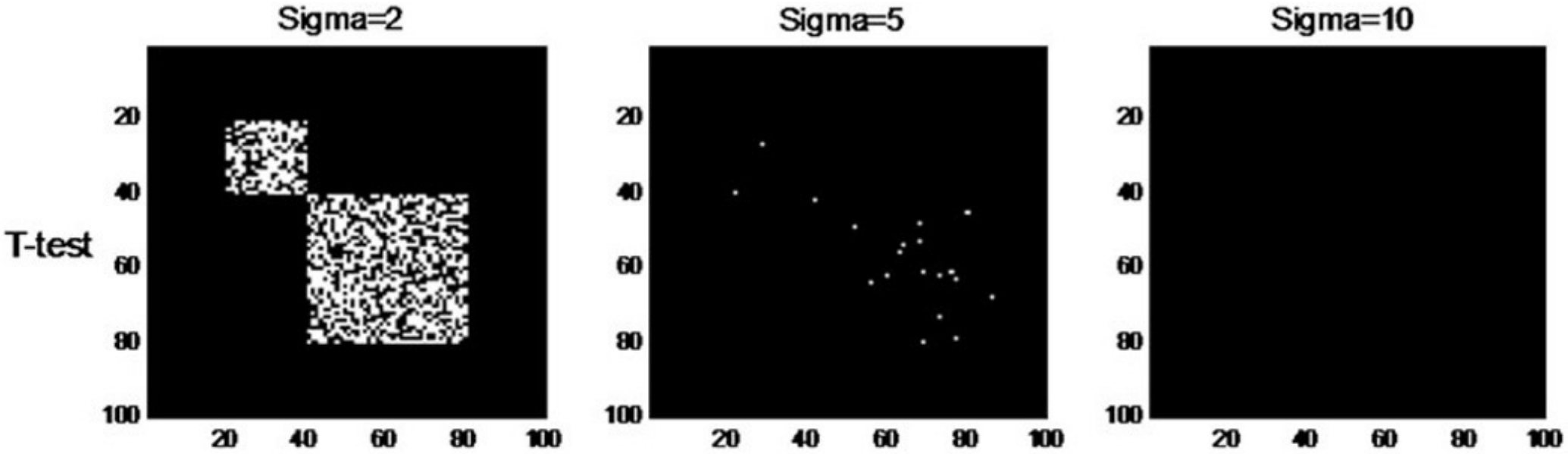

v = pattern

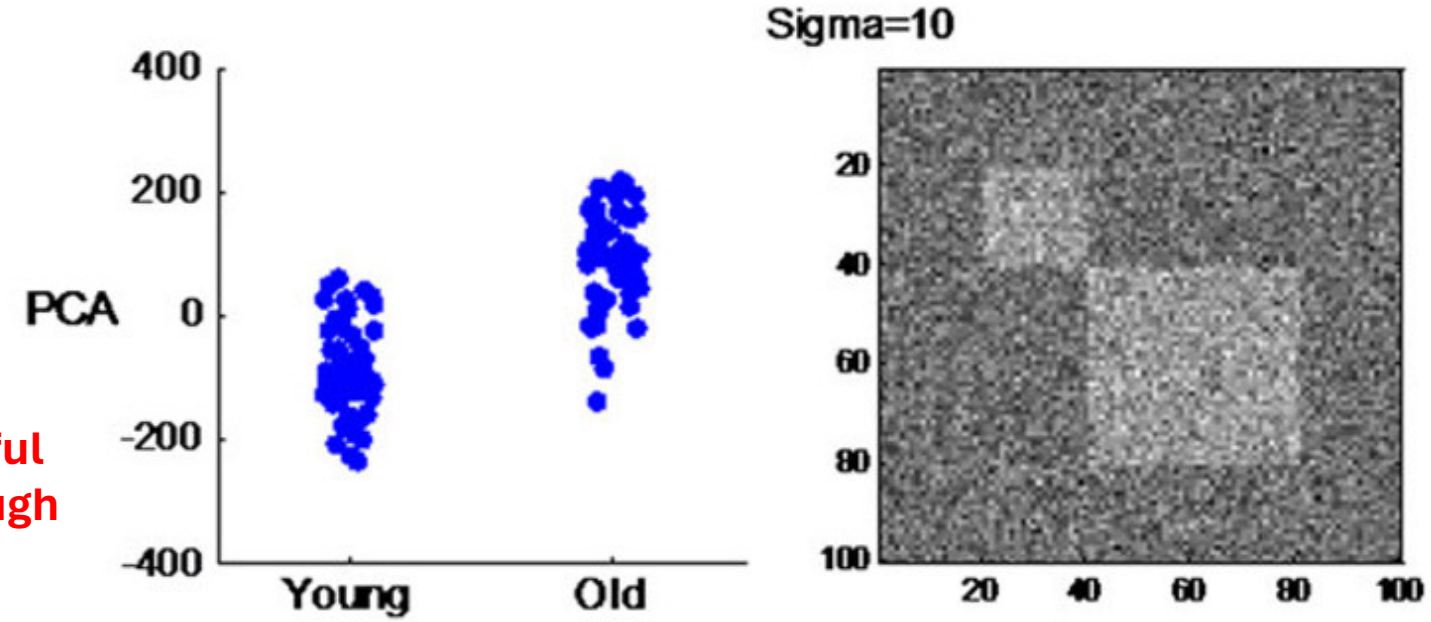



w = subject score vector

**Gaussian pixel noise:**
$\varepsilon \sim N(0,\sigma^2)$

# Different noise levels




➔ Multivariate analysis is more powerful
➔ Distributed nature of activation though is key

# Out-of-sample validation via prospective application

☝ Nice feature of multivariate patterns: **the pattern can be prospectively applied to any data set to yield a pattern score**, and verify correlations between the pattern score and a cognitive or clinical endpoint **even if the pattern was generated from a different data set**

$$score = DATA' * pattern$$

[score] = N x 1
[DATA] = voxels x N
[pattern] = voxels x 1

➔Both pattern and DATA need to be resliced into the same voxel space, obviously

# Recent “caveats” about PCA

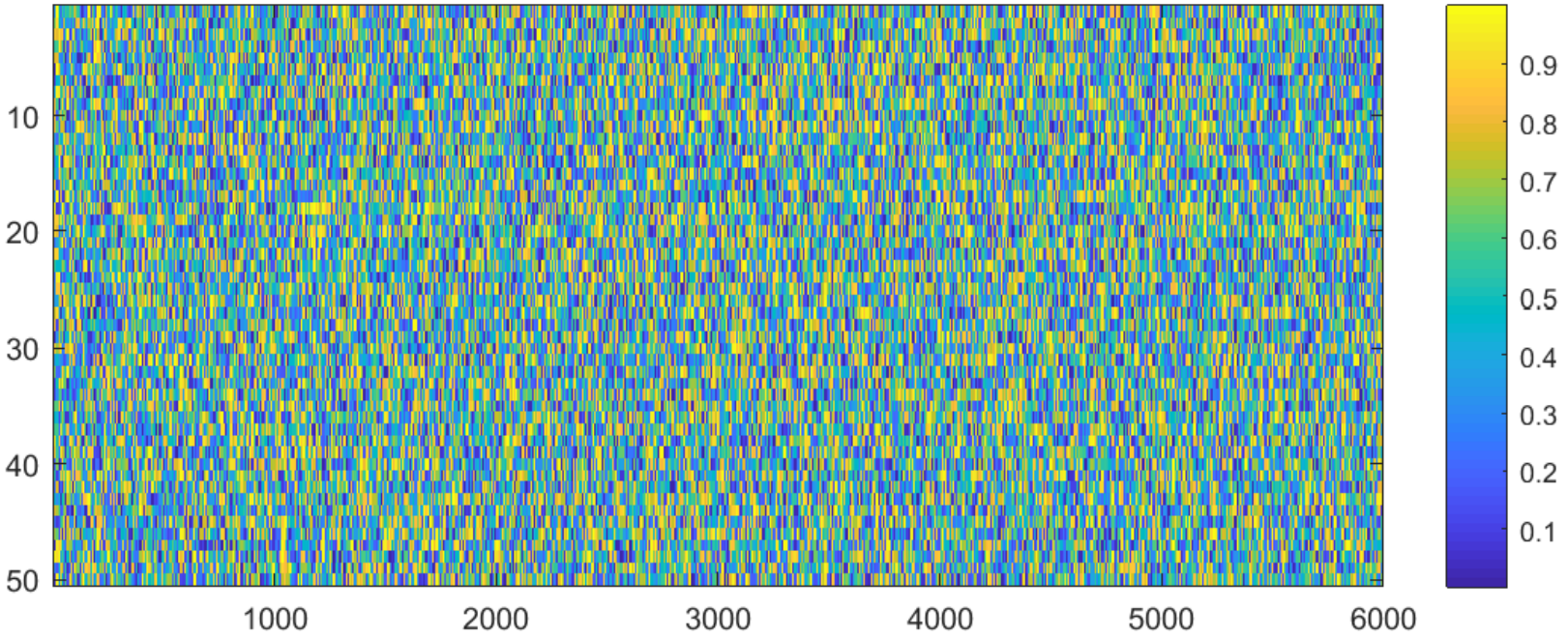
Some caveats about PCAs have been highlighted recently and mainly address well-documented overreach in the interpretation

- PCA (or other multivariate decompositions like ICA, NMF, clustering) produce components regardless of neurobiological meaning – without outside information for validation, i.e., association with a meaningful endpoint, the evidence is much weaker

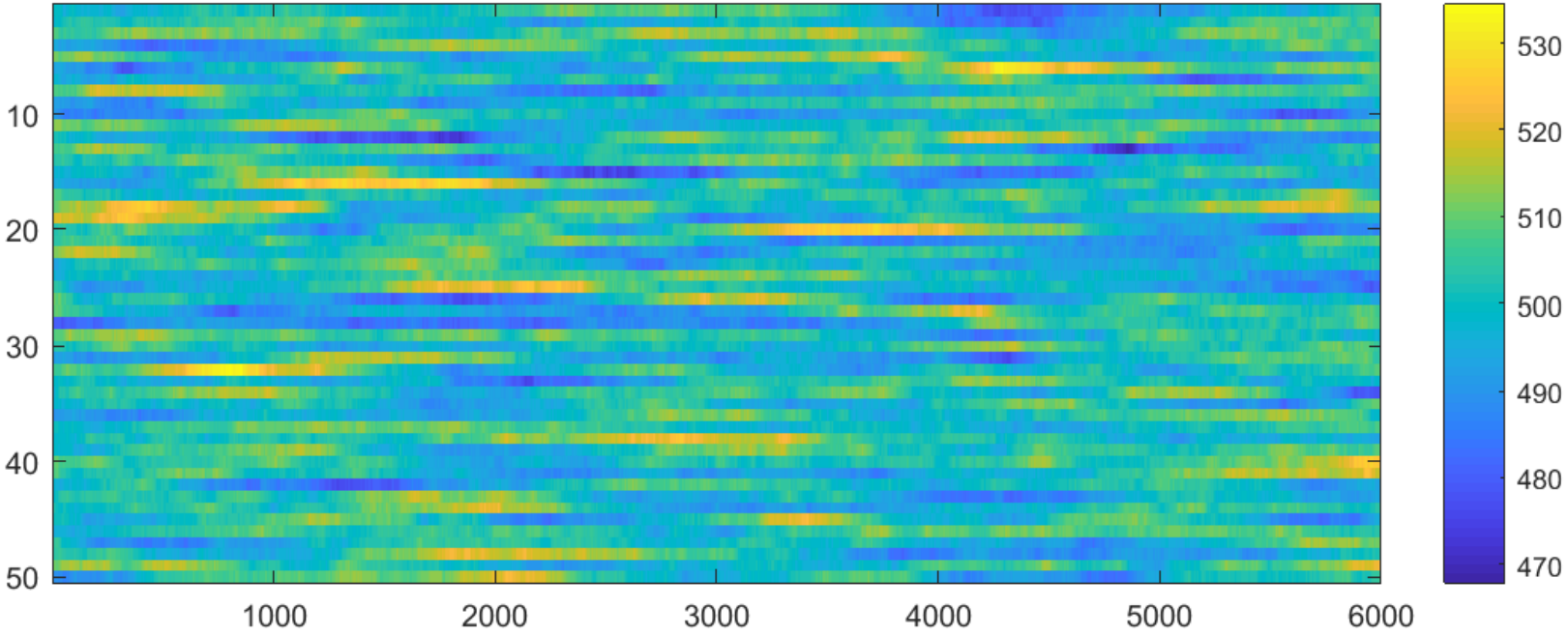
- Picking single principal components, particularly with small variance contribution, is problematic because of sampling variability and noise

**“One man's trash is another man's treasure” – PCA works on <u>every</u> kind of signal**

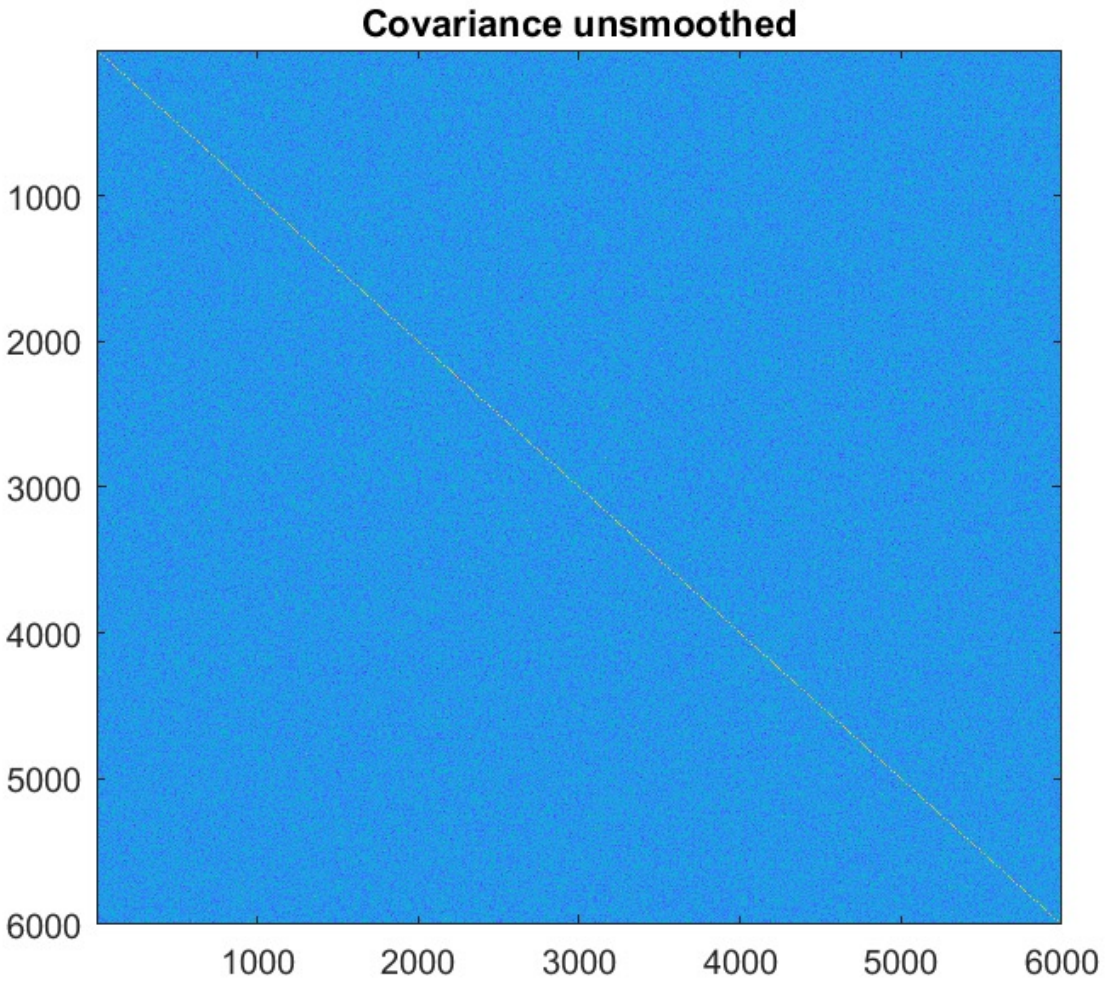
50 temporal null signals without correlation

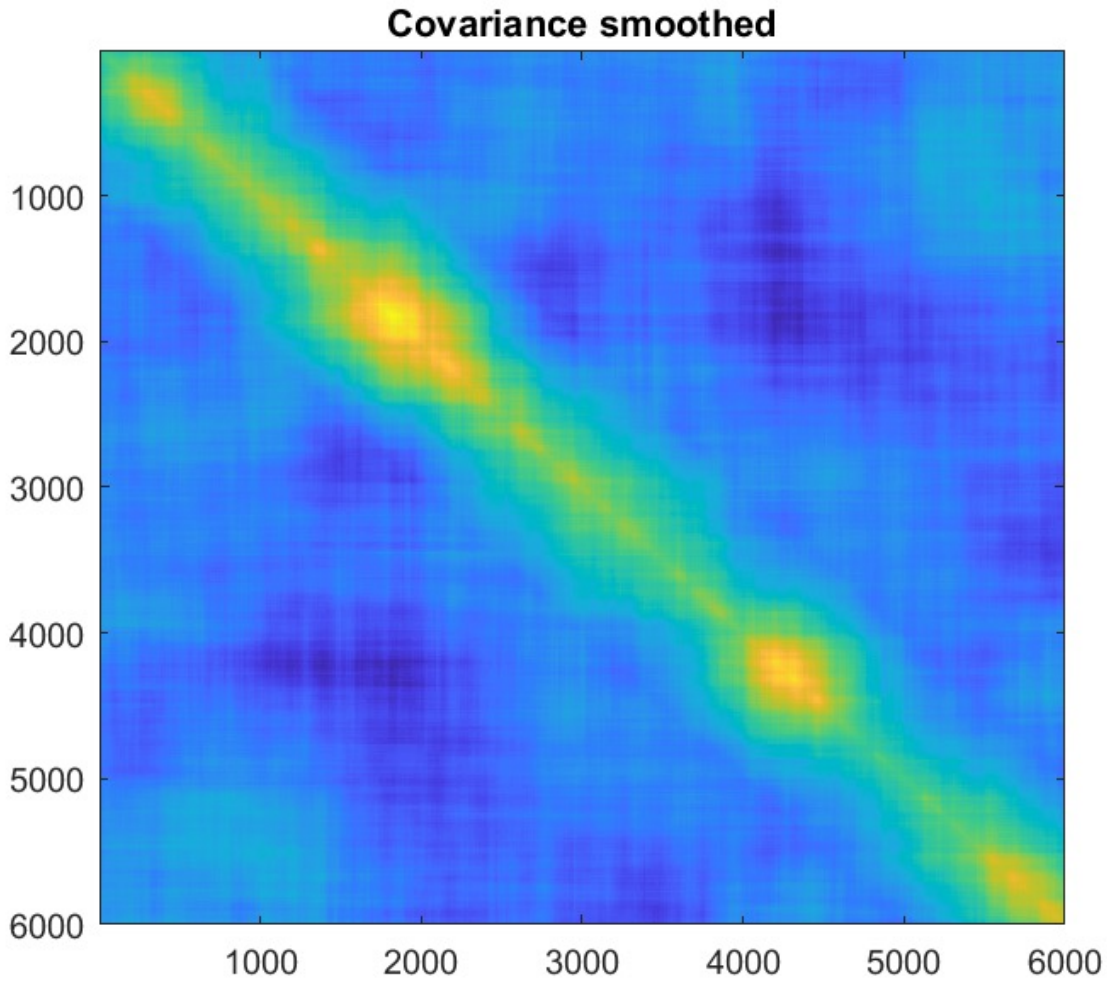



50 null signals after smoothing with a 1000-step moving window average



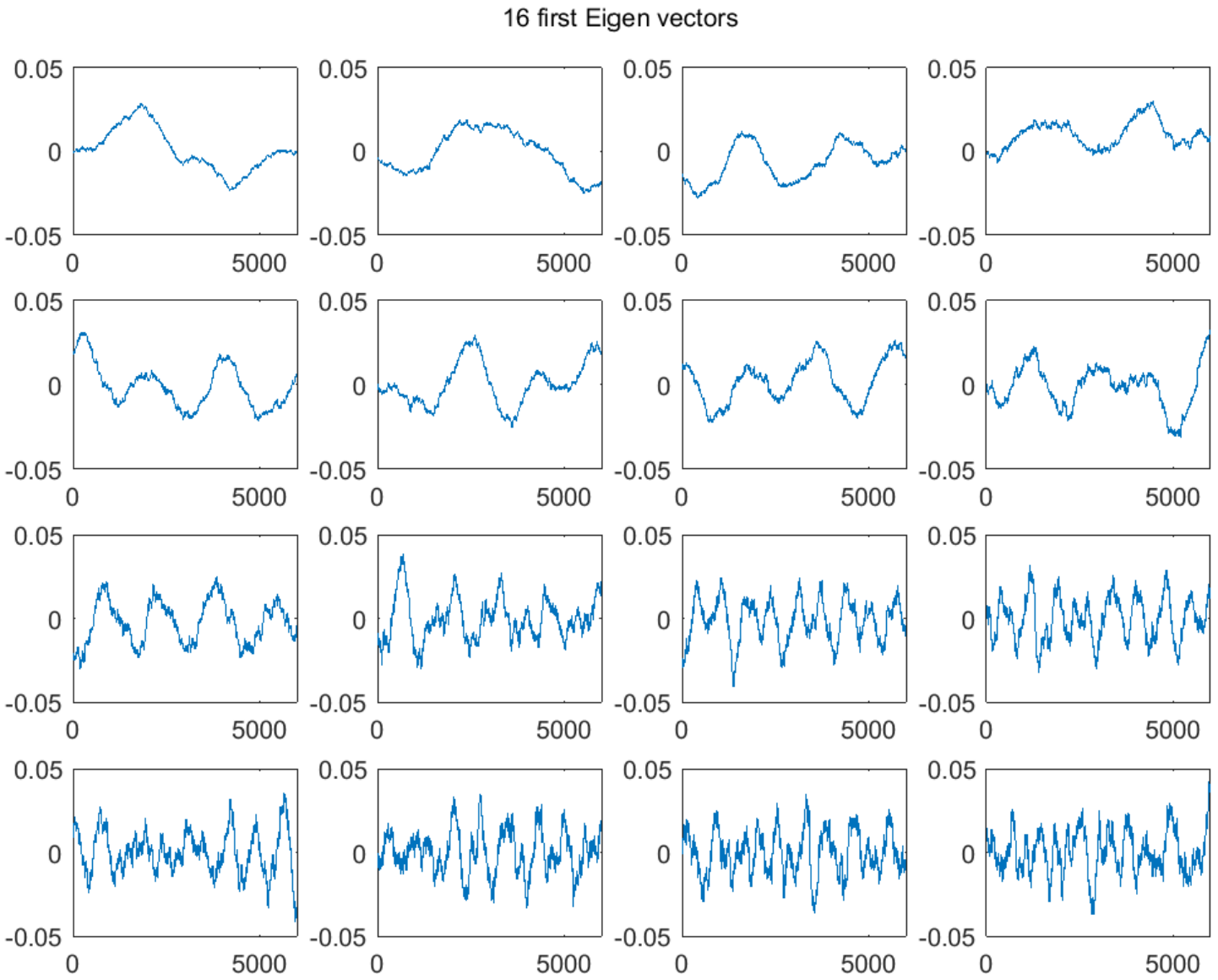
# Computing time-domain covariance matrix



**Eigen vectors show harmonics, despite any meaningful correlation structure in null signals**

**Picking one isolated component would yield meaningless sinusoidal signal**

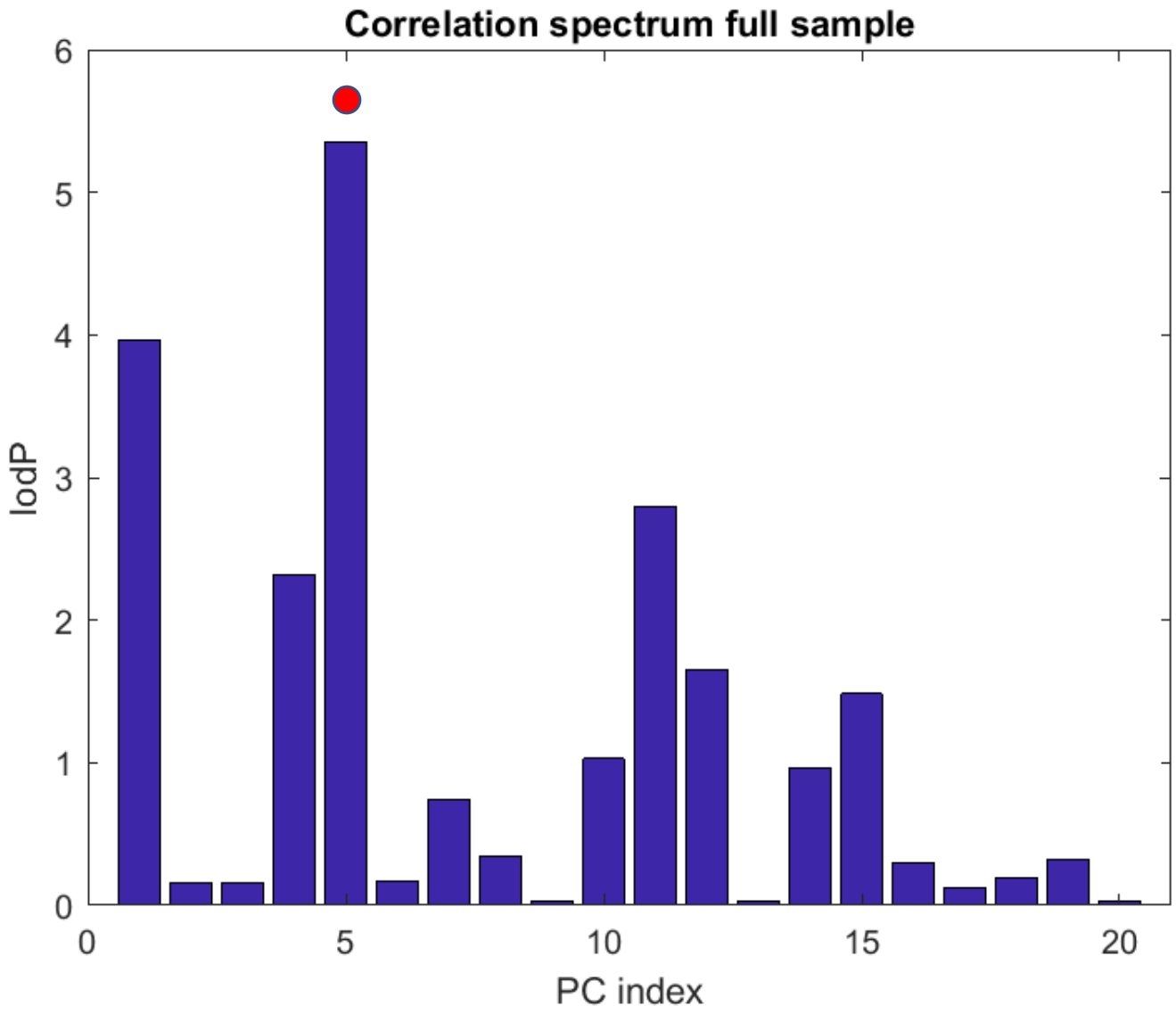



➔ **Simple safeguard: replication out of sample, including components loadings AND prediction of an endpoint!**

# Real-world example in a Fluid Reasoning task (“Paper Folding”) in 324 people aged 20-80

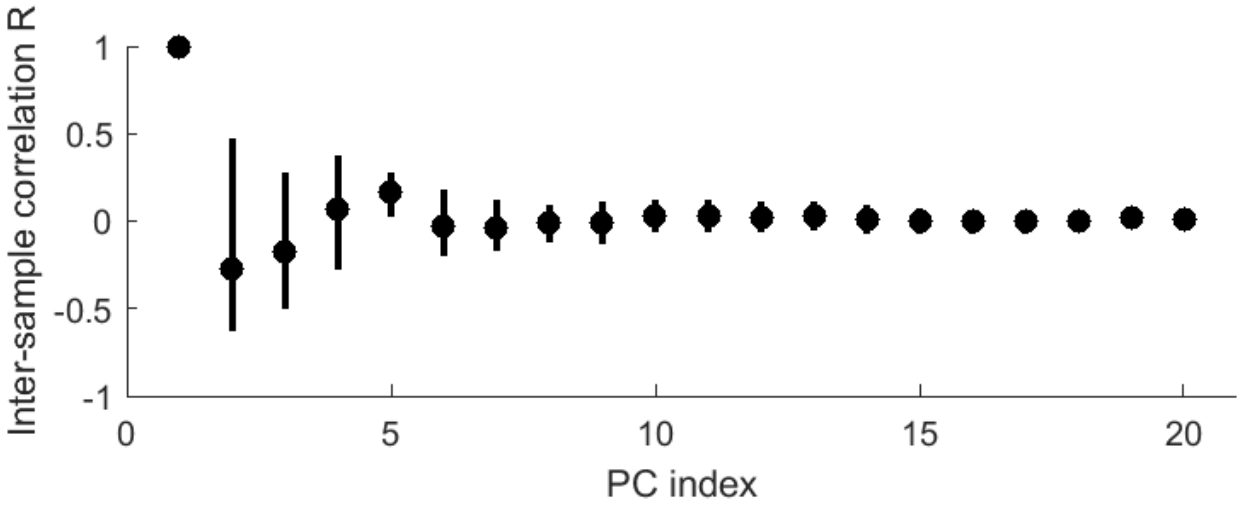
- Activation maps from event-related designs
- Recognition accuracy is the behavioral variable

➔ Perform PCA and check first 20 PC-scores and their correlation with recognition

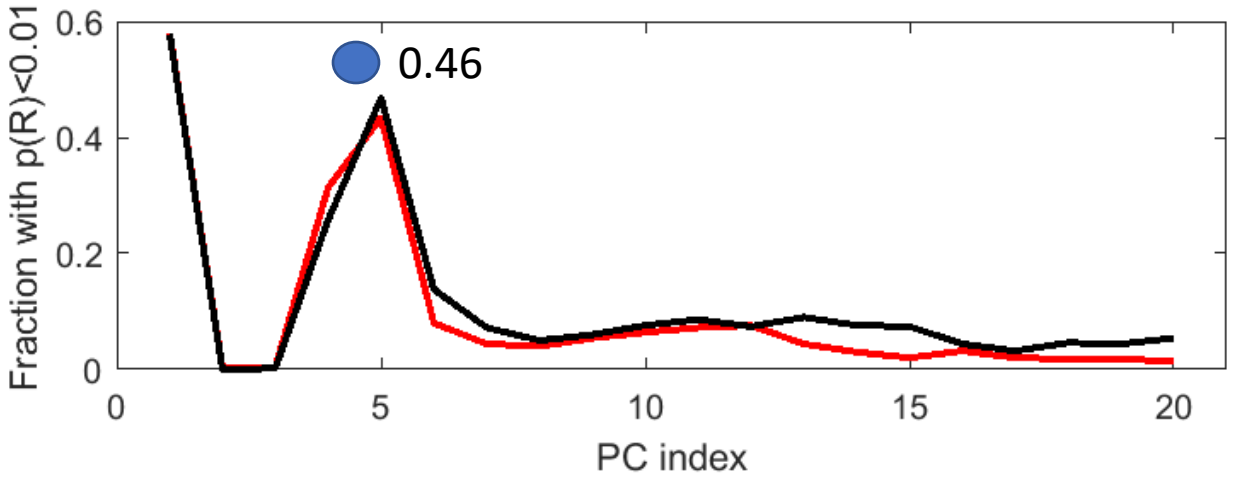


Divide randomly into non-overlapping training and test  of 150/150
Test behavioral correlation of subject scores in both samples, subject scores are obtained from PCs from training sample

**Spatial correlation between two PC-sets samples**




**Correlation with behavior in both training and test samples**



# Multivariate Analysis Framework

Multivariate Analysis framework

- Step 1: PCA on fMRI data

  $\mathbf{Y} = \mathbf{V}\,\mathbf{W}'$

  **Y=** fMRI Data**, V =** PC components**, W=** scores

- Step 2**:** Brain-behavior modeling

  $\mathbf{Cognition} = [\mathbf{W}(:,1{:}K)\ \ \mathbf{1}] * \boldsymbol{\beta} + \boldsymbol{\varepsilon}$

  K = number of included PCs determined with AIC or LOOCV

- Step 3**:** Construction of corresponding brain pattern

  $\mathbf{pattern} = \mathbf{V}(1{:}k) * \boldsymbol{\beta}(1{:}K)$

- Step 4: Application of brain behavioral model to held-out data **Z** and compute $R^2$

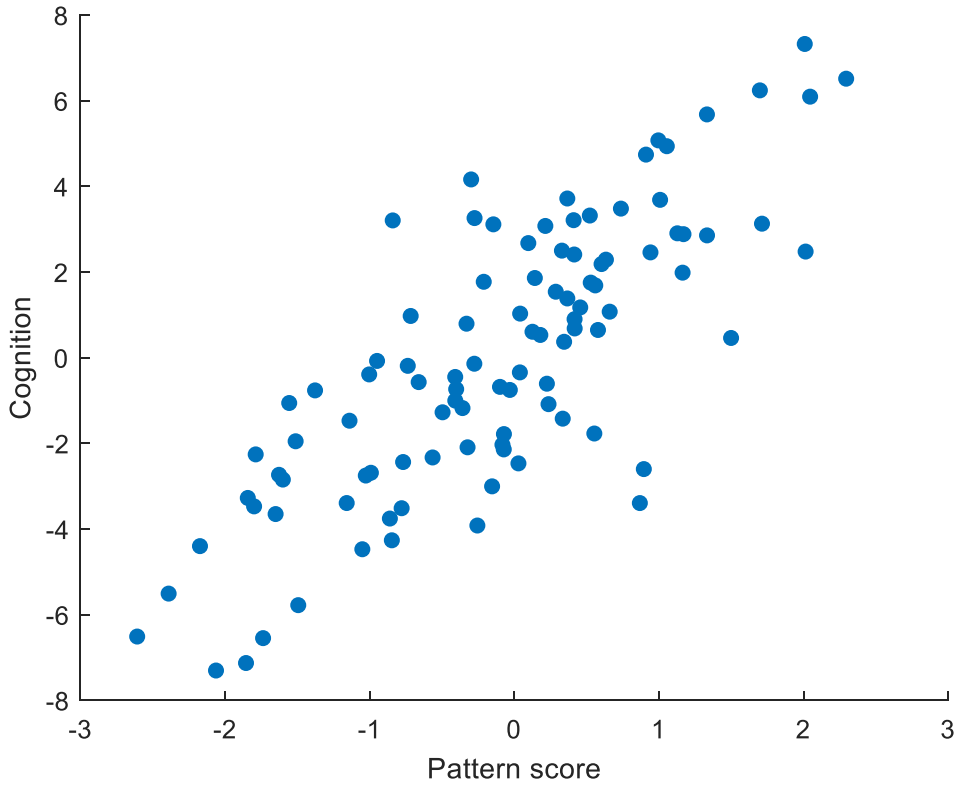
  **R( Z' pattern, cognition** in **Z)**

doi: 10.1038/jcbfm.1987.118.

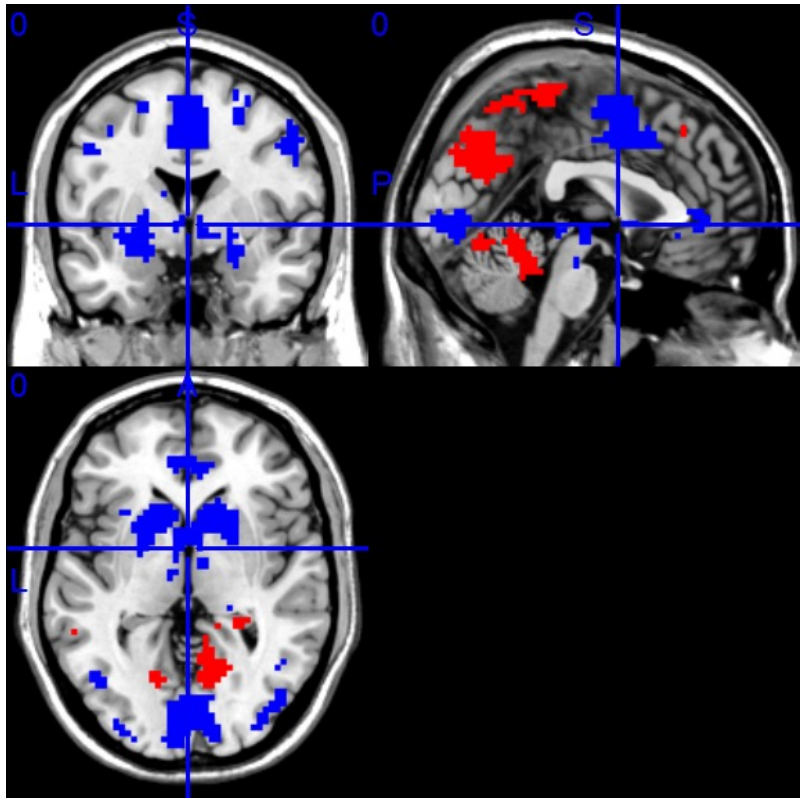
**Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data**

J R Moeller[1], S C Strother, J J Sidtis, D A Rottenberg

# SSM objectives and outcomes

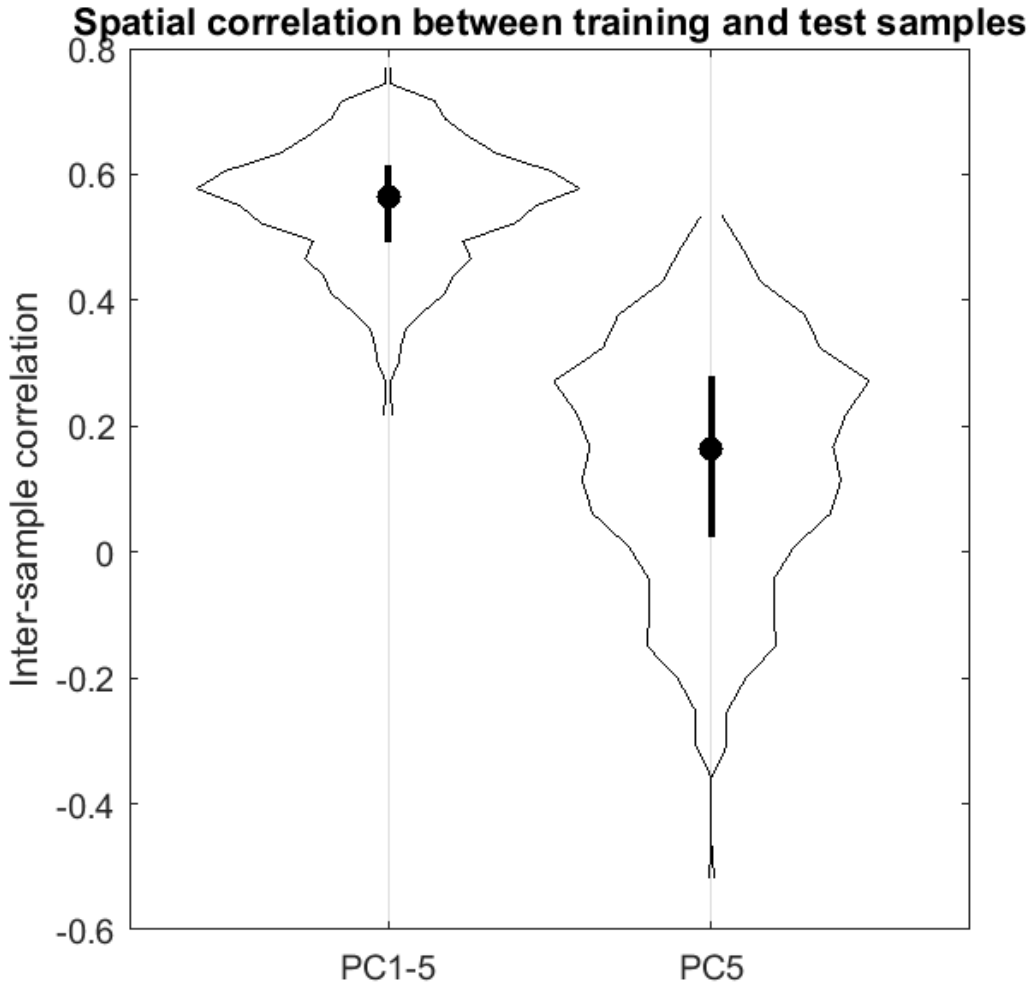


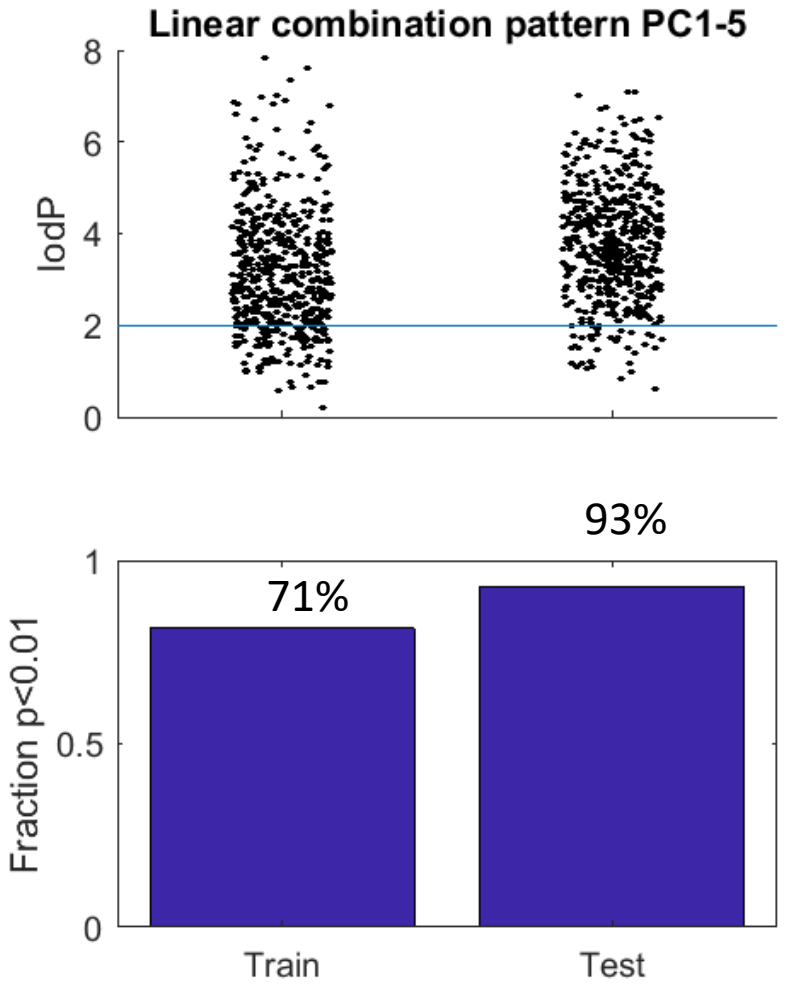

Topographic robustness assessed with bootstrap



$$Y = score * pattern' + \varepsilon$$

**Use set PC1-5 instead of just PC 5**




Replication of performance correlation in test sample: P(PC1-5) = 0.93,  P(PC5) = 0.46

# Data

fMRI Data Sample

- We use 3 fluid-reasoning tasks which have been acquired in 290 people in the context of the Reference ability neural network study, age 20-80

- The tasks are fMRI recognition versions of:
  - (1) **Matrix Reasoning** (= a visuo-spatial completion task like Raven's Advanced Matrices),
  - (2) **Letter Sets** (= 5 groups of letter sets with identification of one set being the odd one and violating an implicit to-be-discerned rule),
  - (3) **Paper Folding** (= depiction of a paper-folding process with subsequent punching by a stapler and a selection of possible punch-patterns from which the participant chooses)

- The cognitive outcome is the fraction of correct responses

# Univariate Analysis Framework

Univariate Analysis framework

- Step 1: Univariate regression

  **y ~ Y(:,k) + 1**

  k = voxel index

- Step 2: assembly of brain maps **β1**(k) and **β0**(k)

- Step 3: out-of-sample prediction of cognitive end point for each voxel

  **y**(:,k)=**Y**(:,k) **β1**(k) + **β0**(k)

- Step 4: form a "vote" that weights and averages the voxel prediction in held-out data according to whole-model significance F in training sample

- Step 5: Application of brain behavioral model to held-out data **Z** and compute $R^2$

  **R ( vote, cognition** in **Z)**

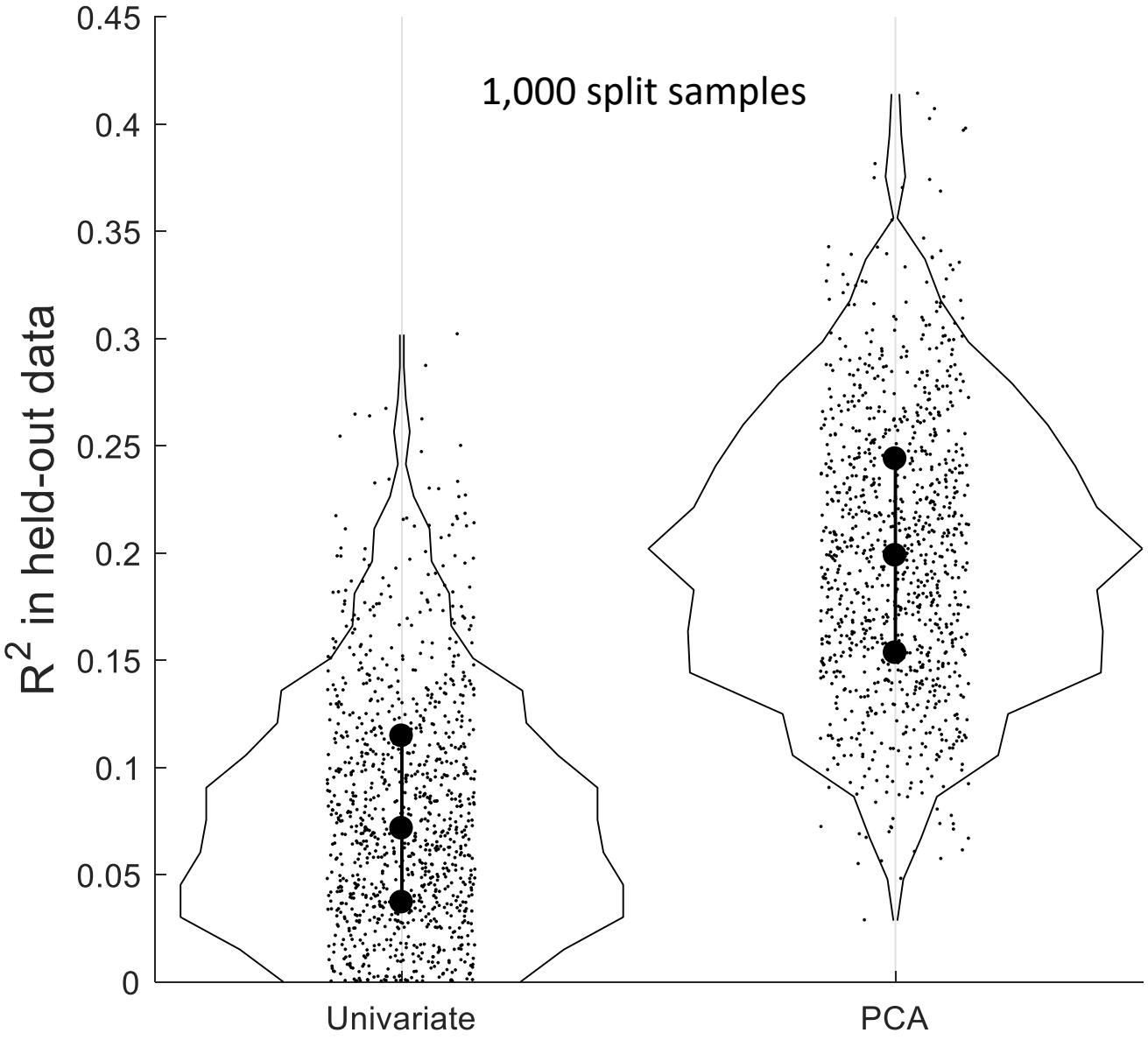
# Check in comparison to univariate approach in real-world fMRI data

Training N=600
Test N=100

290 subjects in 3 Fluid Reasoning tasks (=870 observations)

Endpoint: accuracy on recognition task

Compare SSM and univariate approach, with pooling of predictions weighted according to significance in training sample



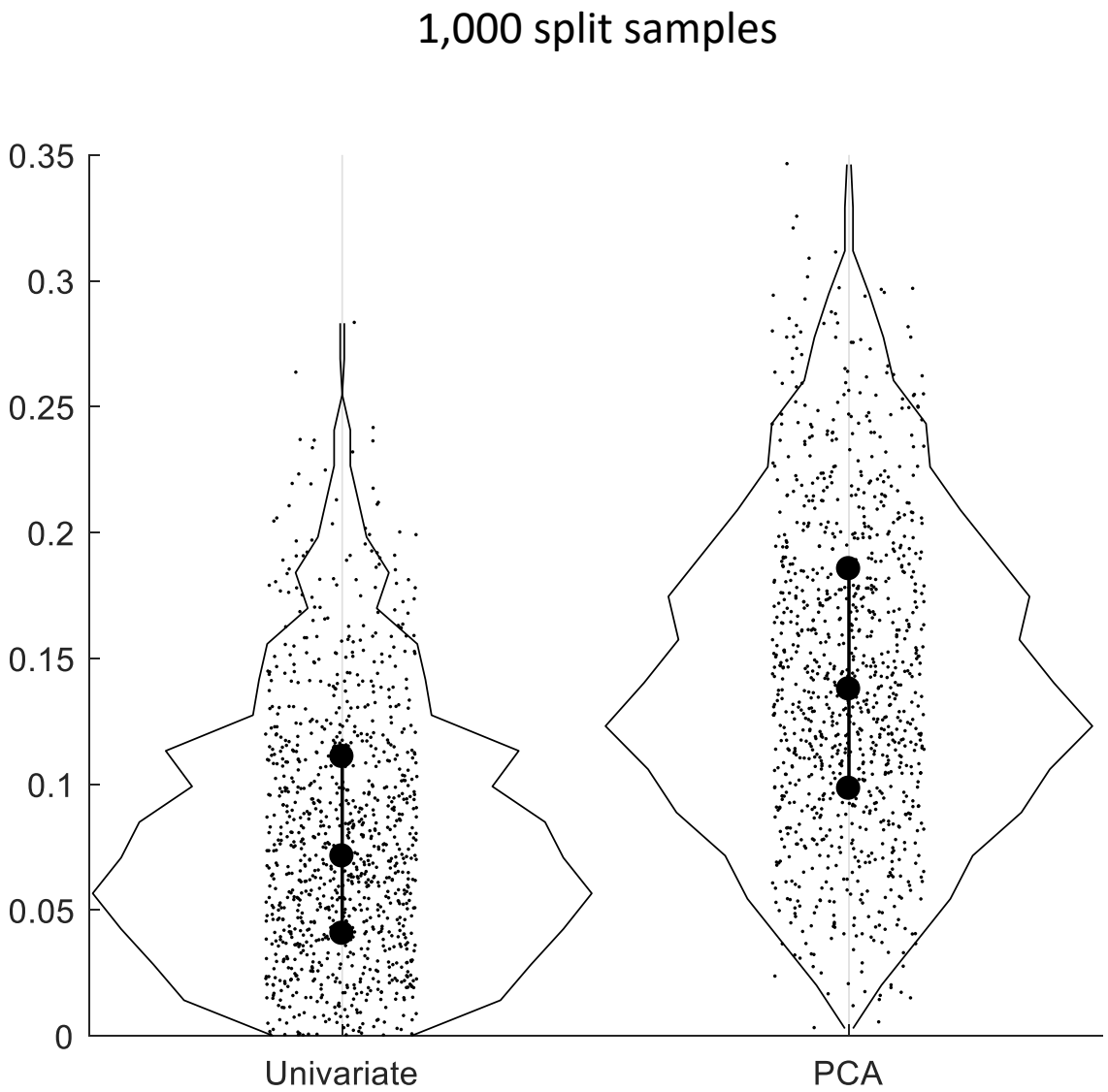
# Avoid rank deficiency and pick 500 randomly sampled voxels and repeat

Training N=600
Test N=100

290 subjects in 3 Fluid Reasoning tasks (=870 observations)

Endpoint: accuracy on recognition task

Compare SSM and univariate approach, with pooling of predictions weighted according to significance in training sample




➔ **PCA loses information, but still beats univariate prediction**

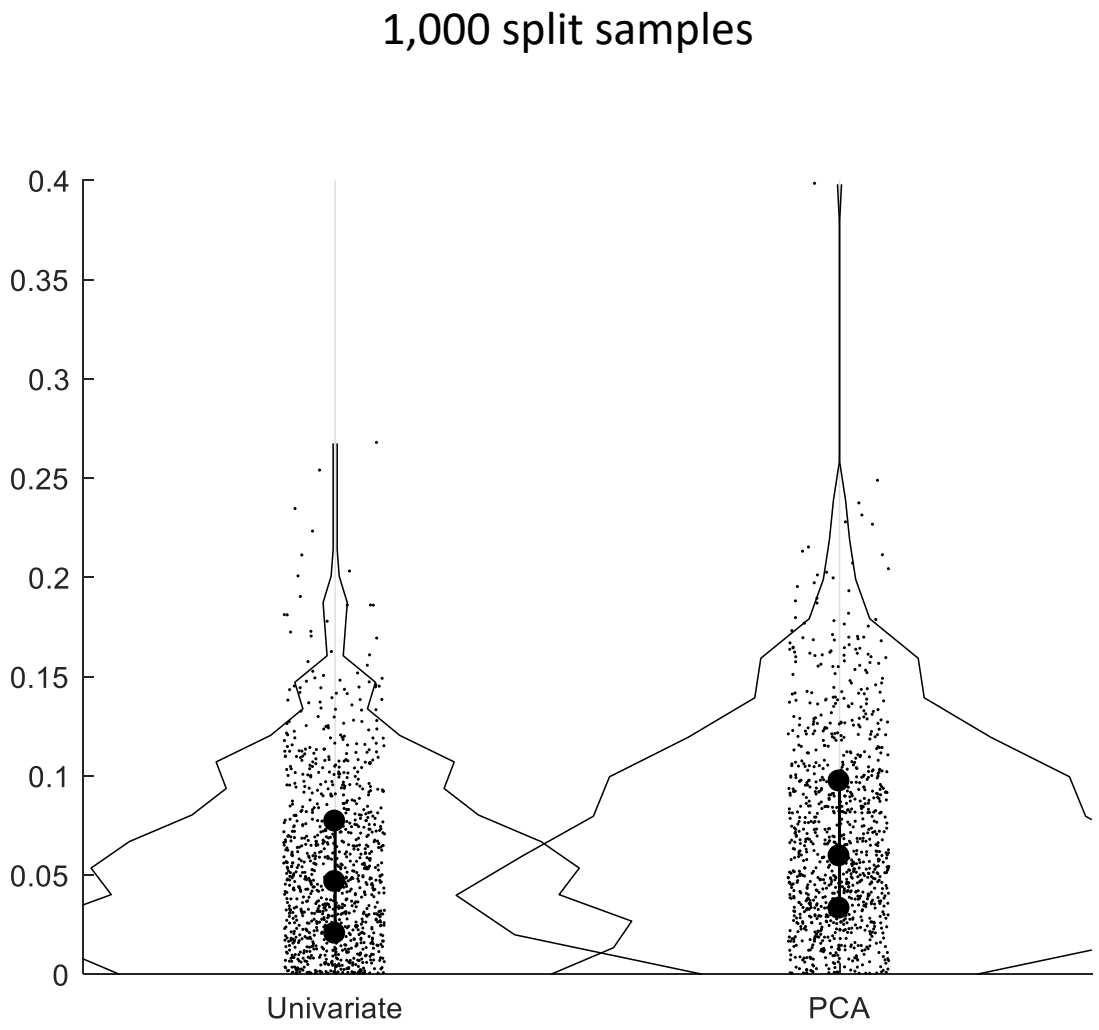
# Avoid rank deficiency and pick 50 randomly sampled voxels and repeat

Training N=600
Test N=100

290 subjects in 3 Fluid Reasoning tasks (=870 observations)

Endpoint: accuracy on recognition task

Compare SSM and univariate approach, with pooling of predictions weighted according to significance in training sample




➔ **PCA loses information and approaches univariate performance**

# Broad takeaway points

- PCA-regression (=SSM) technique achieves better held-out data replication for a Fluid reasoning  fMRI task

- This generalizes to other tasks as well (although I did not show the data)

- Reducing the rank deficiency degrades the multivariate predictive utility, whereas the univariate prediction does not suffer; however, multivariate analysis still performs better

- Sampling variability dictates to pick a set of principal components, rather than picking isolated ones

- Rank deficiency necessitates non-parametric statistics, like bootstrap

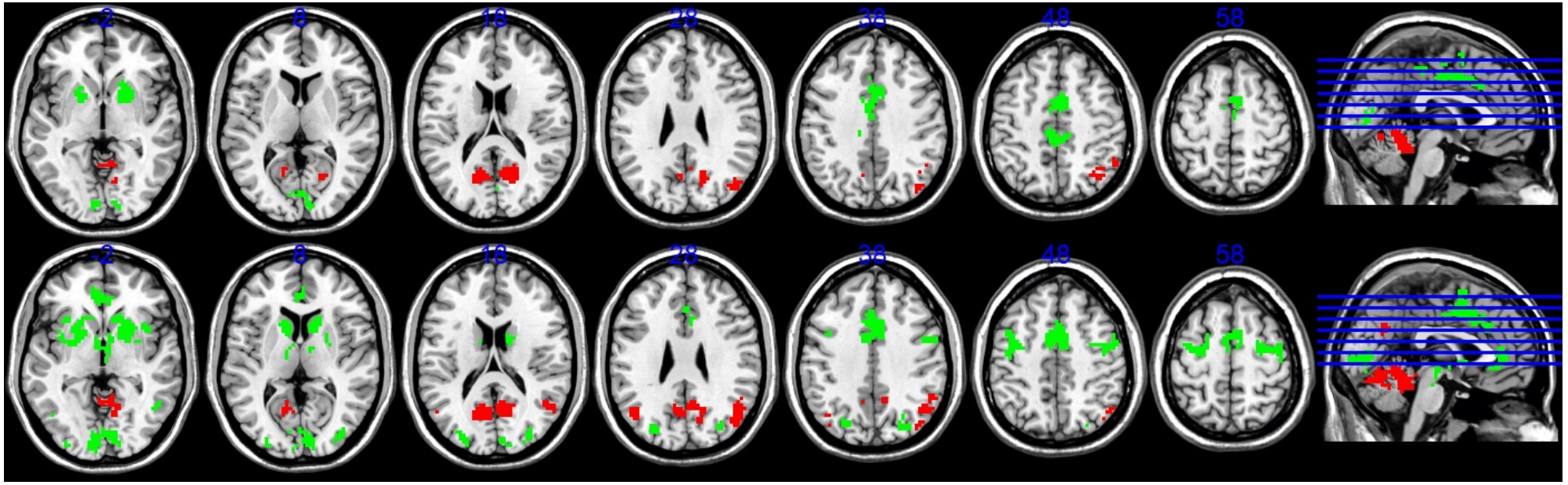
# **How to generate inferential brain maps when there is no parametric theory?** (number of voxels >> number of observations)

- Point estimate pattern coming from SSM technique

**v = SSM(Y,Cog)** - pseudo-function notation

- Perform semi-parametric bootstrap, i.e., resample with replacement ~100-1,000 times **v*=SSM(Y*,Cog*) and** observe variability about point estimate and generate Z-robustness map

**Z(voxel) ~ loading(voxel) / STD of variability around loading (voxel)**




|Z|>2

P=0.0083
FDR=0.05

# Conclusions

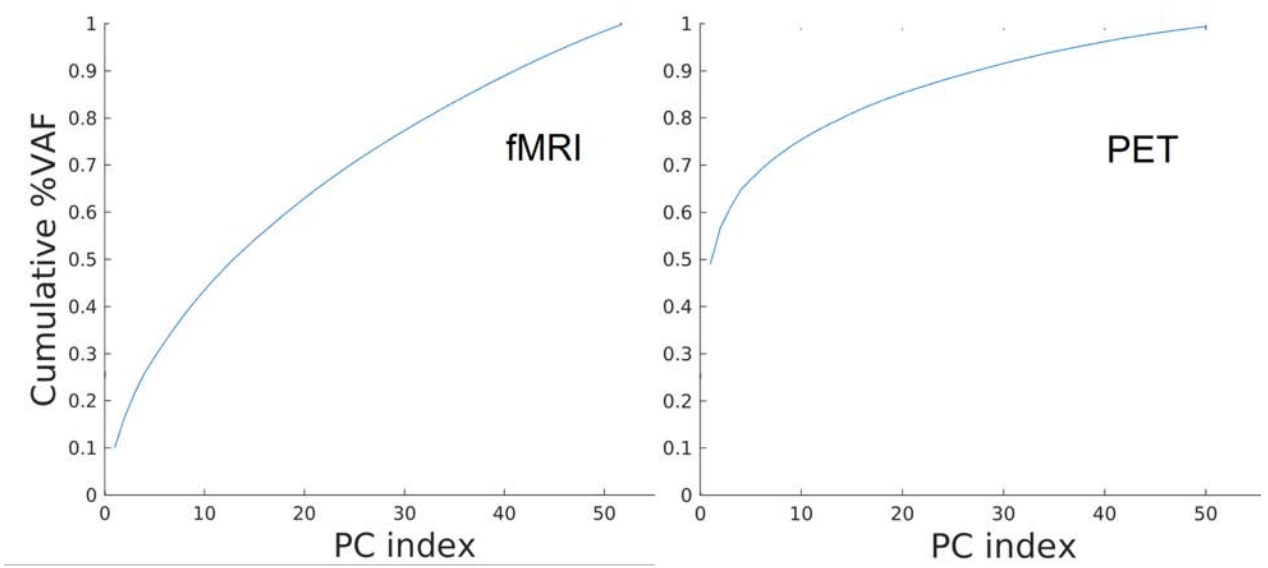
- Multivariate techniques with PCA-regression (=SSM) mark the most basic complication of traditional univariate analysis

- As a “shallow” and one-shot technique, PCA-regression offers relatively easy inferential statistics with non-parametric resampling tests, and topographic information is easily rendered
  - Attractiveness for medium sample sizes that do not yet permit deep-learning and CNNs
  - Ensemble methods like bootstrap aggregating (BAGGING) or boosting can be additionally applied for better predictive utility to better accommodate low signal-to-noise ratios

- Good reference benchmark to evaluate Deep-Learning networks in terms of predictive utility gain

# Implications for molecular connectivity?

- MC potentially offers betters ground-truth knowledge about regions via specific tracers and offers better variance concentration than fMRI




*Courtesy of Igor Yakushev*

*Yakushev, Drzezga, Habeck, 2018*

- For region-specific ground-truth of specific nuclei (Raphe, Nuclear Accumbens, Locus Coeruleus), seed locations could be used as “endpoints” for the derivation of patterns with the SSM (PCA-regression) technique, and further tested with mediation analysis for the relation to symptom scales/cognition

Functional notation: (pattern, pattern score) = SSM(Y, y-seed)

y-seed ➔ pattern score ➔ symptom scales/cognition

# Thank you for your attention

**ch629@columbia.edu**