

# fMRI Reproducibility in *R*

**Brian B. Avants**

PENN Image Computing & Science Laboratory  
Dept. of Radiology, University of Pennsylvania  
Philadelphia, PA, 19104 <sup>1</sup>

November 7, 2013

---

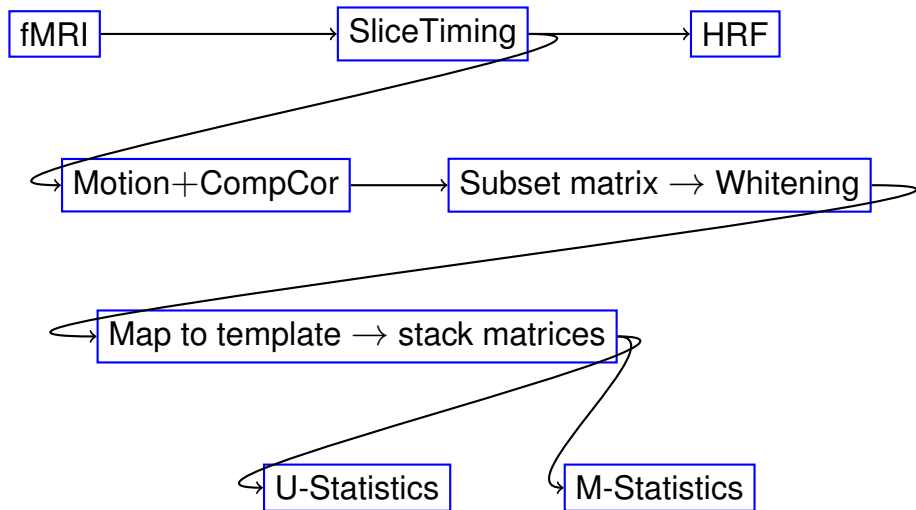
<sup>1</sup>for KRNS project

This presentation is copyrighted by  
The **ANTs software consortium**

distributed under the  
**Creative Commons by Attribution License 3.0**

http:  
[//creativecommons.org/licenses/by/3.0](http://creativecommons.org/licenses/by/3.0)

# ANTsR group-wise fMRI: BOBO<sup>2</sup>



# Reproducibility Datasets

Dataset 1 : [Gorgolewski](#)  $n=10$

Dataset 2 : [Duncan](#)  $n > 35$

```
run1 <- rnorm(100)
run2 <- rnorm(100)
cor.test(run1, run2)

##
## Pearson's product-moment correlation
##
## data:  run1 and run2
## t = -1.112, df = 98, p-value = 0.2689
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.30142  0.08671
## sample estimates:
##      cor
## -0.1116
```

# R Processing

Use *R* to test processing strategies for fMRI

- ▶ Pre-processing **minimal connectome strategies**
- ▶ Univariate: GLM with **CompCor** and **ANTs** motion correction.
- ▶ Multivariate: fMRI application of **SCCAN**
- ▶ Multi/Univariate use same pre-processing and includes outlier detection based on global signal.
- ▶ Employ a group-wise fixed effects analysis — requires mapping to a common template space.
- ▶ Template is BOLD and includes AAL neuroanatomical labels.

# R Processing

Use *R* to test processing strategies for fMRI

- ▶ Pre-processing **minimal connectome strategies**
- ▶ Univariate: GLM with **CompCor** and **ANTs** motion correction.
- ▶ Multivariate: fMRI application of **SCCAN**
- ▶ Multi/Univariate use same pre-processing and includes outlier detection based on global signal.
- ▶ Employ a group-wise fixed effects analysis — requires mapping to a common template space.
- ▶ Template is BOLD and includes AAL neuroanatomical labels.

# R Processing

Use *R* to test processing strategies for fMRI

- ▶ Pre-processing **minimal connectome strategies**
- ▶ Univariate: GLM with **CompCor** and **ANTs** motion correction.
- ▶ Multivariate: fMRI application of **SCCAN**
- ▶ Multi/Univariate use same pre-processing and includes outlier detection based on global signal.
- ▶ Employ a group-wise fixed effects analysis — requires mapping to a common template space.
- ▶ Template is BOLD and includes AAL neuroanatomical labels.

# R Processing

Use *R* to test processing strategies for fMRI

- ▶ Pre-processing [minimal connectome strategies](#)
- ▶ Univariate: GLM with [CompCor](#) and [ANTs](#) motion correction.
- ▶ Multivariate: fMRI application of [SCCAN](#)
- ▶ Multi/Univariate use same pre-processing and includes outlier detection based on global signal.
- ▶ Employ a group-wise fixed effects analysis — requires mapping to a common template space.
- ▶ Template is BOLD and includes AAL neuroanatomical labels.



# R Processing

Use *R* to test processing strategies for fMRI

- ▶ Pre-processing [minimal connectome strategies](#)
- ▶ Univariate: GLM with [CompCor](#) and [ANTs](#) motion correction.
- ▶ Multivariate: fMRI application of [SCCAN](#)
- ▶ Multi/Univariate use same pre-processing and includes outlier detection based on global signal.
- ▶ Employ a group-wise fixed effects analysis — requires mapping to a common template space.
- ▶ Template is BOLD and includes AAL neuroanatomical labels.

# R Processing

Use *R* to test processing strategies for fMRI

- ▶ Pre-processing [minimal connectome strategies](#)
- ▶ Univariate: GLM with [CompCor](#) and [ANTs](#) motion correction.
- ▶ Multivariate: fMRI application of [SCCAN](#)
- ▶ Multi/Univariate use same pre-processing and includes outlier detection based on global signal.
- ▶ Employ a group-wise fixed effects analysis — requires mapping to a common template space.
- ▶ Template is BOLD and includes AAL neuroanatomical labels.

# Signal Reproducibility Measurements

## Validation Mechanisms

- ▶ Univar: Threshold statistical map consistently: top ~~x% of  $\beta$  map~~  $\rightarrow$  threshold at constant  $\beta = 1.5, 2.0, 2.5 \dots$  across two different runs.
- ▶ Multivar: Measure spatial coincidence of sparse multivariate predictors across runs.
- ▶ Measure signal overlap (Dice) between clusters in thresholded beta map.
- ▶ ~~Minimum distance sum (MDS) between clusters in thresholded beta map between two runs~~

## Prediction

Subject-level and group-wise feature selection followed by subject-specific prediction.

# Signal Reproducibility Measurements

## Validation Mechanisms

- ▶ Univar: Threshold statistical map consistently: top ~~x% of  $\beta$  map~~  $\rightarrow$  threshold at constant  $\beta = 1.5, 2.0, 2.5 \dots$  across two different runs.
- ▶ Multivar: Measure spatial coincidence of sparse multivariate predictors across runs.
- ▶ Measure signal overlap (Dice) between clusters in thresholded beta map.
- ▶ ~~Minimum distance sum (MDS) between clusters in thresholded beta map between two runs~~

## Prediction

Subject-level and group-wise feature selection followed by subject-specific prediction.

# Signal Reproducibility Measurements

## Validation Mechanisms

- ▶ Univar: Threshold statistical map consistently: top  $x\%$  of  $\beta$  map  $\rightarrow$  threshold at constant  $\beta = 1.5, 2.0, 2.5 \dots$  across two different runs.
- ▶ Multivar: Measure spatial coincidence of sparse multivariate predictors across runs.
- ▶ Measure signal overlap (Dice) between clusters in thresholded beta map.
- ▶ Minimum distance sum (MDS) between clusters in thresholded beta map between two runs

## Prediction

Subject-level and group-wise feature selection followed by subject-specific prediction.

# Signal Reproducibility Measurements

## Validation Mechanisms

- ▶ Univar: Threshold statistical map consistently: top  $x\%$  of  $\beta$  map  $\rightarrow$  threshold at constant  $\beta = 1.5, 2.0, 2.5 \dots$  across two different runs.
- ▶ Multivar: Measure spatial coincidence of sparse multivariate predictors across runs.
- ▶ Measure signal overlap (Dice) between clusters in thresholded beta map.
- ▶ Minimum distance sum (MDS) between clusters in thresholded beta map between two runs

## Prediction

Subject-level and group-wise feature selection followed by subject-specific prediction.

# Signal Reproducibility Measurements

## Validation Mechanisms

- ▶ Univar: Threshold statistical map consistently: top  $x\%$  of  $\beta$  map  $\rightarrow$  threshold at constant  $\beta = 1.5, 2.0, 2.5 \dots$  across two different runs.
- ▶ Multivar: Measure spatial coincidence of sparse multivariate predictors across runs.
- ▶ Measure signal overlap (Dice) between clusters in thresholded beta map.
- ▶ Minimum distance sum (MDS) between clusters in thresholded beta map between two runs

## Prediction

Subject-level and group-wise feature selection followed by subject-specific prediction.

# Analysis Scripts in RfMRI

- ▶ `ants_2_template.sh` — maps BOLD to template space.
- ▶ `process_bold.R` — bold processing for one run ... outputs hrf, matrices and (thresholded) beta maps
- ▶ After above steps, `univar_multivar_fmri_consistency.sh` produces group-wise reproducibility numbers.
- ▶ Models are of form:  $\text{voxel} \approx \text{hrf} + \text{motion1} + \text{motion2} + \text{motion3} + \text{compcor1} + \text{compcor2} + \text{compcor3} + \text{globalsignal} + \text{SubjectID}$
- ▶ Input data is the "stacked" matrix i.e. if we have  $n$  subjects, each with a  $t \times p$  ( time by space ) matrix , then the input matrix for this study would be of size  $nt \times p$ .



# Analysis Scripts in RfMRI

- ▶ `ants_2_template.sh` — maps BOLD to template space.
- ▶ `process_bold.R` — bold processing for one run ... outputs hrf, matrices and (thresholded) beta maps
- ▶ After above steps, `univar_multivar_fmri_consistency.sh` produces group-wise reproducibility numbers.
- ▶ Models are of form:  $\text{voxel} \approx \text{hrf} + \text{motion1} + \text{motion2} + \text{motion3} + \text{compcor1} + \text{compcor2} + \text{compcor3} + \text{globalsignal} + \text{SubjectID}$
- ▶ Input data is the "stacked" matrix i.e. if we have  $n$  subjects, each with a  $t \times p$  ( time by space ) matrix , then the input matrix for this study would be of size  $nt \times p$ .

# Analysis Scripts in RfMRI

- ▶ `ants_2_template.sh` — maps BOLD to template space.
- ▶ `process_bold.R` — bold processing for one run ... outputs hrf, matrices and (thresholded) beta maps
- ▶ After above steps, `univar_multivar_fmri_consistency.sh` produces group-wise reproducibility numbers.
- ▶ Models are of form:  $\text{voxel} \approx \text{hrf} + \text{motion1} + \text{motion2} + \text{motion3} + \text{compcor1} + \text{compcor2} + \text{compcor3} + \text{globalsignal} + \text{SubjectID}$
- ▶ Input data is the "stacked" matrix i.e. if we have  $n$  subjects, each with a  $t \times p$  ( time by space ) matrix , then the input matrix for this study would be of size  $nt \times p$ .

# Analysis Scripts in RfMRI

- ▶ `ants_2_template.sh` — maps BOLD to template space.
- ▶ `process_bold.R` — bold processing for one run ... outputs hrf, matrices and (thresholded) beta maps
- ▶ After above steps, `univar_multivar_fmri_consistency.sh` produces group-wise reproducibility numbers.
- ▶ Models are of form:  $\text{voxel} \approx \text{hrf} + \text{motion1} + \text{motion2} + \text{motion3} + \text{compcor1} + \text{compcor2} + \text{compcor3} + \text{globalsignal} + \text{SubjectID}$
- ▶ Input data is the "stacked" matrix i.e. if we have  $n$  subjects, each with a  $t \times p$  ( time by space ) matrix , then the input matrix for this study would be of size  $nt \times p$ .

# Analysis Scripts in RfMRI

- ▶ `ants_2_template.sh` — maps BOLD to template space.
- ▶ `process_bold.R` — bold processing for one run ... outputs hrf, matrices and (thresholded) beta maps
- ▶ After above steps, `univar_multivar_fmri_consistency.sh` produces group-wise reproducibility numbers.
- ▶ Models are of form:  $\text{voxel} \approx \text{hrf} + \text{motion1} + \text{motion2} + \text{motion3} + \text{compcor1} + \text{compcor2} + \text{compcor3} + \text{globalsignal} + \text{SubjectID}$
- ▶ Input data is the "stacked" matrix i.e. if we have  $n$  subjects, each with a  $t \times p$  ( time by space ) matrix , then the input matrix for this study would be of size  $nt \times p$ .

# Repeatability: Univariate

10 subjects

- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do  $\beta$  maps overlap? Depends on threshold and task.
- ▶ Finger tapping overlap: Max  $\approx 0.4$  at threshold Y.
- ▶ Covert verb: Less sensitive to threshold. Max  $\approx 0.7$  at threshold 2.

# Repeatability: Univariate

10 subjects

- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do  $\beta$  maps overlap? Depends on threshold and task.
- ▶ Finger tapping overlap: Max  $\approx 0.4$  at threshold Y.
- ▶ Covert verb: Less sensitive to threshold. Max  $\approx 0.7$  at threshold 2.

# Repeatability: Univariate

10 subjects

- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do  $\beta$  maps overlap? Depends on threshold and task.
- ▶ Finger tapping overlap: Max  $\approx 0.4$  at threshold Y.
- ▶ Covert verb: Less sensitive to threshold. Max  $\approx 0.7$  at threshold 2.

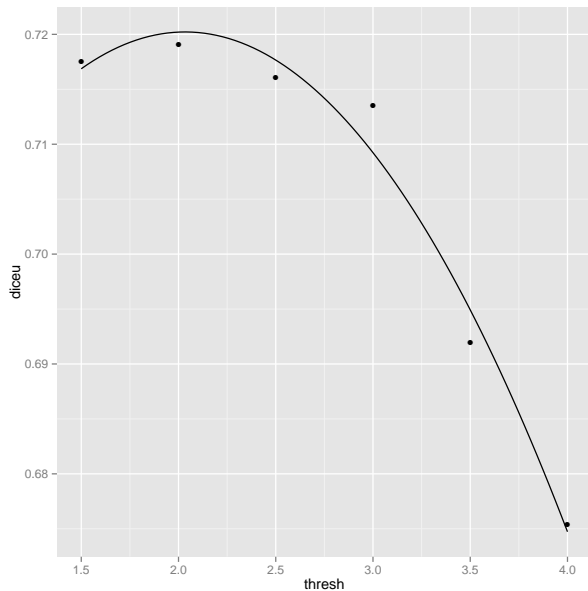
# Repeatability: Univariate

10 subjects

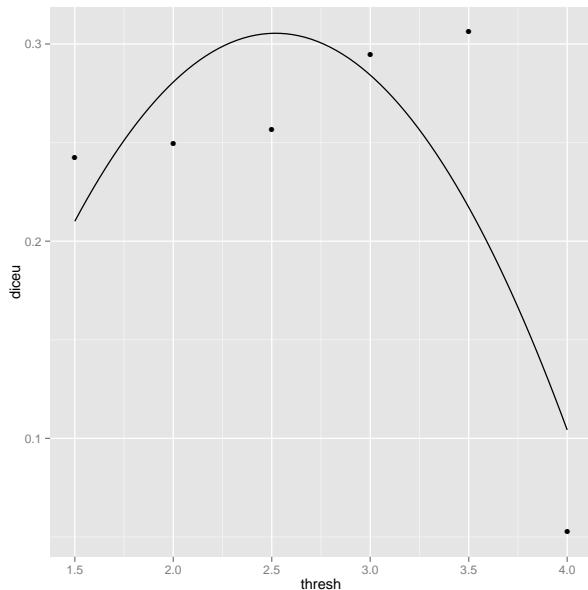
- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do  $\beta$  maps overlap? Depends on threshold and task.
- ▶ Finger tapping overlap: Max  $\approx 0.4$  at threshold Y.
- ▶ Covert verb: Less sensitive to threshold. Max  $\approx 0.7$  at threshold 2.



# Univariate Overlap: Covert Verb



# Univariate Overlap: Finger Tapping



# Repeatability: Multivariate

10 subjects

- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do sparse-components overlap? Depends on threshold and task.
- ▶ A sparse component,  $u$ , maximizes  $\text{PearsonCorrelation}(Xu, \text{HRF}_{\text{model}(s)}v) - \omega \|u\|_1$  where we  $\ell_1$  penalize non-zero components of  $u$  and  $X$  is the BOLD matrix.
- ▶ Finger tapping overlap: "focal network," max  $\approx 0.7$  at threshold 2% of brain.
- ▶ Covert verb: "bigger network," max  $\approx 0.8$  at threshold 12% of brain.

# Repeatability: Multivariate

10 subjects

- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do sparse-components overlap? Depends on threshold and task.
- ▶ A sparse component,  $u$ , maximizes  $\text{PearsonCorrelation}(Xu, \text{HRF}_{\text{model(s)}}v) - \omega \|u\|_1$  where we  $\ell_1$  penalize non-zero components of  $u$  and  $X$  is the BOLD matrix.
- ▶ Finger tapping overlap: "focal network,"  $\text{max} \approx 0.7$  at threshold 2% of brain.
- ▶ Covert verb: "bigger network,"  $\text{max} \approx 0.8$  at threshold 12% of brain.

# Repeatability: Multivariate

10 subjects

- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do sparse-components overlap? Depends on threshold and task.
- ▶ A sparse component,  $u$ , maximizes  $\text{PearsonCorrelation}(Xu, \text{HRF}_{\text{model(s)}}v) - \omega \|u\|_1$  where we  $\ell_1$  penalize non-zero components of  $u$  and  $X$  is the BOLD matrix.
- ▶ Finger tapping overlap: "focal network," max  $\approx 0.7$  at threshold 2% of brain.
- ▶ Covert verb: "bigger network," max  $\approx 0.8$  at threshold 12% of brain.

# Repeatability: Multivariate

10 subjects

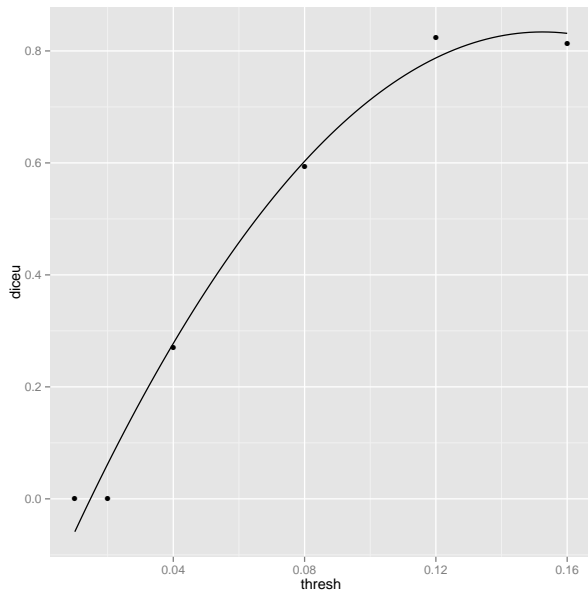
- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do sparse-components overlap? Depends on threshold and task.
- ▶ A sparse component,  $u$ , maximizes  $\text{PearsonCorrelation}(Xu, \text{HRF}_{\text{model(s)}}v) - \omega \|u\|_1$  where we  $\ell_1$  penalize non-zero components of  $u$  and  $X$  is the BOLD matrix.
- ▶ Finger tapping overlap: "focal network,"  $\text{max} \approx 0.7$  at threshold 2% of brain.
- ▶ Covert verb: "bigger network,"  $\text{max} \approx 0.8$  at threshold 12% of brain.

# Repeatability: Multivariate

10 subjects

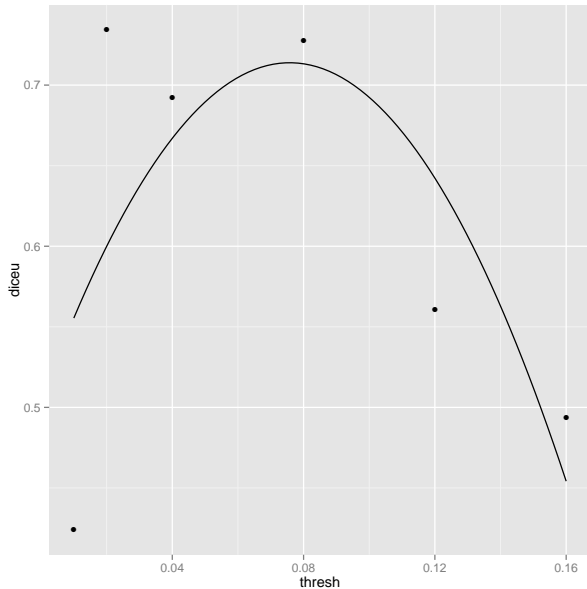
- ▶ Analyzed covert verb generation and finger tapping.
- ▶ Do sparse-components overlap? Depends on threshold and task.
- ▶ A sparse component,  $u$ , maximizes  $\text{PearsonCorrelation}(Xu, \text{HRF}_{\text{model(s)}}v) - \omega \|u\|_1$  where we  $\ell_1$  penalize non-zero components of  $u$  and  $X$  is the BOLD matrix.
- ▶ Finger tapping overlap: "focal network," max  $\approx 0.7$  at threshold 2% of brain.
- ▶ Covert verb: "bigger network," max  $\approx 0.8$  at threshold 12% of brain.

# Multivariate Overlap: Covert Verb



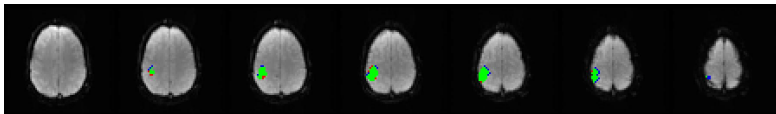


# Multivariate Overlap: Finger Tapping



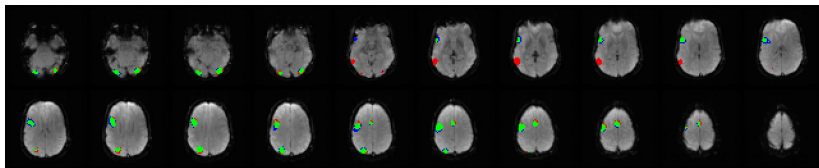
# Finger Tapping: M-Spatial Maps

Figure : Finger tapping reproducibility — 3 color, red:run1, blue:run2, green:both. Dice  $\approx 0.87$ .



# Covert Verb Generation: M-Spatial Maps

Figure : Covert verb generation reproducibility — 3 color, red:run1, blue:run2, green:both. Dice  $\approx 0.83$ .



# Discussion

- ▶ Multivariate approach appears to be more repeatable: 0.8 vs 0.7 (covert verb) and 0.7 vs 0.4 (finger tapping).
- ▶ This may be due to the data-driven and spatially-informed smoothing within the multivariate optimization.
- ▶ Optimized spatial smoothing for univariate data may work better.
- ▶ Spatiotemporal smoothing improves overlap for both univariate (vg:0.74, ft:0.38) & multivariate (vg:0.83, ft: 0.87) approaches.

# Discussion

- ▶ Multivariate approach appears to be more repeatable: 0.8 vs 0.7 (covert verb) and 0.7 vs 0.4 (finger tapping).
- ▶ This may be due to the data-driven and spatially-informed smoothing within the multivariate optimization.
- ▶ Optimized spatial smoothing for univariate data may work better.
- ▶ Spatiotemporal smoothing improves overlap for both univariate (vg:0.74, ft:0.38) & multivariate (vg:0.83, ft: 0.87) approaches.

# Discussion

- ▶ Multivariate approach appears to be more repeatable: 0.8 vs 0.7 (covert verb) and 0.7 vs 0.4 (finger tapping).
- ▶ This may be due to the data-driven and spatially-informed smoothing within the multivariate optimization.
- ▶ Optimized spatial smoothing for univariate data may work better.
- ▶ Spatiotemporal smoothing improves overlap for both univariate (vg:0.74, ft:0.38) & multivariate (vg:0.83, ft: 0.87) approaches.

# Discussion

- ▶ Multivariate approach appears to be more repeatable: 0.8 vs 0.7 (covert verb) and 0.7 vs 0.4 (finger tapping).
- ▶ This may be due to the data-driven and spatially-informed smoothing within the multivariate optimization.
- ▶ Optimized spatial smoothing for univariate data may work better.
- ▶ Spatiotemporal smoothing improves overlap for both univariate (vg:0.74, ft:0.38) & multivariate (vg:0.83, ft: 0.87) approaches.

# Discussion 2

- ▶ SCCAN allows combination of block and HRF models to drive prediction — may also help reproducibility.
- ▶ Subject-level reproducibility: *not bad* (needs work to quantify further via prediction).
- ▶ Many processing decisions made by observation, knowledge from literature and educated guesses.



## Discussion 2

- ▶ SCCAN allows combination of block and HRF models to drive prediction — may also help reproducibility.
- ▶ Subject-level reproducibility: *not bad* (needs work to quantify further via prediction).
- ▶ Many processing decisions made by observation, knowledge from literature and educated guesses.

## Discussion 2

- ▶ SCCAN allows combination of block and HRF models to drive prediction — may also help reproducibility.
- ▶ Subject-level reproducibility: *not bad* (needs work to quantify further via prediction).
- ▶ Many processing decisions made by observation, knowledge from literature and educated guesses.

# Summary of Findings

## Novel BOLD-fMRI study comparing the reliability of multivariate and univariate group-level analysis

- ▶ Both finger-tapping and CovVerbGen result in Dice overlap  $> 0.83$ . The upper limit in reliability is *unknown*.<sup>a</sup>
- ▶ CovVerbGen involves a more distributed network of effects and is therefore less reliable yet is consistent @ both group-level and subject-level.

---

<sup>a</sup>Including GlobalSignal appears to increase or, at minimum, not hurt, reliability.

# Summary of Findings

## Novel BOLD-fMRI study comparing the reliability of multivariate and univariate group-level analysis

- ▶ Both finger-tapping and CovVerbGen result in Dice overlap  $> 0.83$ . The upper limit in reliability is *unknown*.<sup>a</sup>
- ▶ CovVerbGen involves a more distributed network of effects and is therefore less reliable yet is consistent @ both group-level and subject-level.

---

<sup>a</sup>Including GlobalSignal appears to increase or, at minimum, not hurt, reliability.

# Future Work

- ▶ Study impact of pre-processing for both univariate and multivariate approaches.
- ▶ Random effects in univariate? Optimized smoothing in univariate?
- ▶ Test decoding of 3 simple tasks with both univariate and multivariate feature selection: *“dictionary” or “template”*?
- ▶ Given results of above, process Haxby data.

# Future Work

- ▶ Study impact of pre-processing for both univariate and multivariate approaches.
- ▶ Random effects in univariate? Optimized smoothing in univariate?
- ▶ Test decoding of 3 simple tasks with both univariate and multivariate feature selection: *“dictionary” or “template”*?
- ▶ Given results of above, process Haxby data.

# Future Work

- ▶ Study impact of pre-processing for both univariate and multivariate approaches.
- ▶ Random effects in univariate? Optimized smoothing in univariate?
- ▶ Test decoding of 3 simple tasks with both univariate and multivariate feature selection: *“dictionary” or “template”*?
- ▶ Given results of above, process Haxby data.

# Future Work

- ▶ Study impact of pre-processing for both univariate and multivariate approaches.
- ▶ Random effects in univariate? Optimized smoothing in univariate?
- ▶ Test decoding of 3 simple tasks with both univariate and multivariate feature selection: *“dictionary” or “template”*?
- ▶ Given results of above, process Haxby data.



# Trivial Preliminary Results: Decoding $S1$

Train on Run 1 and test on Run 2 — finger, mouth, foot interleaved with rest. Prediction from group and subject features via  $R$  and  $SCCAN$ .

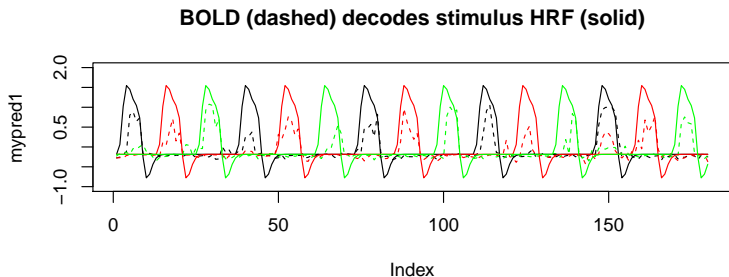


Figure : Subject 1.

Prior constraints and voting will improve results.

# Trivial Preliminary Results: Decoding S2

Train on Run 1 and test on Run 2 — finger, mouth, foot interleaved with rest. Prediction from group and subject features via  $R$  and  $SCCAN$ .

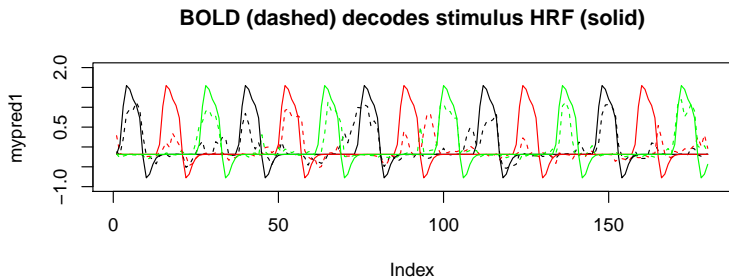


Figure : Subject 2.

Prior constraints and voting will improve results.