Introduction to, Examples and (towards a) Philosophy of Skeptical Neuroimaging

Victor Solo*,†

*School of Electrical Engineering
University of New South Wales
Sydney, AUSTRALIA

HBM Educational Workshop, June, 2013

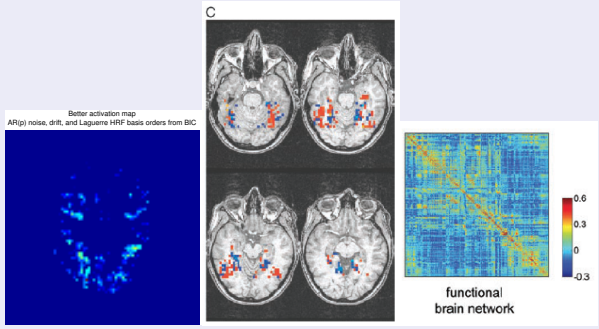
How Not to Analyze Your Data:
A Skeptical Introduction to Modeling Methods

†Work partly supported by an NIH P41 grant at the Martinos Center for Biomedical Imaging, Harvard Medical School.

Skeptical Neuroimaging: A Process of Constructive Criticism

Topics

1. Neuroimaging Data Analysis Paradigms:
   Algorithm Stream Approach.
   Statistical Modelling Approach.
2. Statistics:
   a Software Bag of Tricks?
   or a Scientific Method of Inference?
3. Macro Criticism (Long Term):
   Statistical Modelling Approach versus Algorithm Stream Approach.
4. Micro Criticism (Short Term):
   Skeptical Neuroimaging at Work.
   Activation Maps; Multivoxel Pattern Analysis; Network Analysis.
5. Conclusion.



FSL

FMRIB Software Library v5.0 September 2012

Created by the Analysis Group, FMRIB, Oxford, UK.

FREESURFER

SPM

By members & collaborators of the Wellcome Trust Centre for Neuroimaging

Introduction | Software | Documentation | Courses | Email list | Data | Extensions

The tendency in science to reinvent the wheel is especially prevalent in literatures that are inherently interdisciplinary and where experimental and statistical methods are borrowed intermittently across domains. One purpose

Statisticians have long bemoaned the inadequacies of status quo approaches to functional neuroimaging data analysis and have succeeded in publishing, but not popularizing, sensible alternatives. Pattern-based classifiers

Source: O'Toole etal, 2007.

Algorithm Stream Approach

Temporal dynamics of spontaneous MEG activity in brain networks

Francesco de Pasquale[a,b,1], Stefania Della Penna[a,b], Abraham Z. Snyder[c,d], Christopher Lewis[a,b], Dante Mantini[a,b,2], Laura Marzetti[a,b], Paolo Belardinelli[a,b], Luca Ciancetta[a,b], Vittorio Pizzella[a,b], Gian Luca Romani[a,b], and Maurizio Corbetta[a,b,c,d]

Materials and Methods

See *SI Text* for additional details.

Subjects, Procedures, and Acquisition. A total of 13 fMRI (four runs, 6 min each) and MEG (three runs, 5 min each) datasets were acquired in healthy young adult subjects; 10 subjects (mean age 29 ± 6 years, five females) contributed both MEG and fMRI datasets in separate sessions. BOLD time series were acquired using a 1.5 T Siemens Vision scanner (TR = 2.163 s; 3.75 × 3.75 mm in-plane resolution; slice thickness = 8 mm). Neuromagnetic signals were recorded with the MEG system developed at the University of Chieti (47) that includes 153 dc SQUID integrated magnetometers and coregistered to the fMRI data following a procedure described in *SI Text*.

fMRI Analysis. fMRI correlation maps were generated using the pipeline developed at Washington University (3, 11, 13).

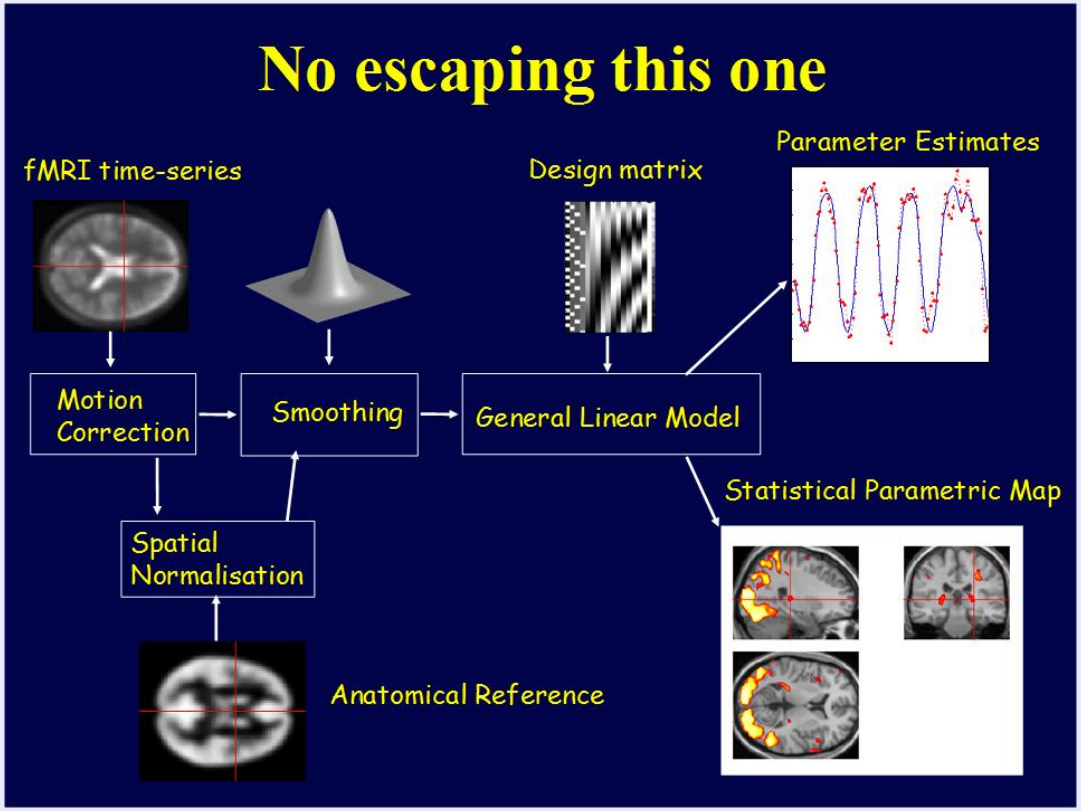
MEG Analysis. An extension of the Independent Component Analysis (ICA) algorithm described in ref. 48 was employed to automatically classify and remove artifactual MEG components. Artifact-free MEG signals were reconstituted from the remaining ICs, and source-space current was reconstructed by a weighted minimum-norm least squares (WMNLS) procedure implemented in Curry 6.0 (Neuroscan). This step yielded

$$\mathbf{q}_j(t) = [\, q_{jx}(t) \quad q_{jy}(t) \quad q_{jz}(t)\,]', \qquad [1]$$

the source-space current density vector at voxel j at time t. Power time series at voxel j was computed as

$$p_j(t) = (1/T_p) \int_t^{t+T_p} |\mathbf{q}_j(\tau)|^2 d\tau, \qquad [2]$$

where $T_p = 400$ms. Power time series were reconstructed from wide-band (1-150 Hz) MEG signals and on the basis of $\mathbf{q}_j(t)$ (see [1]) restricted to the theta (3.5–7 Hz), alpha (8-13 Hz), beta (14-25 Hz) and gamma (27-70 Hz) bands. Correlation time series between voxels j and s (the seed) were computed using the Pearson product moment formula. Thus,



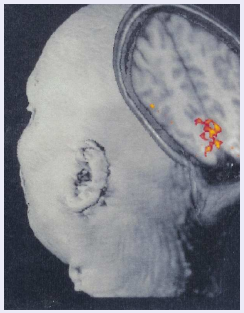
Source: Jesper Andersson, KI, Stockholm

This approach would seem a logical consequence of the scientific paradigm. i.e. here is a precise description of how the experiment was done as well as how the data was analysed. With care you can repeat the results. What else is necessary?

What's Statistics got to do with it?

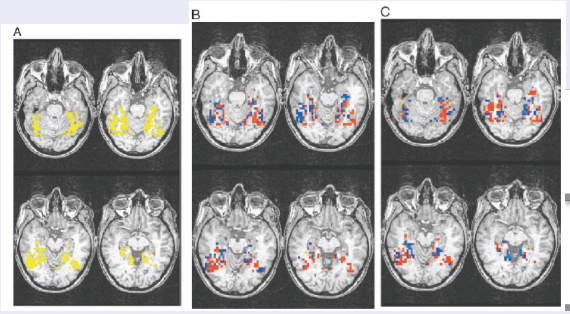
Activation Maps:

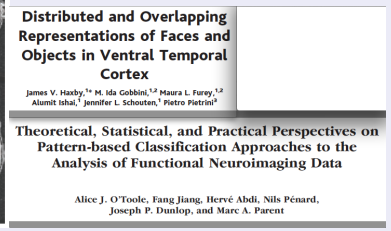
are Spatial Detection Statistics



Multivoxel Pattern Analysis

is Statistical Classification Analysis




Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex

James V. Haxby,[1*] M. Ida Gobbini,[1,2] Maura L. Furey,[1,2] Alumit Ishai,[1] Jennifer L. Schouten,[1] Pietro Pietrini[3]

Theoretical, Statistical, and Practical Perspectives on Pattern-based Classification Approaches to the Analysis of Functional Neuroimaging Data

Alice J. O'Toole, Fang Jiang, Hervé Abdi, Nils Pénard, Joseph P. Dunlop, and Marc A. Parent

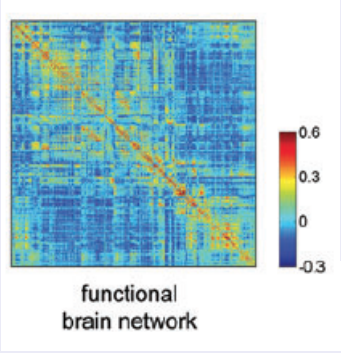
OHBM SEATTLE 2013

Since its inception in 1995 about 20% of HBM posters have been on (statistical) methodology.

Architectural Network Analysis: Activity Network Analysis:

uses correlation networks;

uses partial correlation networks?!




ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: The Year in Cognitive Neuroscience

The human connectome: a complex network

Olaf Sporns

Statistical Science: too complex to be Automated

Statistics

• Statistical inference is about the search for and validation of pattern in data. It is the original information extraction science.
• It's formalization, Statistical Decision Theory, can handle any conceivable statistical question.
Whatever your statistical circumstance it has probably already been raised.
• Like Science, Statistics has a History. Basic concepts [a] such as histograms (1840s,1892), correlation & regression (1880s), models (1922), maximum likelihood (1922), randomization (1922) **etc,etc,etc**. developed in and spread from Statistics.
• No matter how strong your quantitative background you can't rediscover/ understand all this in a short time (10,000 hours makes an expert!).

Incarnations

Biostatistics
Epidemiology
Statistical Signal Processing (Elec Eng)
System Identification (Engineering)
Econometrics
Pattern Recognition
(predates Computer Science!)
Assimilation (Oceanography, Meteorology)
Ill-conditioned Inverse Problems (Physics, Applied Math, Eng)
Photogrammetry (Surveying)
Psychometrics
Chemometrics
Interpretation (Geophysics)
Machine Learning (Computer Science)
Data Mining
etc etc etc!

[a] All discoveries have precursors.

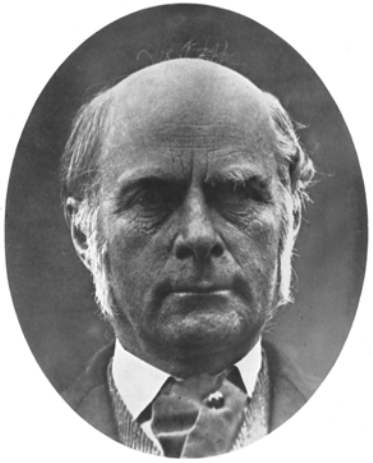
Creators of Mathematical Statistics 1880-1950

Fisher

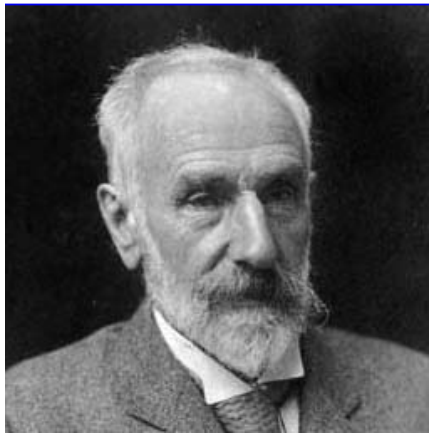
Wald

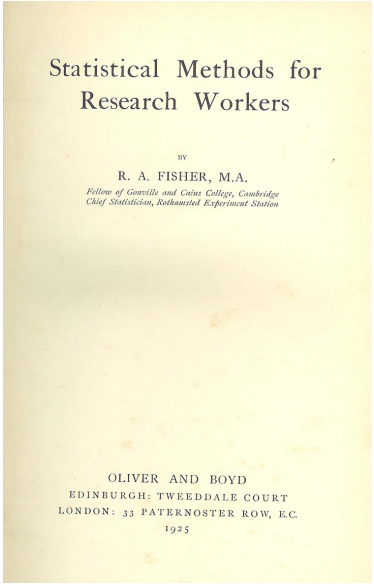
Galton

Sir Francis Galton F.R.S. 1822-1911



Edgeworth





Statistical Methods for Research Workers

BY

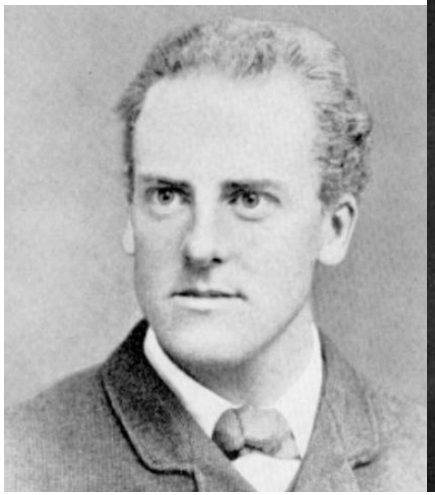
R. A. FISHER, M.A.

Fellow of Gonville and Caius College, Cambridge
Chief Statistician, Rothamsted Experiment Station

OLIVER AND BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 33 PATERNOSTER ROW, E.C.
1925

Statistical
Decision Functions

CHELSEA PUBLISHING COMPANY
BRONX, NEW YORK



Pearson

Yule

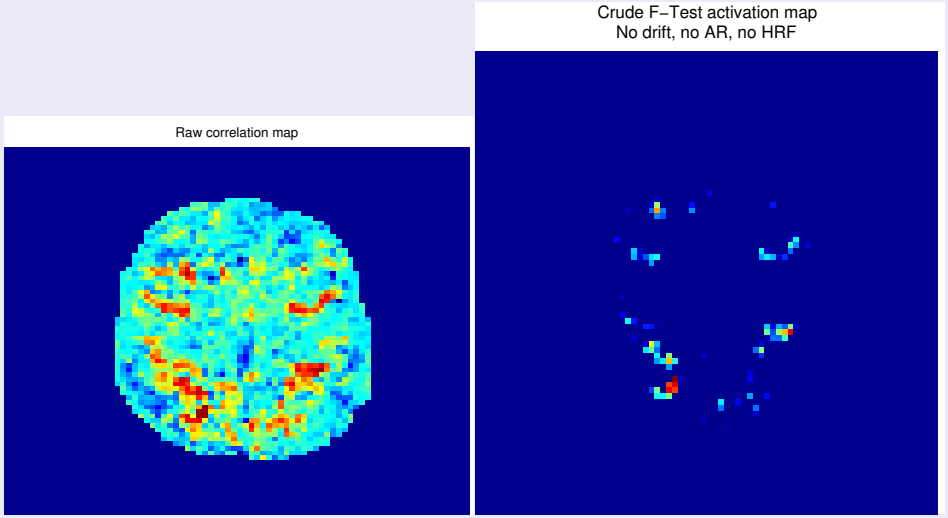
http://psychclassics.yorku.ca/Fisher/Methods/index.htm

An Example: Activation Maps

Viusal/Motor Experiment; 3T Philips Scanner.

Algorithm Stream

Use 'intuition' to 'create' a plot
e.g. correlate stimulus with BOLD.
or regress BOLD on stimulus.

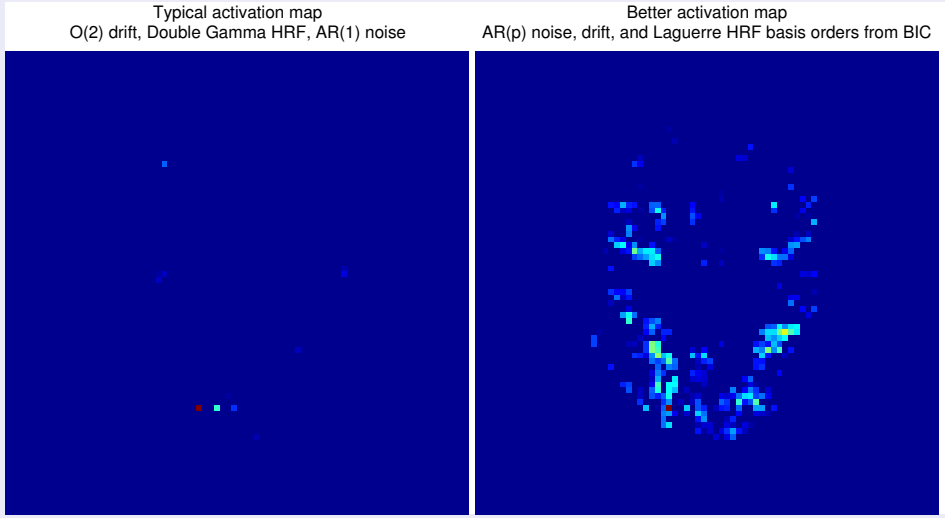


Identifying fMRI Model Violations
With Lagrange Multiplier Tests

Ben Cassidy*, *Student Member, IEEE*, Christopher J Long, Caroline Rae, and Victor Solo, *Fellow, IEEE*

Statistical-Empirical Model

Start with/develop a Model;
Derive a Test Statistic
y=drift+HR+noise
HR=activation×(HRF∗stimulus)



Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1

Geoffrey M. Boynton,[1] Stephen A. Engel,[1] Gary H. Glover,[2] and David J. Heeger[1]

Statistical Modelling versus Algorithm Stream

Statistical Modelling

Provides framework/methods for any conceivable data related question.

1. Experiment Design.
2. Construction of Models:
   black box $\leftrightarrow$ grey box $\leftrightarrow$ physical
3. Test & estimator construction
   e.g. maximum likelihood, likelihood ratio etc;[good null distributions].
4. Test & estimator (comparative) performance analysis e.g. bias, variance, power etc.
5. Theory e.g. Theorem: There is no best estimator and no best test!
6. Model criticism/statistical diagnostics (e.g. residuals plots) $\Rightarrow$ change model.
7. Assessing impact of preprocessing

Algorithm Stream

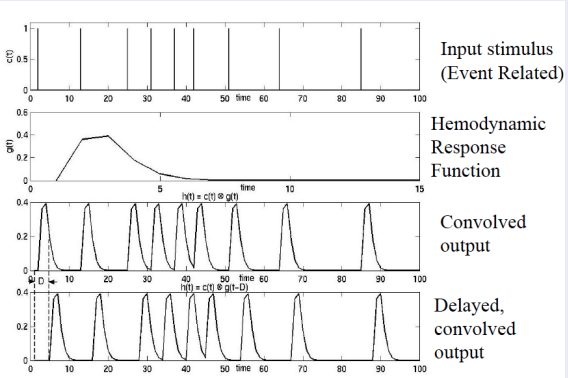
Provides:

1. Few experimental design guidelines.
2. No general methods.
   It's all adhoc/intuitive; whatever 'works';[bad null distributions].
3. No framework to measure performance or compare alternative methods; just endless simulation.
4. No theoretical analysis.
5. No notion of/framework for criticism/improvement.
6. Preprocessing is forgotten about once done & described.

Skeptical Neuroimaging at Work: Activation Maps

Statistical-Empirical (Grey-Box) Model

• Basic dynamic linear model (DLM not GLM) due to Boynton etal. (1996).



• Activation map is a spatially plotted & thresholded likelihood ratio (LR) test [a].

• But model must be specified e.g. double gamma, FIR, etc.

[a] But SPM gets this wrong

Model Criticism

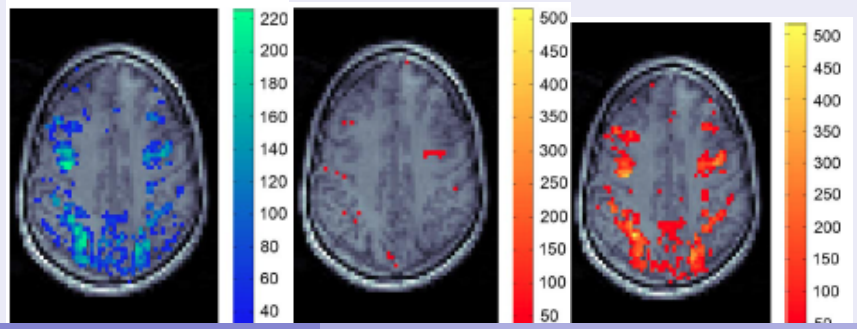
Two approaches:

• Exploratory based on residual plots & sensitivity analysis.

• Confirmatory based on hypothesis tests for model violations.

But confirmatory LR test is computationally demanding needing specialised software whereas LM (aka Score) test needs minor additional computation following e.g. an SPM fit: see Cassidy etal. (2012).

LM Double Gamma Laguerre



Skeptical Neuroimaging at Work: Multi-voxel Pattern Analysis

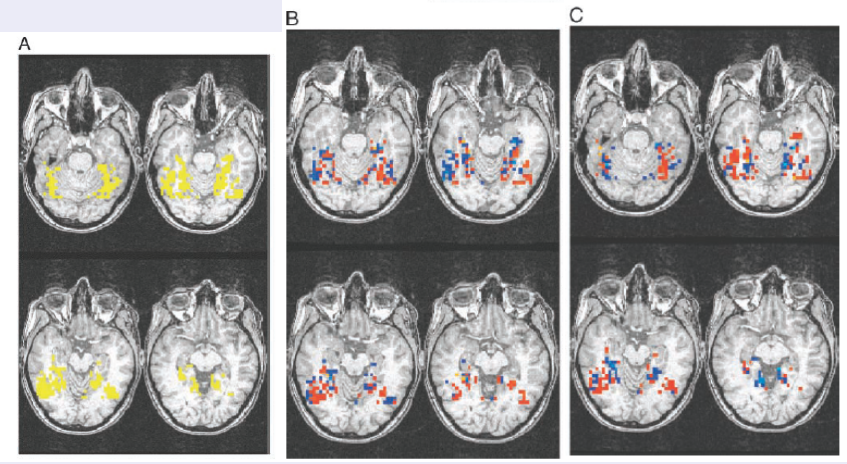
Theoretical, Statistical, and Practical Perspectives on Pattern-based Classification Approaches to the Analysis of Functional Neuroimaging Data

Alice J. O'Toole, Fang Jiang, Hervé Abdi, Nils Pénard, Joseph P. Dunlop, and Marc A. Parent

- MVPA uses statistical classification(/discriminant) analysis to distinguish between different stimuli that may activate the same cortical regions.
- But most such analysis is preceded by a region dimension reduction to a few time series typically by PCA.
- But: A classic statistical problem [Joliffe]. The dimension reduction ignores the stimuli.
- Indeed Haxby etal. found the discriminative PC to be the one explaining only 3% of the signal variation!

- Haxby etal. study of object recognition (faces, houses, etc).

A: Activation regions that discriminate between stimuli.
B: 50% PC that activates in the right places (fusiform, parahippocampal place areas) but cannot distinguish stimuli.
C: 3% PC - discriminates between stimuli!

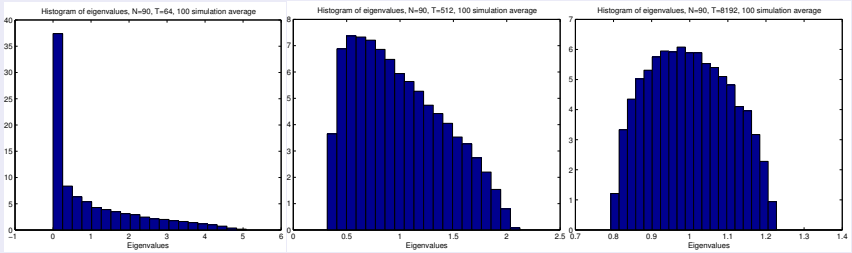


Other Criticisms: White Noise Methods are Problematic

Skeptical Neuroimaging at Work: Network Analysis

Architectural Analysis of Activity Networks

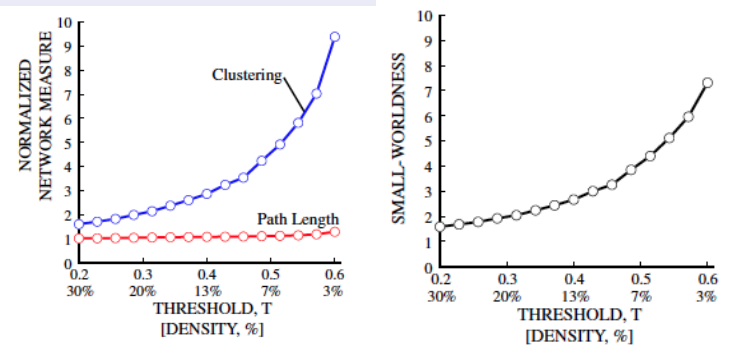
d = # nodes (parcels) = 90.
p = # parameters $= \frac{d(d+1)}{2} = 4050$.
n = # time points =64,128,512,8192.
d/n $\sim$ 1.5,.7,.35,.1.
= Random Matrix Theory (RMT) Regime.
The sample covariance/correlation matrix is very noisy but exhibits non-standard stable structure!
Eigenvalue histograms: n=64,512,8192

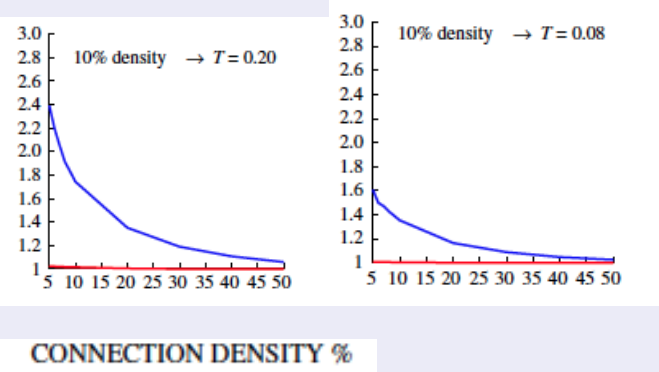


On the use of correlation as a measure of network connectivity

Andrew Zalesky [a,*], Alex Fornito [a,b,c], Ed Bullmore [d,e]

Thresholded white noise simulation exhibits small world network structure! Cluster coefficient & path length relative to randomly rewired network (n=10,64,512).



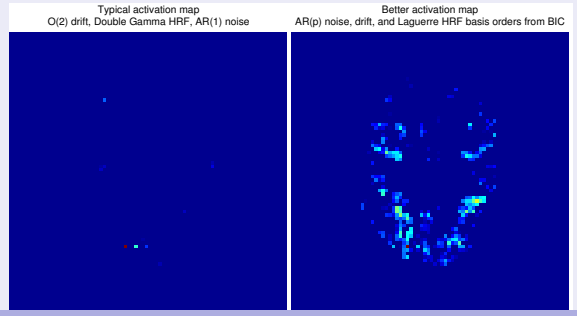


Conclusions

Skeptical Neuroimaging Credo:
Statistical Methodology and Neuro-science are Synergistic and so Equally Important.

1. Currently there are two paradigms of Neuroimaging Data Analysis:
   Statistical Modelling & Algorithm Stream.
   2 problems follow: Macro-scale & Micro-scale.
2. Macro: In the long term Statistical Modelling will dominate Algorithm Stream. But this needs to be managed in a manner suited to Neuroimaging.
3. Micro: In the short term we need critical tools for each of them.
4. Model criticism tools are already well developed for Statistical Modelling. But need further development for the Neuromaging setting.
5. Tools of criticism need to be urgently developed for Algorithm Stream. Candidates include: sensitivity analysis; residuals analysis; random matrix theory.

y=drift+HR+noise
HR=activation$\times$
(HRF$*$stimulus)



References

1. G.M. Boynton, S.E. Engel, G.A. Glover, D.J. Heeger (1996), Linear Systems Analysis of Functional Magnetic Resonance Imaging in Human V1, Jl. Neuroscience, 16, 4207-4221.
2. B Cassidy, C J Long, C Rae, and V Solo, Identifying fMRI Model Violations With Lagrange Multiplier Tests IEEE Trans. Medical Imaging, 31, 2012, 1481-1492.
3. Haxby, J. V. et al. (2001), Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425-2430.
4. I. T. Joliffe, 2002, Principal Component Analysis, 2nd edn, Springer, Berlin.
5. A J. OToole, F Jiang, H Abdi, N Penard, J P. Dunlop, and M A. Parent, (2007), Theoretical, Statistical, and Practical Perspectives on Pattern-based Classification Approaches to the Analysis of Functional Neuroimaging Data, Journal of Cognitive Neuroscience ,19, 1735-1752.
6. F de Pasquale, S Della Penna, A Z. Snyder, C Lewis, D Mantini, L Marzetti, P Belardinelli, L Ciancetta, V Pizzella, G L Romani, and M Corbett, (2010), Temporal dynamics of spontaneous MEG activity in brain networks, PNAS, 107, 6040-6045.
7. O. Sporns The human connectome: a complex network, Ann. N.Y. Acad. Sci. 1224 (2011) 109125.
8. A Zalesky , A Fornito , E Bullmore, (2012), On the use of correlation as a measure of network connectivity, NeuroImage 60, 2096-2106.

Postscript: A Challenge

Temporal dynamics of spontaneous MEG activity in brain networks

Francesco de Pasquale[a,b,1], Stefania Della Penna[a,b], Abraham Z. Snyder[c,d], Christopher Lewis[a,b], Dante Mantini[a,b,2], Laura Marzetti[a,b], Paolo Belardinelli[a,b], Luca Ciancetta[a,b], Vittorio Pizzella[a,b], Gian Luca Romani[a,b], and Maurizio Corbetta[a,b,c,d]

Materials and Methods

See *SI Text* for additional details.

Subjects, Procedures, and Acquisition. A total of 13 fMRI (four runs, 6 min each) and MEG (three runs, 5 min each) datasets were acquired in healthy young adult subjects; 10 subjects (mean age 29 ± 6 years, five females) contributed both MEG and fMRI datasets in separate sessions. BOLD time series were acquired using a 1.5 T Siemens Vision scanner (TR = 2.163 s; 3.75 × 3.75 mm in-plane resolution; slice thickness = 8 mm). Neuromagnetic signals were recorded with the MEG system developed at the University of Chieti (47) that includes 153 dc SQUID integrated magnetometers and coregistered to the fMRI data following a procedure described in *SI Text*.

fMRI Analysis. fMRI correlation maps were generated using the pipeline developed at Washington University (3, 11, 13).

MEG Analysis. An extension of the Independent Component Analysis (ICA) algorithm described in ref. 48 was employed to automatically classify and remove artifactual MEG components. Artifact-free MEG signals were reconstituted from the remaining ICs, and source-space current was reconstructed by a weighted minimum-norm least squares (WMNLS) procedure implemented in Curry 6.0 (Neuroscan). This step yielded

$$\mathbf{q}_j(t) = [\, q_{jx}(t) \quad q_{jy}(t) \quad q_{jz}(t) \,]', \qquad [1]$$

the source-space current density vector at voxel j at time t. Power time series at voxel j was computed as

$$p_j(t) = (1/T_p) \int_t^{t+T_p} |\mathbf{q}_j(\tau)|^2 d\tau, \qquad [2]$$

where $T_p = 400$ms. Power time series were reconstructed from wide-band (1-150 Hz) MEG signals and on the basis of $\mathbf{q}_j(t)$ (see [1]) restricted to the theta (3.5–7 Hz), alpha (8-13 Hz), beta (14-25 Hz) and gamma (27-70 Hz) bands. Correlation time series between voxels j and s (the seed) were computed using the Pearson product moment formula. Thus,

Algorithm Stream Criticism

Find 6 points of contention in the methods opposite.

3 from the data collection.

3 from the data analysis.

Suggest how these issues could be investigated in a post data analysis.