Nonparametric methods for correcting the multiple comparisons problem in classification-based fMRI

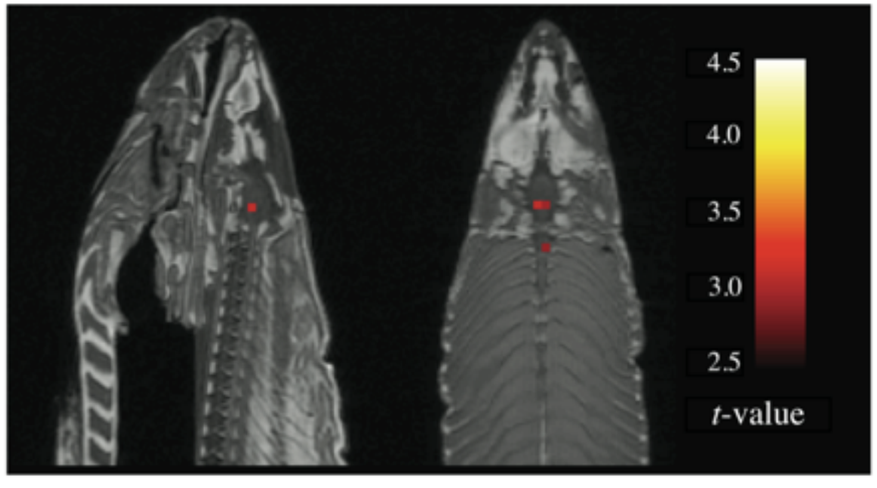
Johannes Stelzer, 17th of June 2015

whole-brain map of classifier's estimates (e.g. decoding accuracies) → thresholded map indicating statistical significance

Assuming the classifier "guesses":
what is the likelihood to observe
a given decoding accuracy (or larger)?

how to derive
null **distribution**

how to correct
for **multiple** testing



how to derive null distribution *parametrically*:

- binomial models
- t-based stats

real label	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
classifier's estimate	b	a	a	b	b	a	a	b	a	a	b	b	a	b	a	b	a	b	b	a

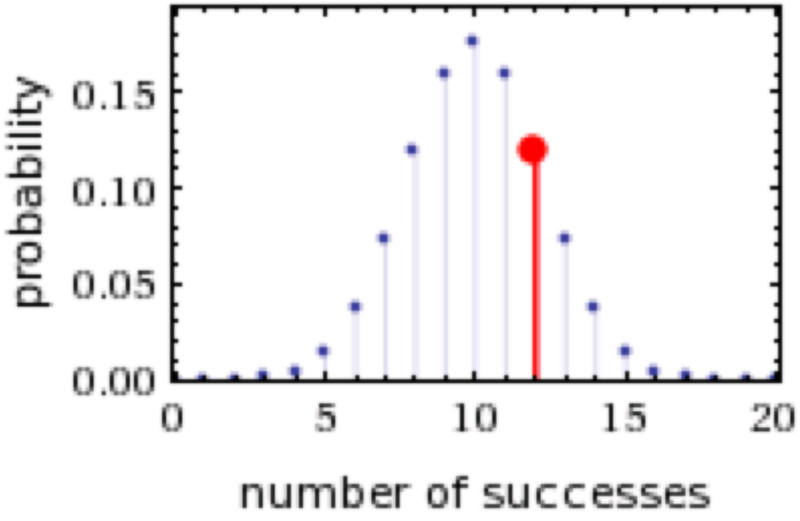
number of samples **N** = 20

correct samples **c** = 12

accuracy = 0.6

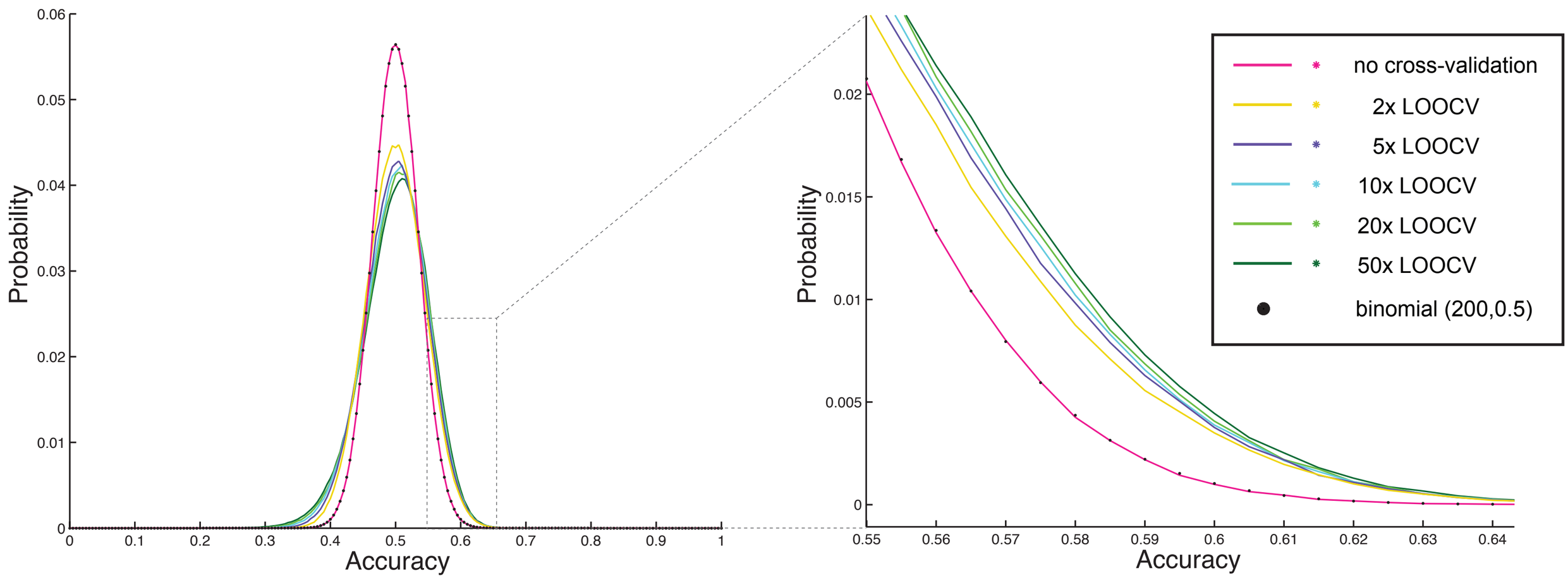
binomial model:

$$f_X(c) = \binom{N}{c} p^c (1-p)^{N-c}$$



real label	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b	a	b
classifier's estimate	b	a	a	b	b	a	a	b	a	a	b	b	a	b	a	b	a	b	b	a
2x CV											T	T	T	T	T	T	T	T	T	T
4x CV																T	T	T	T	T
5x CV																	T	T	T	T
10x CV																			T	T
20x CV																				T

test set T



Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063

problems regarding binomial model

- dependency between cross-validation folds undermines assumption of independency between trials
- (ignoring above problem): solution for one single test - multiple testing solution needed!

t-based methods and random fields procedure

- get voxel-wise accuracy map for each subject
- one sample t-test against 0.5 (2 classes)
- apply gaussian random-fields cluster level correction

problems regarding t-based framework

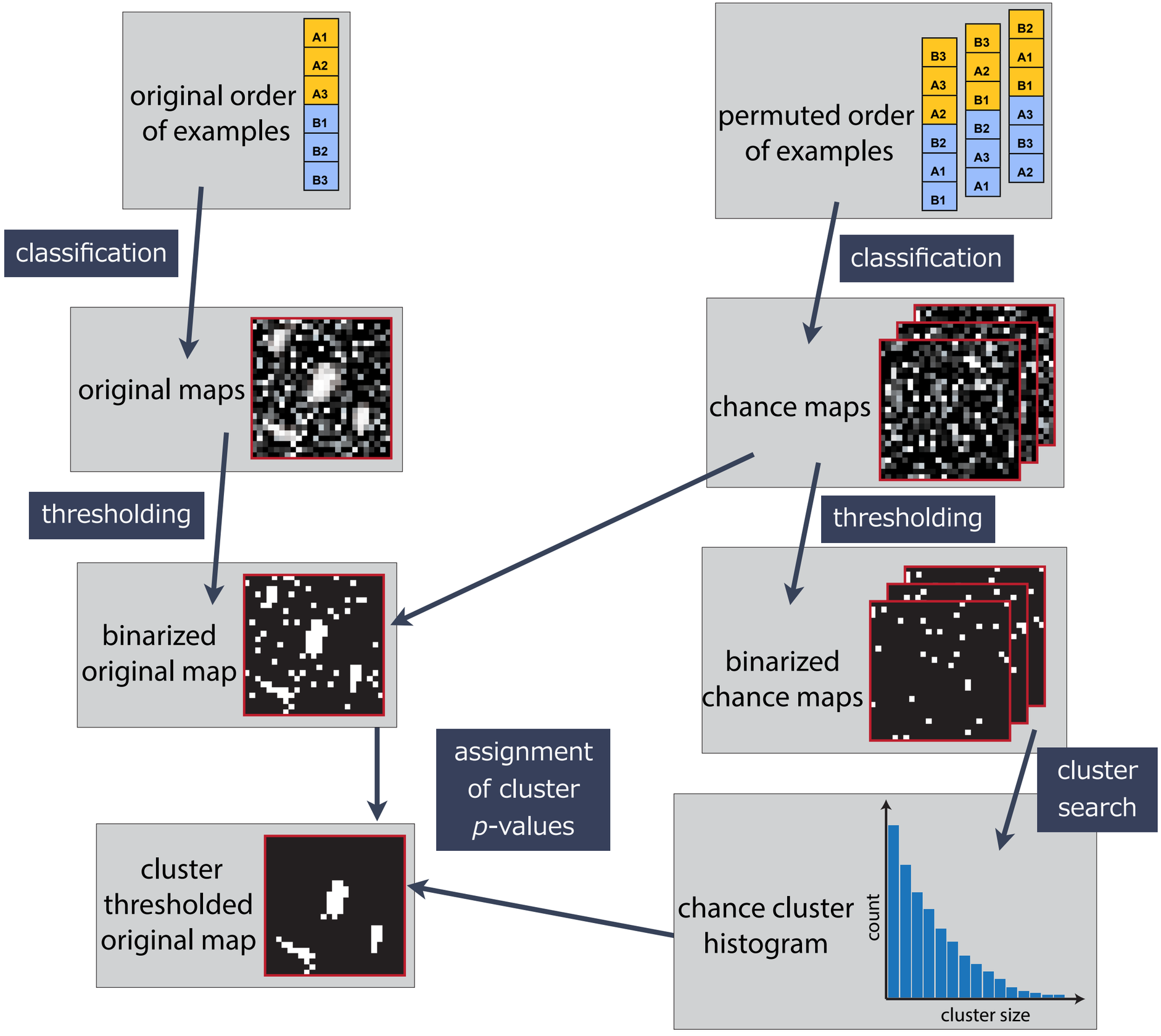
- decoding accuracies discrete (t-test assumes continuous)
- accuracies don’t follow normal distribution
- assumptions regarding random field theory & estimation of spatial smoothness not met

advantages of permutation testing

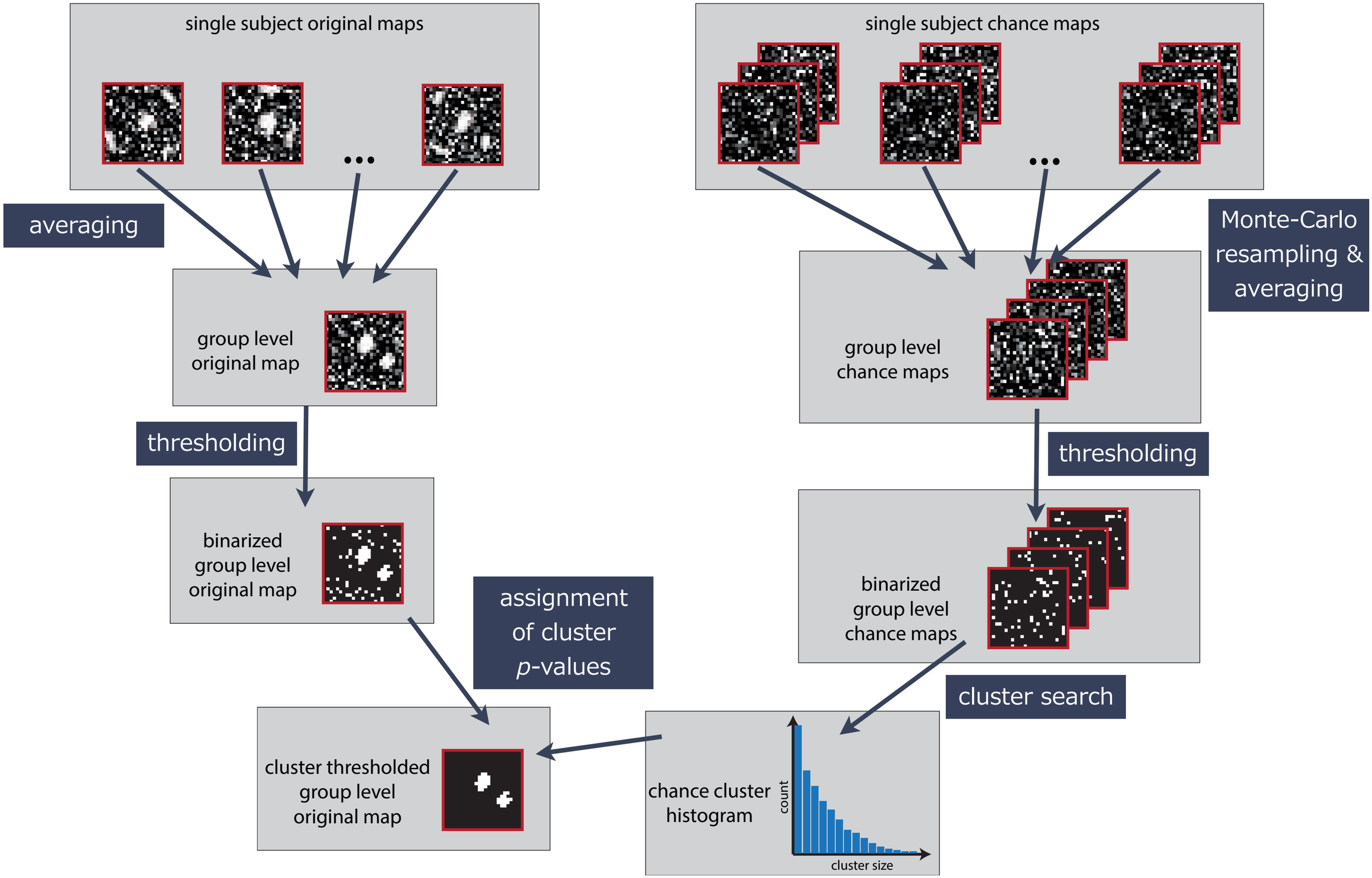
- (practically) assumption free
- yields voxel-wise null distribution
- spatial “chance maps” for free!

Golland, P., Liang, F., Mukherjee, S., and Panchenko, D. (2005). Permutation Tests for Classification. 3559, 501–515. doi: 10.1007/11503415_34.

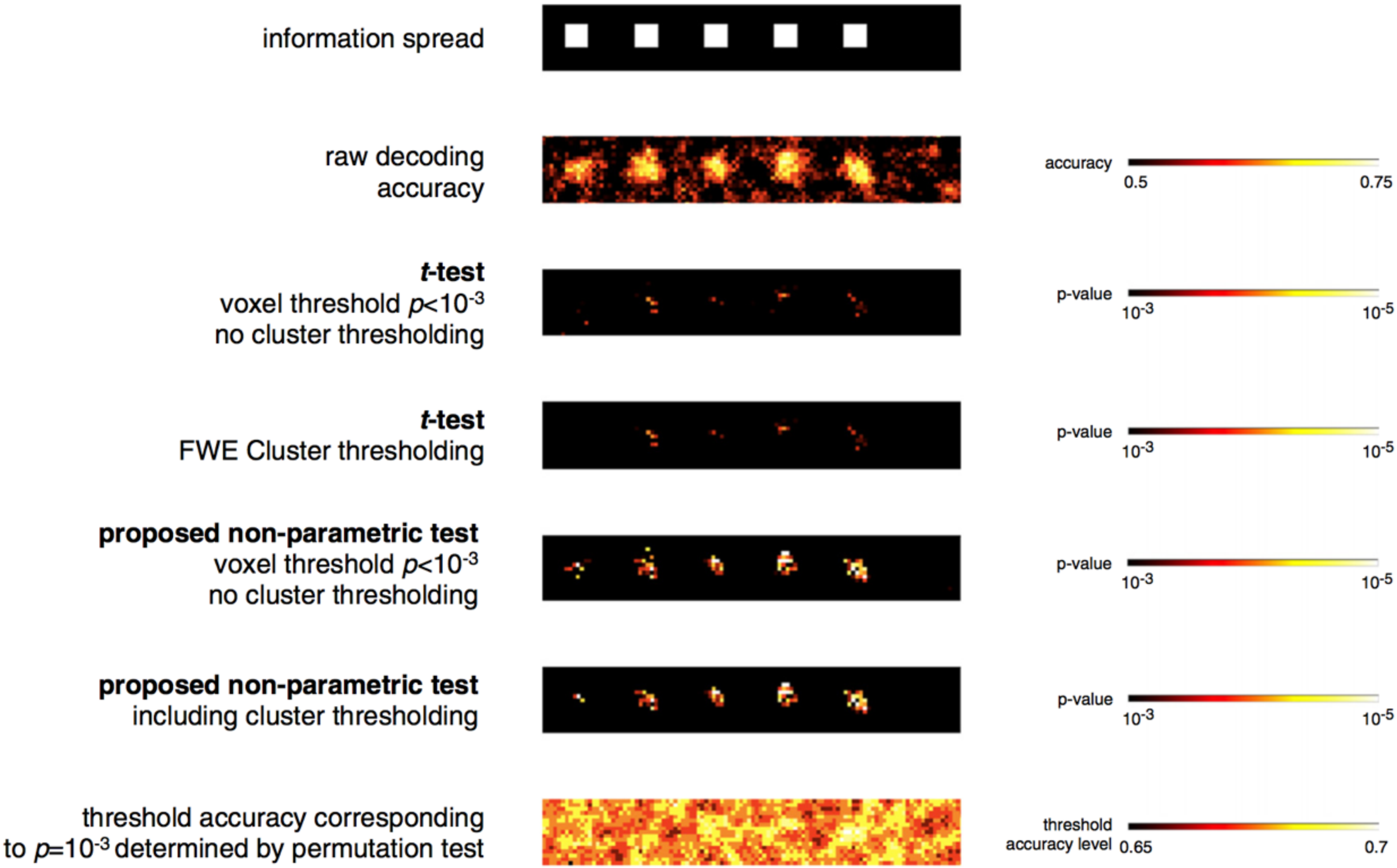
Hayasaka, S., and Nichols, T. E. (2003). Validating cluster size inference: random field and permutation methods. 20, 2343–2356. doi: 10.1016/j.neuroimage.2003.08.003.

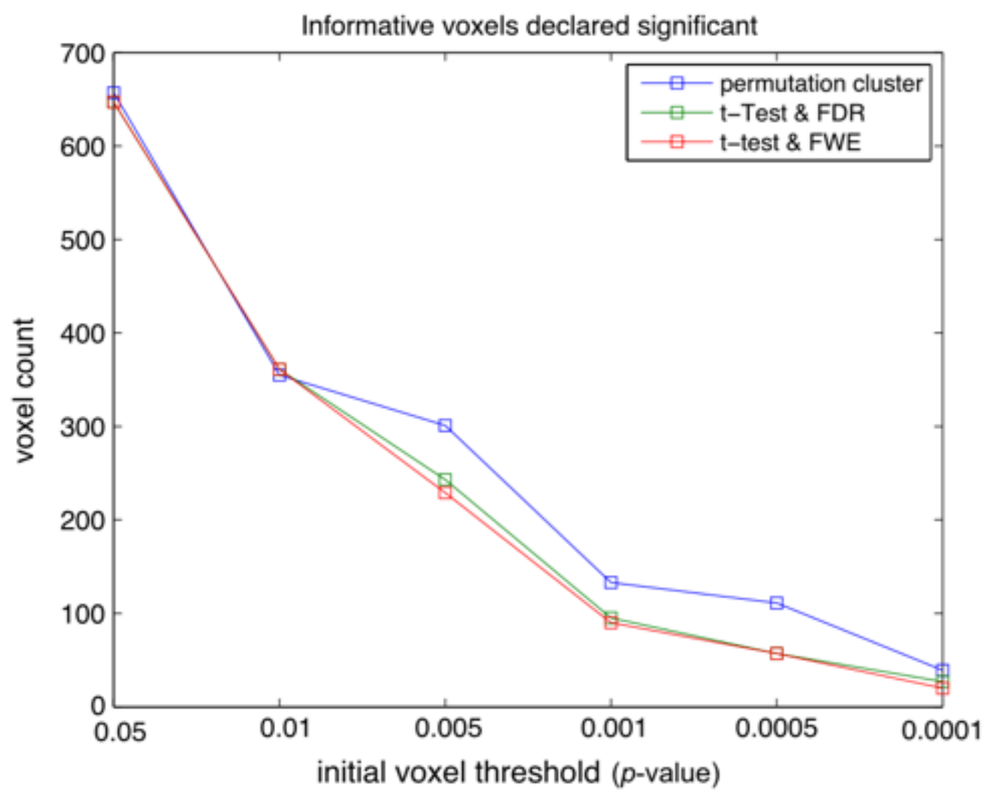


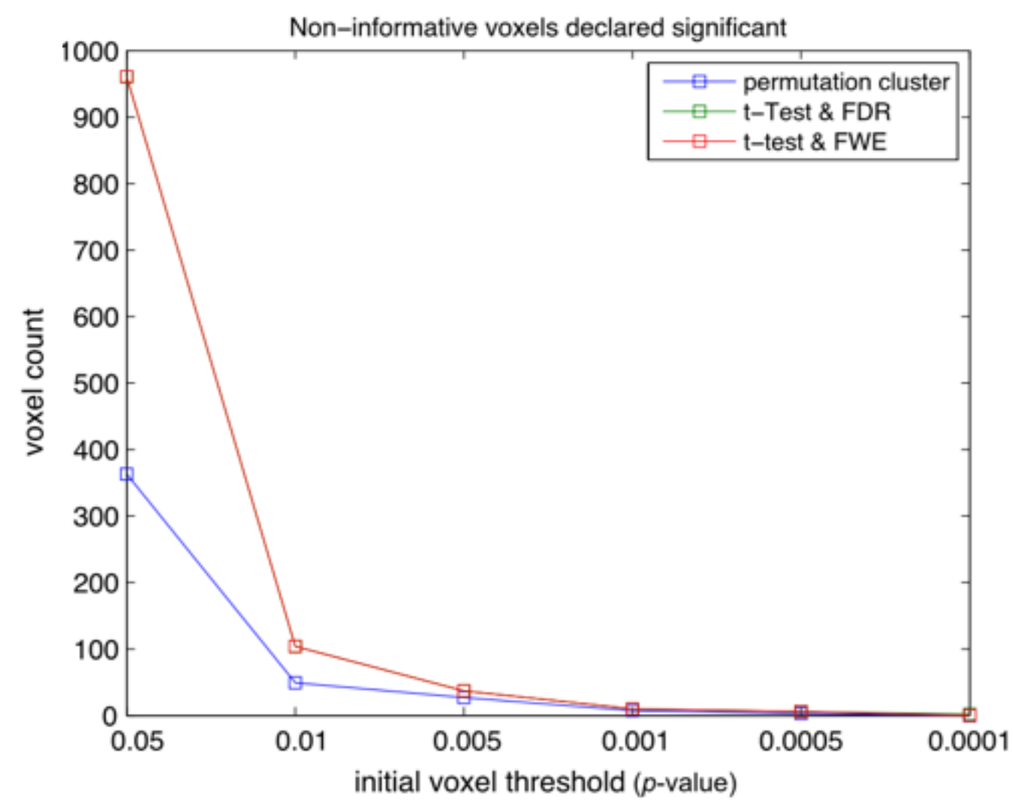
Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063



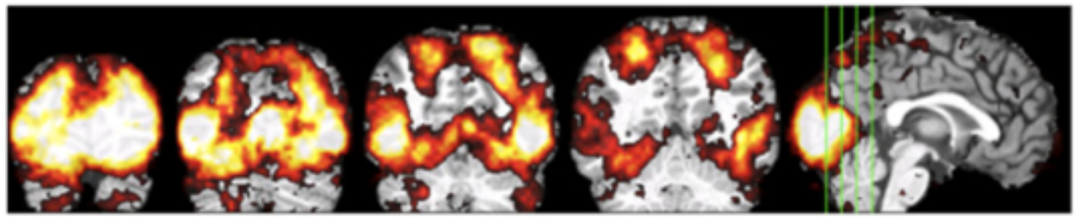
Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063


information spread
raw decoding accuracy
accuracy
0.5
0.75
t-test
voxel threshold p<10-3
no cluster thresholding
p-value
10-3
10-5
t-test
FWE Cluster thresholding
p-value
10-3
10-5
proposed non-parametric test
voxel threshold p<10-3
no cluster thresholding
p-value
10-3
10-5
proposed non-parametric test
including cluster thresholding
p-value
10-3
10-5
threshold accuracy corresponding
to p=10-3 determined by permutation test
threshold
accuracy level
0.65
0.7


Informative voxels declared significant
permutation cluster
t–Test & FDR
t–test & FWE
voxel count
700
600
500
400
300
200
100
0
0.05
0.01
0.005
0.001
0.0005
0.0001
initial voxel threshold (p-value)

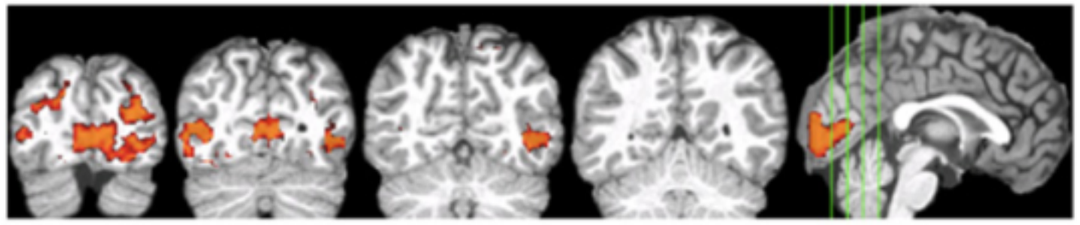

Non–informative voxels declared significant
permutation cluster
t–Test & FDR
t–test & FWE
voxel count
1000
900
800
700
600
500
400
300
200
100
0
0.05
0.01
0.005
0.001
0.0005
0.0001
initial voxel threshold (p-value)

raw decoding accuracy
no cluster size thresholding



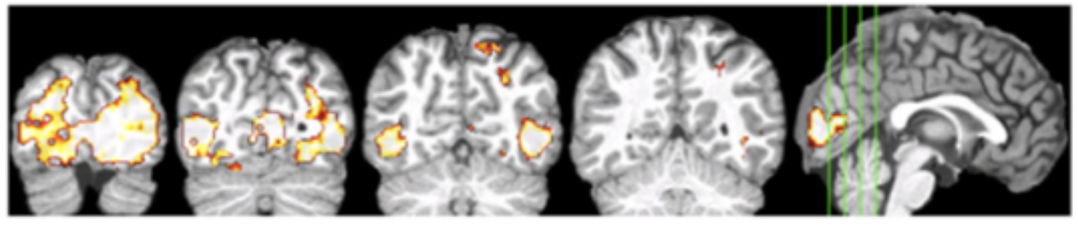
accuracy
0.5 0.75

***t*-test**
FWE cluster thresholding



p-value
10^{-3} 10^{-5}

proposed non-parametric test
including cluster thresholding



p-value
10^{-3} 10^{-5}

issues

- computational burden
- spatial smoothness homogeneous
- small-scale patterns invisible (cluster stats)
- group stat issues

Eklund, A., Dufort, P., Villani, M., and LaConte, S. (2014). BROCCOLI: Software for Fast fMRI Analysis on Many-Core CPUs and GPUs. Front Neuroinform 8. doi:10.3389/fninf.2014.00024.

Stelzer, J., Lohmann, G., Mueller, K., Buschmann, T., and Turner, R. (2014). Deficient approaches to human neuroimaging. Front. Hum. Neurosci. 8. doi:10.3389/fnhum.2014.00462.

thanks for your attention!

& thanks to

Robert Turner

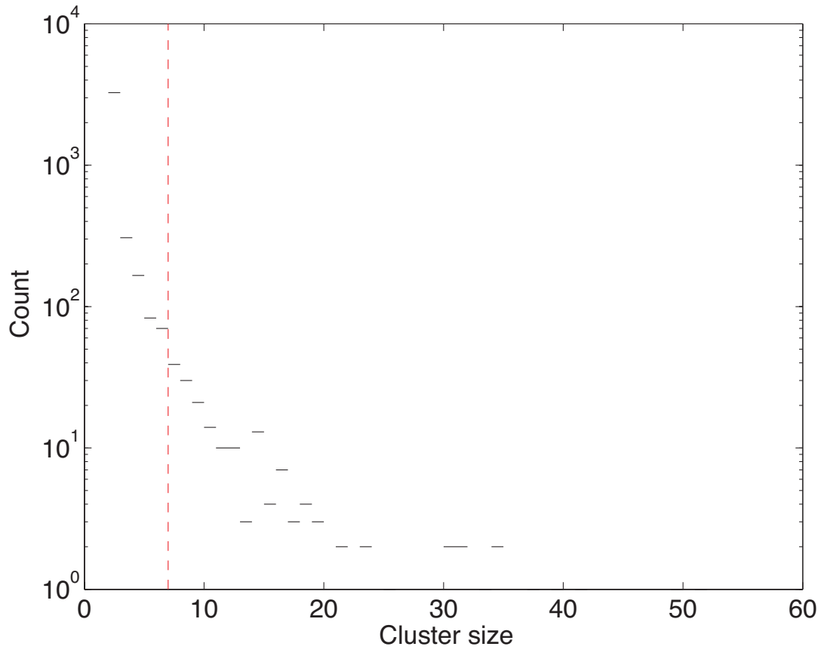
Yi Chen

Gabriele Lohmann

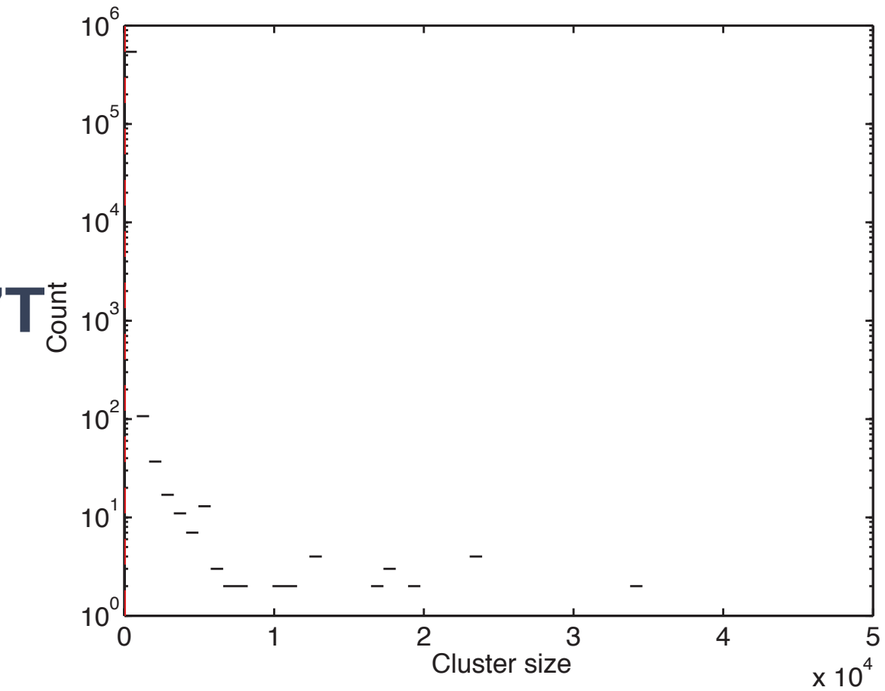
Tilo Buschmann

cluster histograms

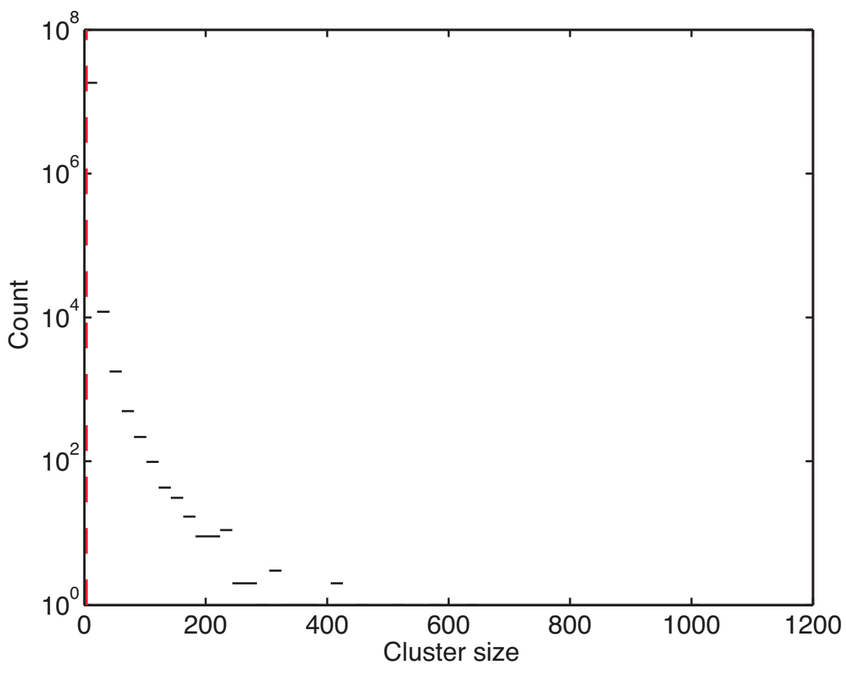
SLD 3T



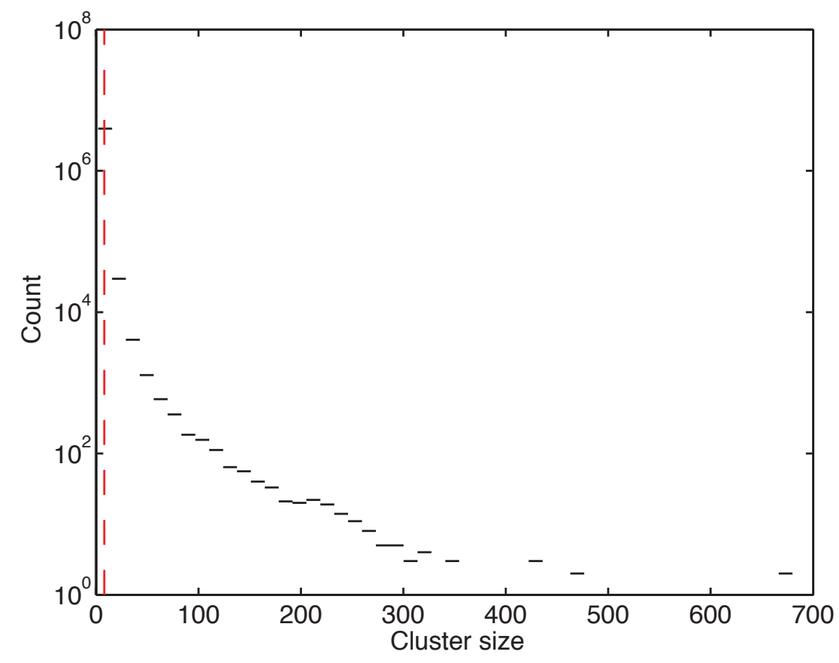
SLD 7T



FWM 3T



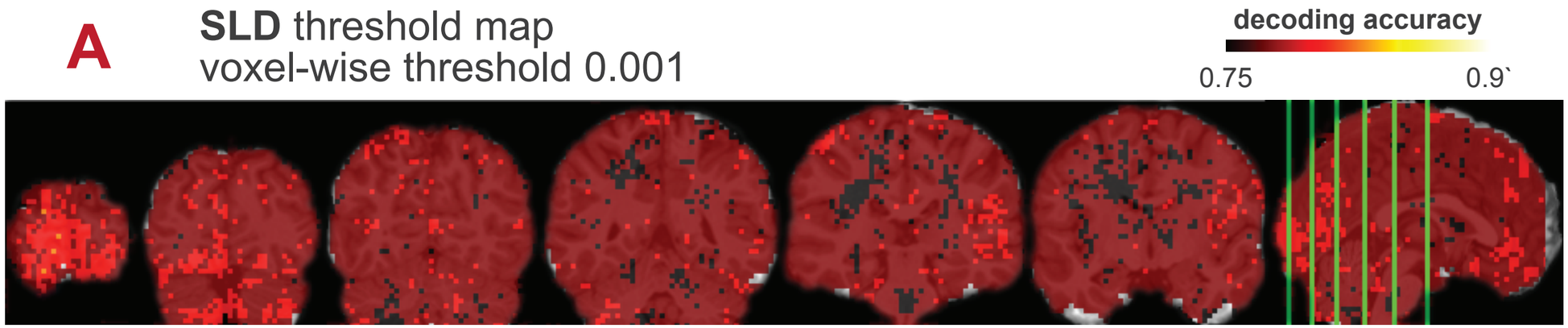
FWM 7T



threshold maps

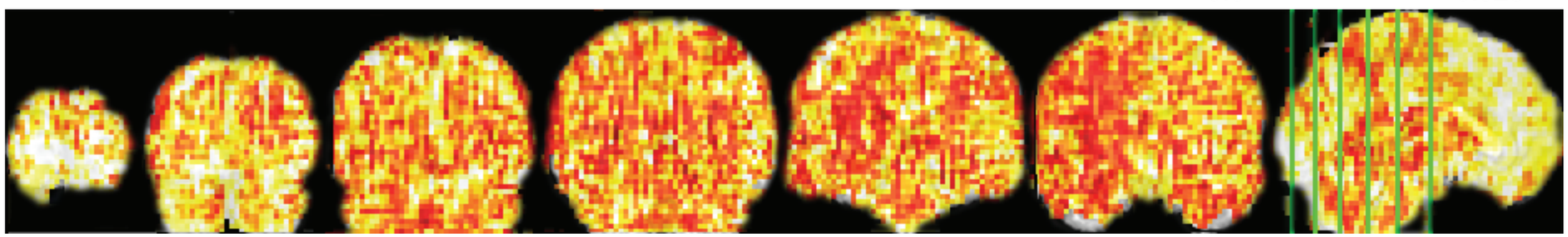
A

SLD threshold map
voxel-wise threshold 0.001



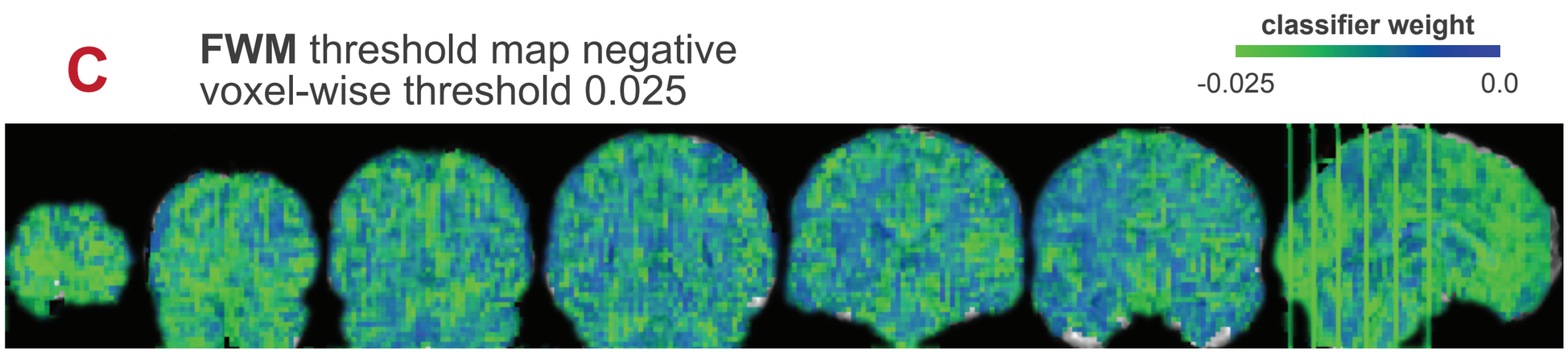
B

FWM threshold map positive
voxel-wise threshold 0.025

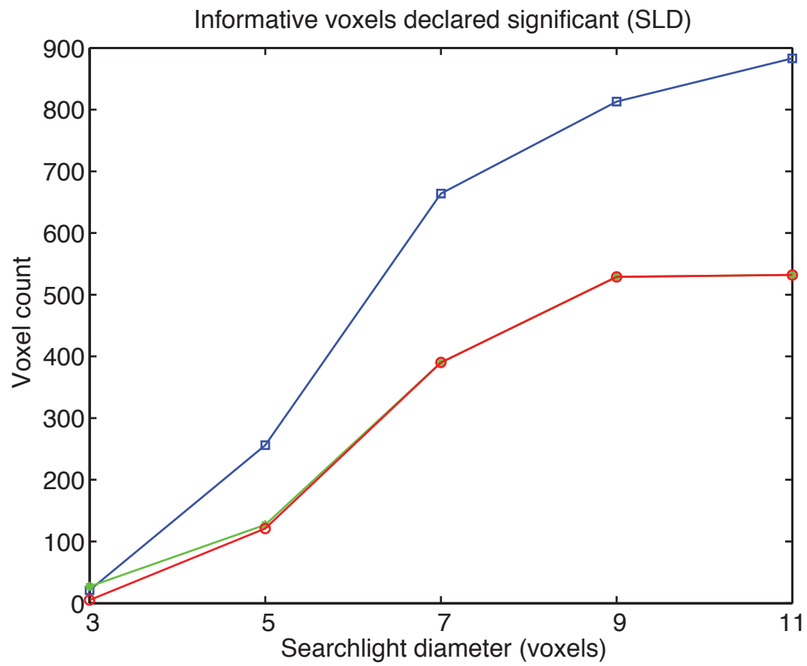


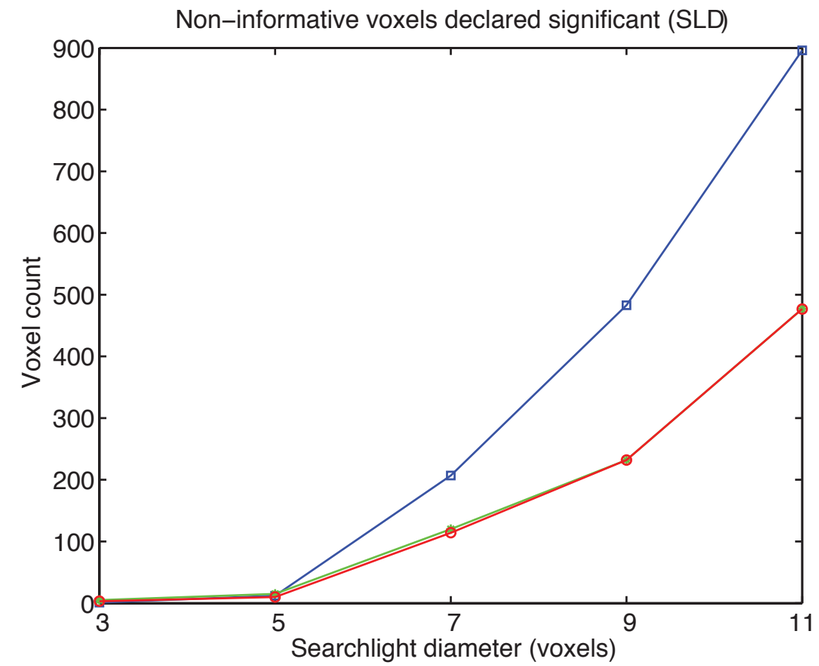
C

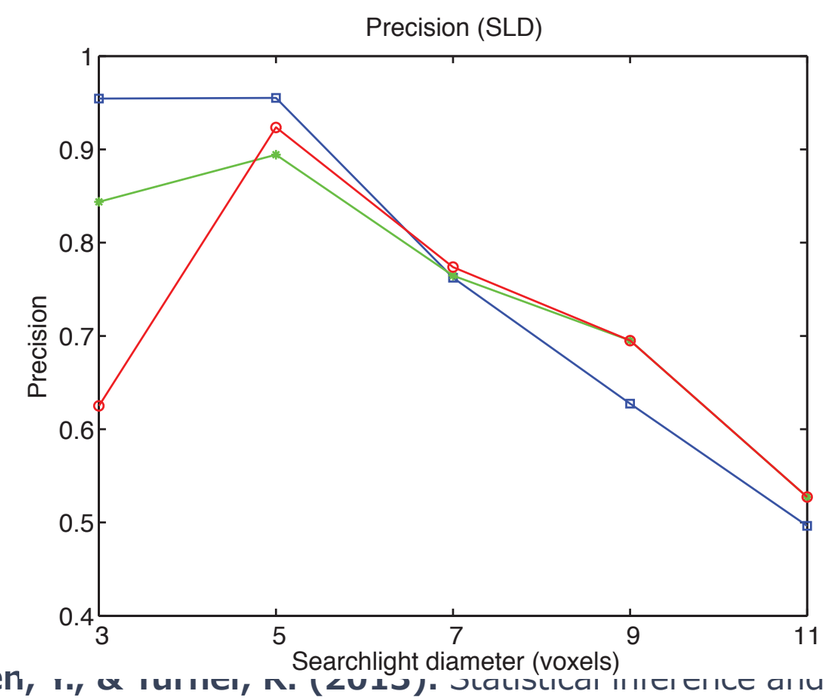
FWM threshold map negative
voxel-wise threshold 0.025

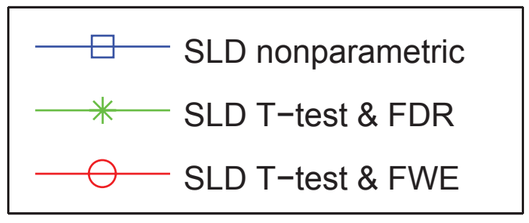


SLD diameter size









Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. Neuroimage, 65, 69–82. doi:10.1016/j.neuroimage.2012.09.063

Binomial model for accuracies

$$f_X(N_{corr}) = \binom{N_{test}}{N_{corr}} p^{N_{corr}} (1-p)^{N_{test}-N_{corr}} \qquad p = 0.5$$

k cross-validations

(~~assuming independency~~)

$$N = N_{test} \cdot k$$

$$f(N_{corr}) = \binom{N}{N_{corr}} p^{N_{corr}} (1-p)^{N-N_{corr}}$$

undersampling of permutation space