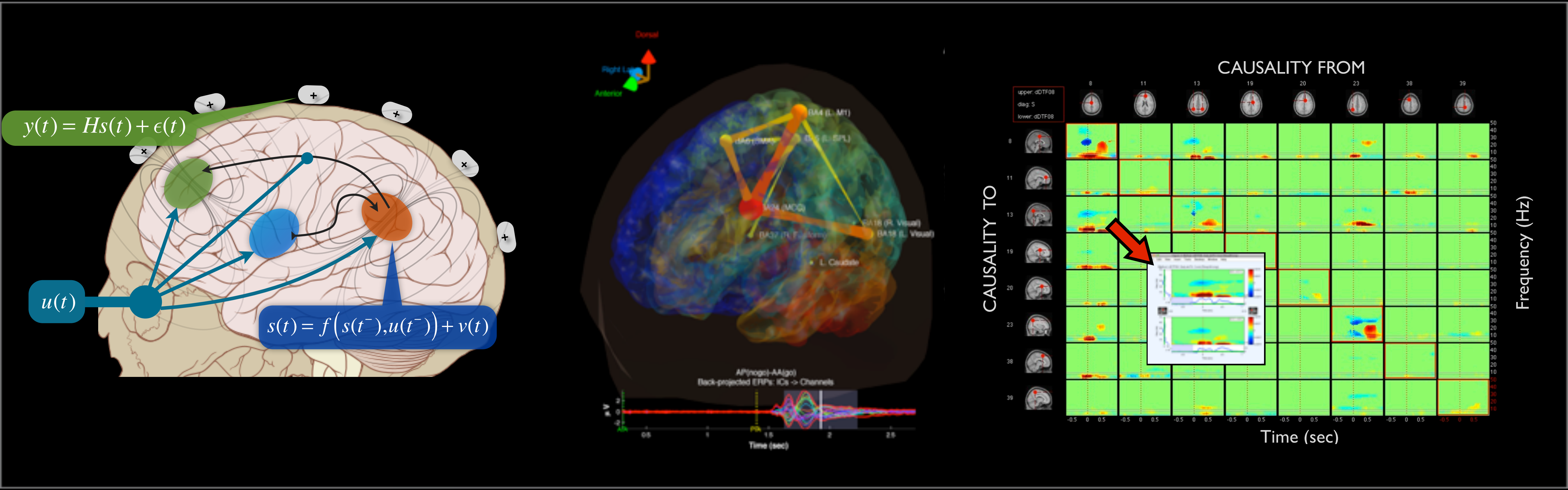
The Dynamic Brain II: Modeling Neural Dynamics and Interactions from M/EEG



Tim Mullen, PhD

Qusp Labs

Swartz Center for Computational Neuroscience
Institute for Neural Computation
UC San Diego

Intro
Theory
SIFT
Apps
To-Do
Fin

Review

Theoretical Foundations I

- Functional Connectivity Measures (PLV, PAC, Coherence)
- Effective Connectivity Measures and Granger Causality
- Scalp versus Source
- Adapting to Time-Varying Dynamics

Practicum: Hands-On Walkthrough of SIFT

Outline

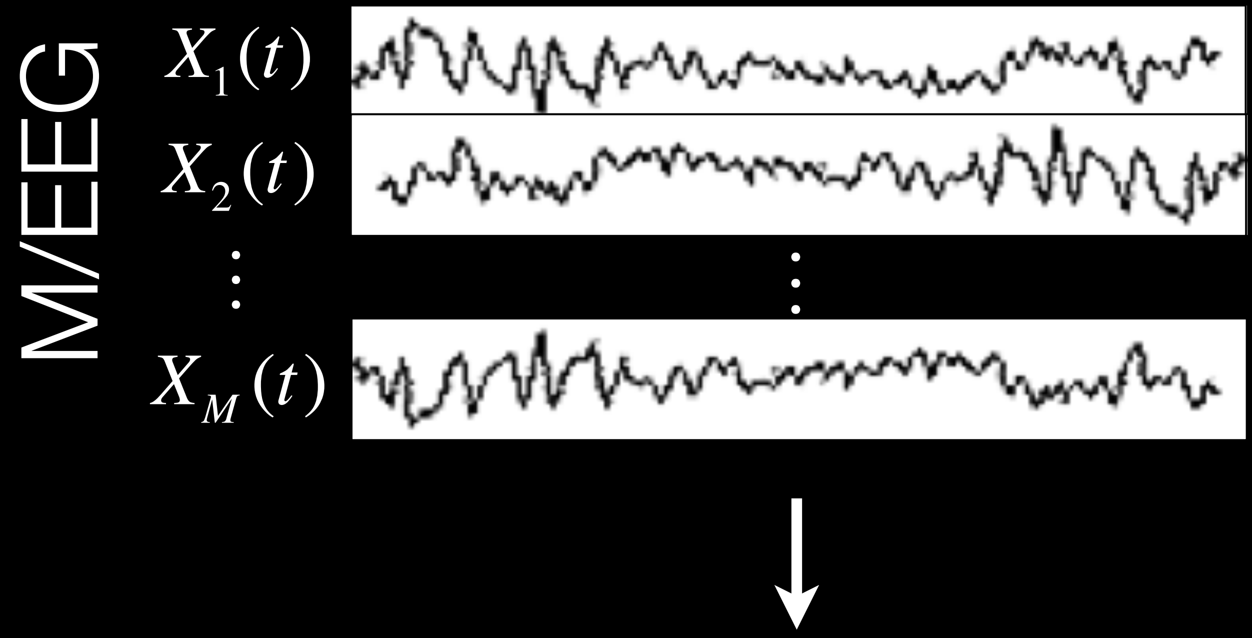
Theoretical Foundations II

- Model Validation
- Multivariate vs. Bivariate
- Imposing Constraints
- Single-trial Estimation and State-Space Models
- Statistical Testing

Practicum: Hands-On Simulation-based training

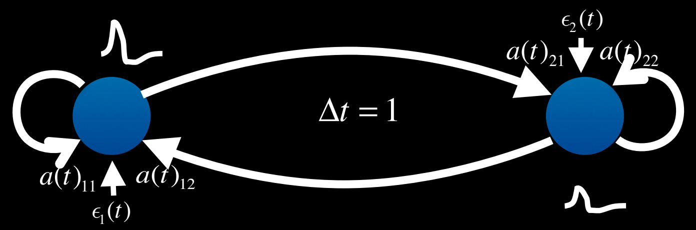
Review: The VAR model

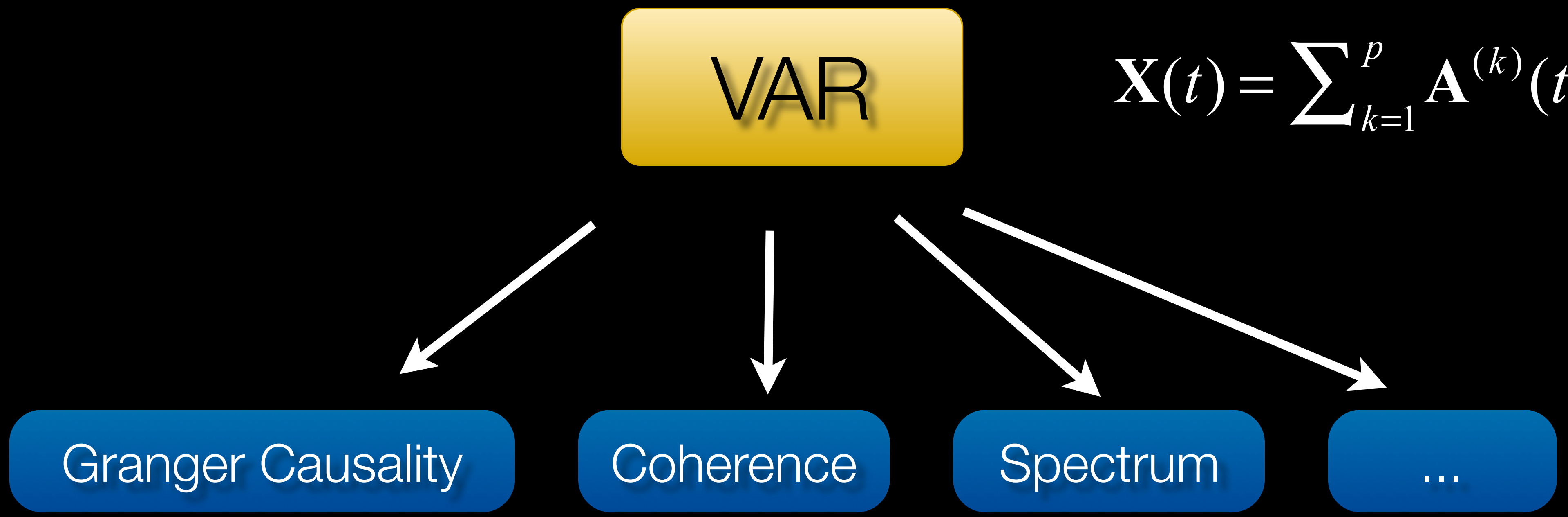
- The Vector Autoregressive (VAR) model as a basis for dynamical estimation



Stochastic Linear Dynamical System

$$X_1(t) = a(t)_{11} X_1(t-1) + a(t)_{12} X_2(t-1) + \epsilon_1(t)$$
$$X_2(t) = a(t)_{22} X_2(t-1) + a(t)_{21} X_1(t-1) + \epsilon_2(t)$$





VAR

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

Granger Causality

Coherence

Spectrum

...

Model Validation

- If a model is poorly fit to data, then few, if any, inferences can be validly drawn from the model.
- There a number of criteria which we can use to determine whether we have appropriately fit our VAR model. Here are three commonly used categories of tests:
- **Whiteness Tests:** checking the residuals of the model for serial and cross-correlation
- **Consistency Test:** testing whether the model generates data with same correlation structure as the real data
- **Stability Test:** checking the stability/stationarity of the model.

Tim Mullen

Whiteness Tests

- We can regard the VAR[p] model coefficients $\mathbf{A}^{(k)}$ as a filter which transforms random (white) noise $\mathbf{E}(t)$, into observed, structured data $\mathbf{X}(t)$:

$$\mathbf{X}(t) = f(L)\mathbf{E}(t)\ , \qquad f(L) = \left(I - \sum_{k=1}^{p} A^{(k)} L^{k}\right)^{-1}, \qquad L^{k} Z(t) = Z(t-k)$$

L is a "lag operator"

- For coefficient estimates $\hat{\mathbf{A}}^{(k)}$, we can obtain the residuals

$$\hat{\mathbf{E}}(t) = \mathbf{X}(t) - \sum_{k=1}^{p} \hat{\mathbf{A}}^{(k)}(t)\mathbf{X}(t-k)$$

- If we have adequately modeled the data, the residuals should be indistinguishable from a white noise process. Correlation structure in the residuals means there is still correlation structure in the data that has not been explained by the model.

- Checking the whiteness of residuals typically involves testing whether the residual **auto-** and **cross-correlation** coefficients up to some desired lag *h* are sufficiently small to ensure that we cannot reject the null hypothesis of white residuals at some desired significance level.

Tim Mullen

Whiteness Tests

$$\mathbf{E}(t) = N(0, \mathbf{V})$$

$$C_l = \left\langle \hat{\mathbf{E}}(t)\hat{\mathbf{E}}'(t-l) \right\rangle$$ autocovariance at lag *l* ...

$$R_l = D^{-1} C_l D^{-1}$$ with correponding autocorrelation *R*

$$D = diag\left(\sqrt{diag(C_0)}\right)$$

$$R_h = (R_1, \ldots, R_h)$$ set of autocorrelations up to lag *h*

We want to test the null hypothesis $H_0 : \mathbf{R}_h = (R_1, \ldots, R_h) = 0$

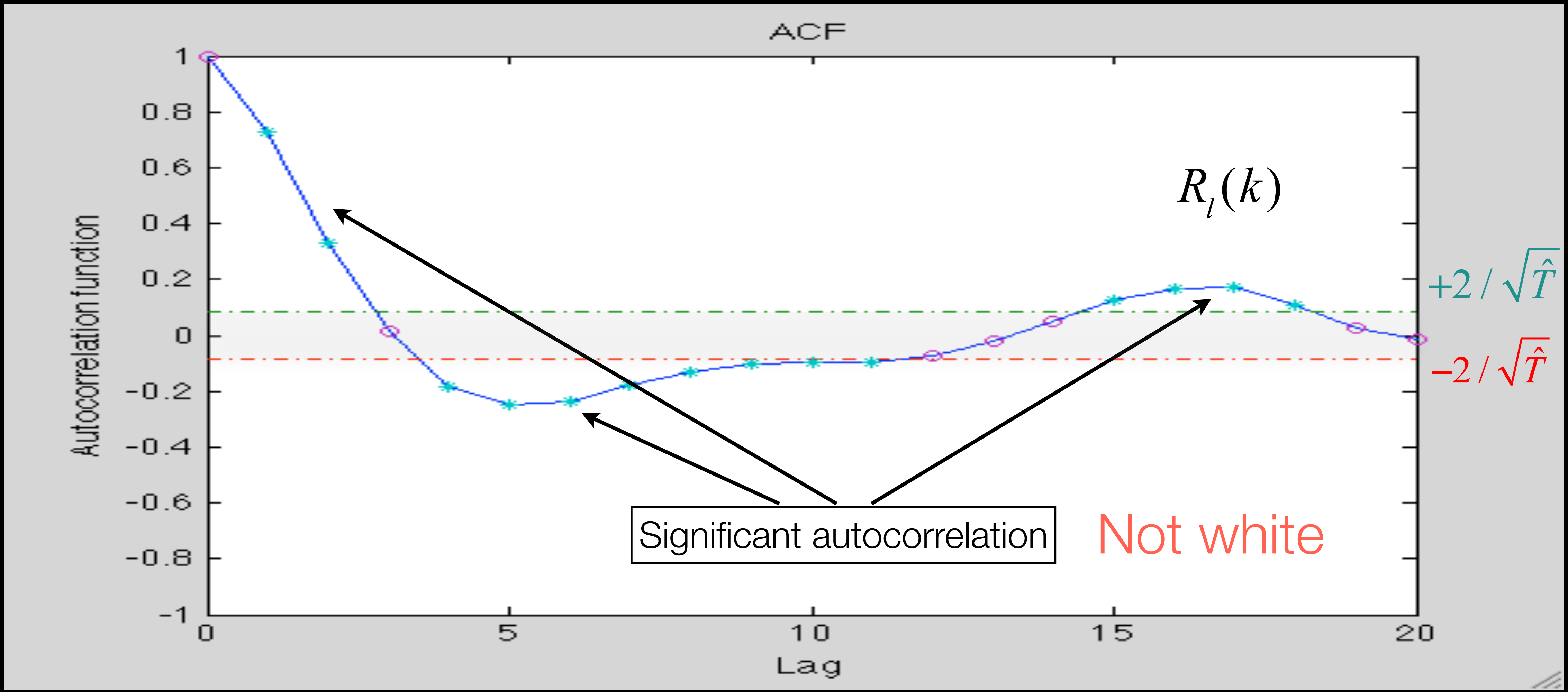
against the alternative: $H_1 : \mathbf{R}_h \neq 0$

Two possible ways to do this:

- Autocorrelation function test
- Portmanteau tests

Tim Mullen

Whiteness Tests: ACF



Significant autocorrelation

Not white

Under the null hypothesis that **Ê**(*t*) is Gaussian white noise, we expect approximately 1/20=5% of a.c.f. coefficients to exceed the threshold $\pm 2/\sqrt{\hat{T}}$ This gives us a p-value for rejecting H_0

$$r = \frac{\text{count}\left(\left|\mathbf{R}_h\right| > 2/\sqrt{\hat{T}}\right)}{\text{count}\left(\mathbf{R}_h\right)} = \frac{\text{count}\left(\left|\mathbf{R}_h\right| > 2/\sqrt{\hat{T}}\right)}{M^2(h+1) - M}$$

If $p<0.05$ ($1-p > .95$) then we cannot reject H_0 at the 5% level and we accept that residuals **Ê**(*t*) are white

Whiteness Tests: ACF

- Problem:
  - Confidence intervals apply to individual coefficients and assume coefficients are asymptotically independent. This may not be the case for multivariate models.
  - In small sample conditions (small *T*), this test may cause us to reject the null hypothesis (residuals indicated as non-white) **less often** than we should for the chosen significance level (Lutkepohl, 2006) -- in other words, we may have a **higher false positive rate** for accepting that the model fits the data.
- This motivates the use of alternate multivariate tests

Tim Mullen

Whiteness Tests: Portmanteau

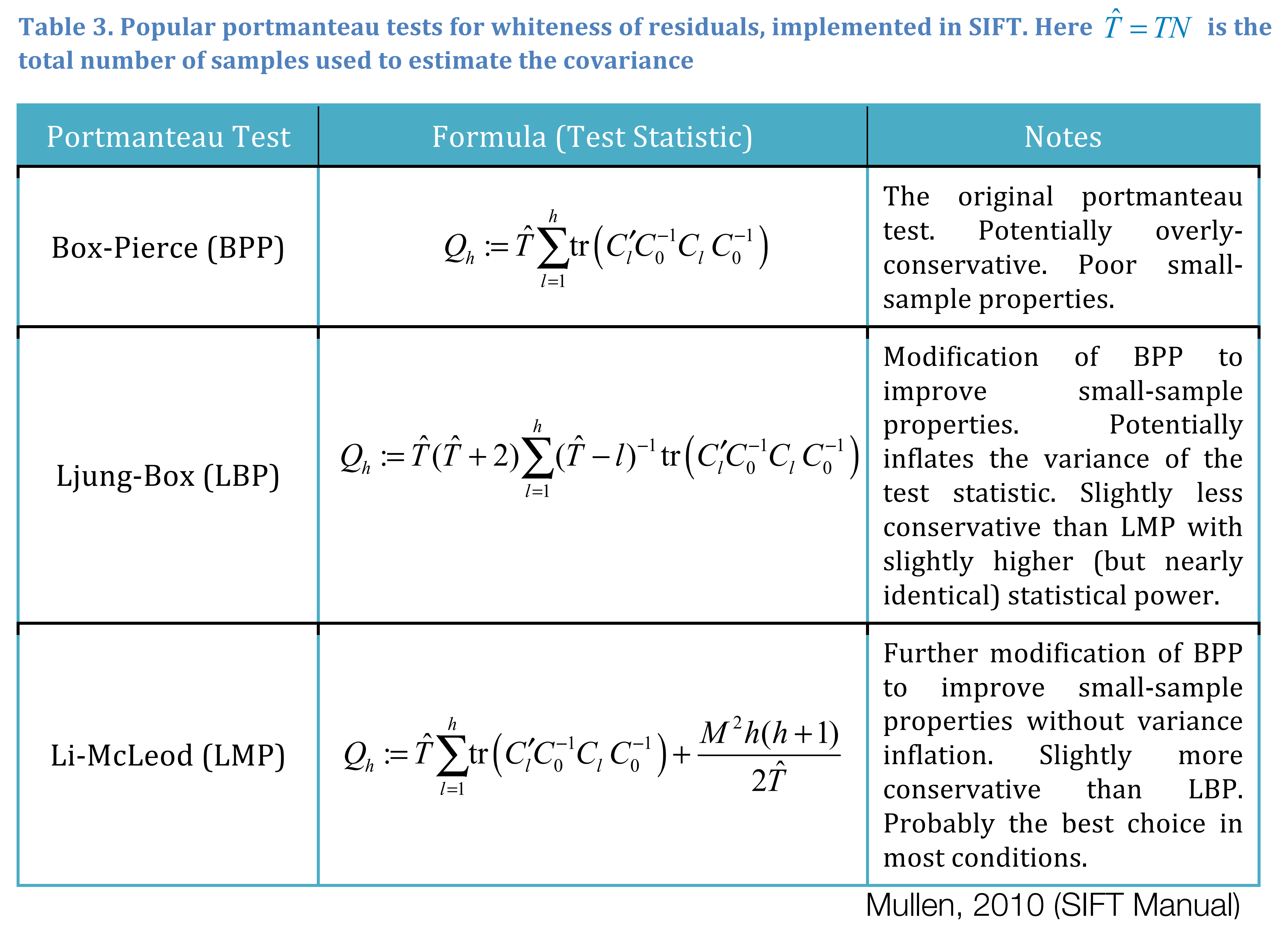


Table 3. Popular portmanteau tests for whiteness of residuals, implemented in SIFT. Here $\hat{T} = TN$ is the total number of samples used to estimate the covariance

Portmanteau Test	Formula (Test Statistic)	Notes
Box-Pierce (BPP)	$Q_h := \hat{T}\sum_{l=1}^{h}\mathrm{tr}\left(C_l' C_0^{-1} C_l C_0^{-1}\right)$	The original portmanteau test. Potentially overly-conservative. Poor small-sample properties.
Ljung-Box (LBP)	$Q_h := \hat{T}(\hat{T}+2)\sum_{l=1}^{h}(\hat{T}-l)^{-1}\mathrm{tr}\left(C_l' C_0^{-1} C_l C_0^{-1}\right)$	Modification of BPP to improve small-sample properties. Potentially inflates the variance of the test statistic. Slightly less conservative than LMP with slightly higher (but nearly identical) statistical power.
Li-McLeod (LMP)	$Q_h := \hat{T}\sum_{l=1}^{h}\mathrm{tr}\left(C_l' C_0^{-1} C_l C_0^{-1}\right) + \frac{M^2 h(h+1)}{2\hat{T}}$	Further modification of BPP to improve small-sample properties without variance inflation. Slightly more conservative than LBP. Probably the best choice in most conditions.

Mullen, 2010 (SIFT Manual)

These test statistics are asymptotically χ^2 distributed with $M^2(h-p)$ d.f.

Tim Mullen

Consistency Tests

- A well-fit model should be able to generate data that has the same correlation structure as the original data.
- One test of this is *percent consistency* (Ding et al, 2000)
- Here we generate simulated data from our fitted model (feeding it white noise) and calculate auto- and cross-correlations up to a fixed lag for both simulated data ($\mathbf{R}_s$) and real data ($\mathbf{R}_r$).
- The percent consistency (PC) is then given by

$$\mathrm{PC} = \left(1 - \frac{\left\|\mathbf{R}_s - \mathbf{R}_r\right\|_2}{\left\|\mathbf{R}_r\right\|_2}\right) \times 100$$

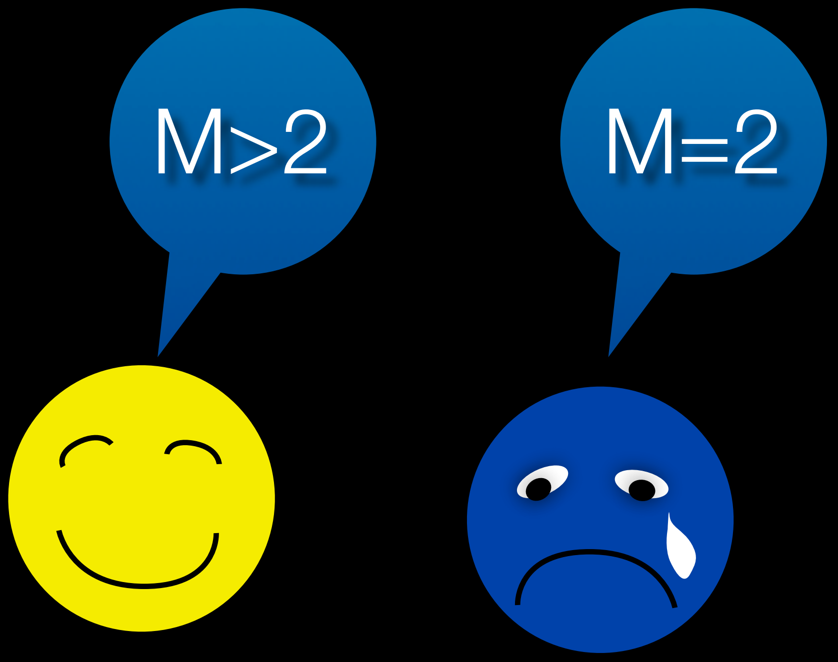
- A PC value near 100% indicates that the model is able to generate data that has a nearly identical correlation structure as the original data. A PC value near 0% indicates a complete failure to model the data.

Stability Tests

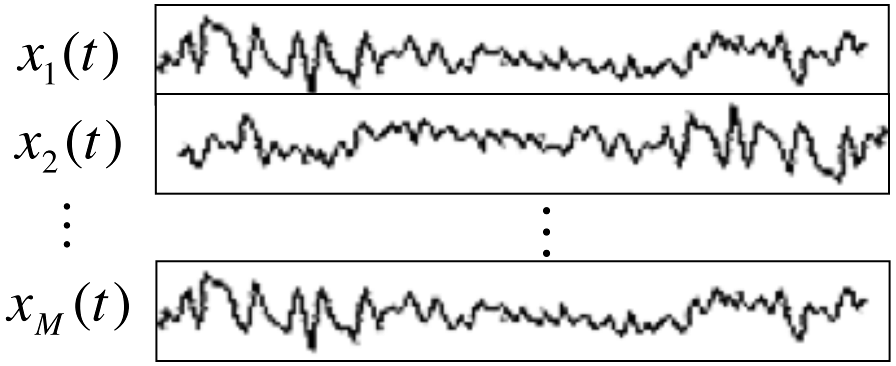
- **Stability**
  - All eigenvalues of the system matrix are ≤1
  - A stable process will not “blow up” (diverge to infinity)
  - A stable model is always a stationary model (however, the converse is not necessarily true). If a stable model adequately fits the data (white residuals), then the data is likewise stationary

Multivariate versus Bivariate

- Exclusion of processes that may exert causal influence on modeled processes increases the risk of causal mis-identification. (c.f. Pearl, *Causality: Models, Inference and Reasoning,* 2009)
  - Multivariate approaches are generally superior to bivariate approaches
    - allow detection of direct versus indirect dependence, reducing false positives
    - allow us to partially control for exogenous/unobserved causes (e.g. Guo, et al., J. Neuro. Methods, 2008)



- In the absence of *a priori* knowledge concerning causal structure, it is advisable to include as many processes as possible in a causal model *(within data/modeling limitations)*

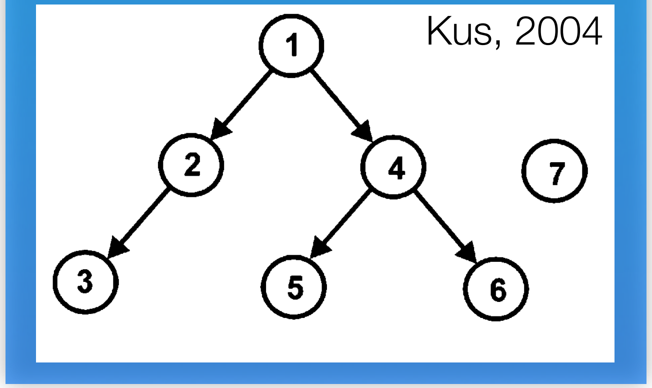


$$\mathbf{X}(t)=\sum_{k=1}^{p}\mathbf{A}^{(k)}(t)\mathbf{X}(t-k)+\mathbf{E}(t)$$

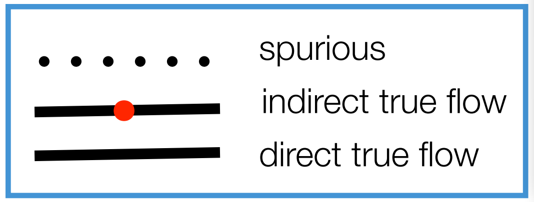
$$\mathbf{A}(f,t)=-\sum_{k=0}^{p}\mathbf{A}^{(k)}(t)e^{-i2pfk};\quad \mathbf{A}^{(0)}=I$$

$$\mathbf{X}(f,t)=\mathbf{A}(f,t)^{-1}\mathbf{E}(f,t)=\mathbf{H}(f,t)\mathbf{E}(f,t)$$

Ground Truth



Kus, 2004



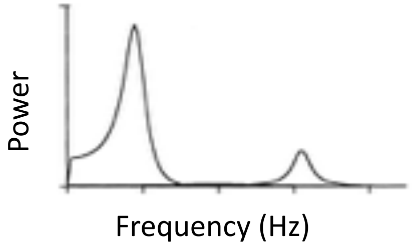
NOTE: time index (t) dropped for convenience

Functional

Spectrum

$$S(f)=\mathbf{X}(f)\mathbf{X}(f)^{*}$$
$$=\mathbf{H}(f)\Sigma\mathbf{H}(f)^{*}$$

(Brillinger, 2001)

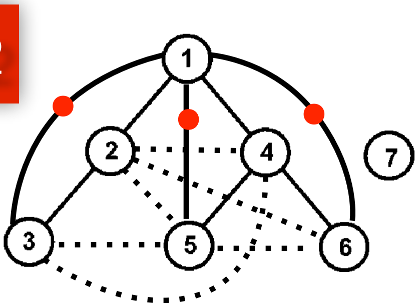


Coherency

M=2

$$C_{ij}(f)=\frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}$$

(Brillinger, 2001)

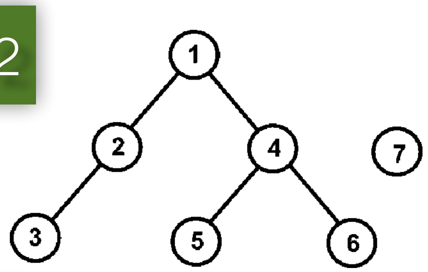


Partial Coherence

M>2

$$P_{ij}(f)=\frac{S_{ij}^{-1}(f)}{\sqrt{S_{ii}^{-1}(f)S_{jj}^{-1}(f)}}$$

(Brillinger, 2001)



Multiple Coherence

M=1

$$G_i(f)=\sqrt{1-\frac{\det(S(f))}{S_{ii}(f)\mathbf{M}_{ii}(f)}}$$

(Brillinger, 2001)

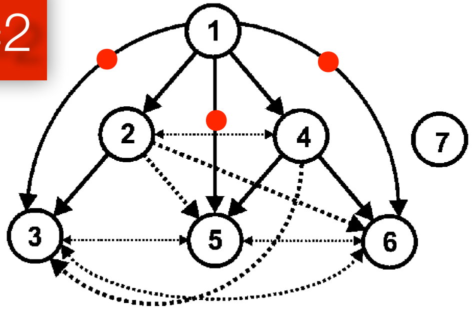
Effective

Granger-Geweke Causality

M=2

$$F_{ij}(f)=\frac{\Sigma_{jj}-(\Sigma_{ij}^{2}/\Sigma_{ii}))\,|H_{ij}(f)|^{2}}{S_{ii}(f)}$$

(Geweke, 1982; Bressler et al., 2007)

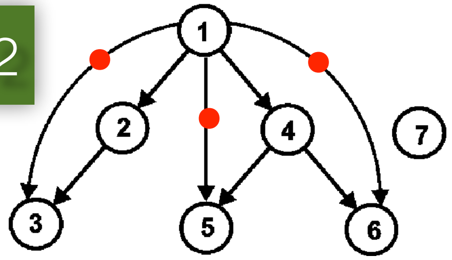


Directed Transfer Function

M>2

$$h_{ij}^{2}(f)=\frac{\left|H_{ij}(f)\right|^{2}}{\sum_{f}\sum_{k=1}^{M}\left|H_{ik}(f)\right|^{2}}$$

(Kaminski and Blinowska, 1991)

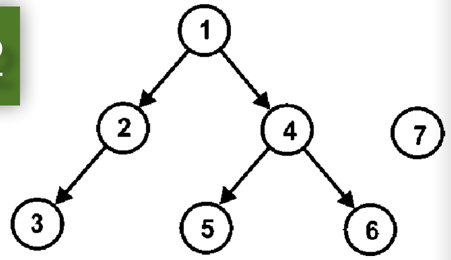


Partial Directed Coherence

M>2

$$p_{ij}^{2}(f)=\frac{|A_{ij}(f)|^{2}}{\sum_{k=1}^{M}\left|A_{kj}(f)\right|^{2}}$$

(Baccalá and Sameshima, 2001)

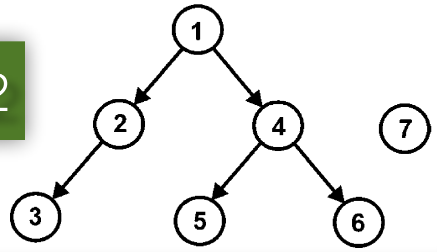


Direct DTF

M>2

$$d_{ij}^{2}(f)=h_{ij}^{2}(f)P_{ij}^{2}(f)$$

(Korzeniewska, 2003)



	Estimator	Formula
Spectral M.	Spectral Density Matrix	$S(f) = X(f)X(f)^* = H(f)\Sigma H(f)^*$
Coherence Measures	Coherency	$C_{ij}(f) = \dfrac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}}$ $0 \le \lvert C_{ij}(f)\rvert^2 \le 1$
	Imaginary Coherence (iCoh)	$iCoh_{ij}(f) = \mathrm{Im}(C_{ij}(f))$
	Partial Coherence (pCoh)	$P_{ij}(f) = \dfrac{\hat{S}_{ij}(f)}{\sqrt{\hat{S}_{ii}(f)\hat{S}_{jj}(f)}}$ $\hat{S}(f) = S(f)^{-1}$ $0 \le \lvert P_{ij}(f)\rvert^2 \le 1$
	Multiple Coherence (mCoh)	$G_i(f) = \sqrt{1 - \dfrac{\det(S(f))}{S_{ii}(f)\mathbf{M}_{ii}(f)}}$ $\mathbf{M}_{ii}(f)$ is the **minor** of *S*(*f*) obtained by removing the ith row and column of *S*(*f*) and returning the determinant.

	Estimator	Formula
Partial Directed Coherence Measures	Normalized Partial Directed Coherence (PDC)	$\pi_{ij}(f) = \dfrac{A_{ij}(f)}{\sqrt{\sum_{k=1}^{M} \lvert A_{kj}(f)\rvert^2}}$ $0 \le \lvert \pi_{ij}(f)\rvert^2 \le 1$ $\sum_{j=1}^{M} \lvert \pi_{ij}(f)\rvert^2 = 1$
	Generalized PDC (GPDC)	$\bar{\pi}_{ij}(f) = \dfrac{\frac{1}{\Sigma_{ii}} A_{ij}(f)}{\sqrt{\sum_{k=1}^{M} \frac{1}{\Sigma_{ii}^2} \lvert A_{kj}(f)\rvert^2}}$ $0 \le \lvert \bar{\pi}_{ij}(f)\rvert^2 \le 1$ $\sum_{j=1}^{M} \lvert \bar{\pi}_{ij}(f)\rvert^2 = 1$
	Renormalized PDC (rPDC)	$\lambda_{ij}(f) = Q_{ij}(f)^* V_{ij}(f)^{-1} Q_{ij}(f)$ where $Q_{ij}(f) = \begin{pmatrix} \mathrm{Re}[A_{ij}(f)] \\ \mathrm{Im}[A_{ij}(f)] \end{pmatrix}$ and $V_{ij}(f) = \sum_{k,l=1}^{p} R_{jj}^{-1}(k,l)\Sigma_{ii} Z(2\pi f,k,l)$ $Z(\omega,k,l) = \begin{pmatrix} \cos(\omega k)\cos(\omega l) & \cos(\omega k)\sin(\omega l) \\ \sin(\omega k)\cos(\omega l) & \sin(\omega k)\sin(\omega l) \end{pmatrix}$ ***R*** is the [(*Mp*)² x (*Mp*)²] covariance matrix of the VAR[*p*] process (Lütkepohl, 2006)
Granger-Geweke	Granger-Geweke Causality (GGC)	$F_{ij}(f) = \dfrac{\left(\Sigma_{jj} - (\Sigma_{ij}^2/\Sigma_{ii})\right)\lvert H_{ij}(f)\rvert^2}{S_{ii}(f)}$

	Estimator	Formula
Directed Transfer Function Measures	Normalized Directed Transfer Function (DTF)	$\gamma_{ij}(f) = \dfrac{H_{ij}(f)}{\sqrt{\sum_{k=1}^{M} \lvert H_{ik}(f)\rvert^2}}$ $0 \le \lvert \gamma_{ij}(f)\rvert^2 \le 1$ $\sum_{j=1}^{M} \lvert \gamma_{ij}(f)\rvert^2 = 1$
	Full-Frequency DTF (ffDTF)	$\eta_{ij}^2(f) = \dfrac{\lvert H_{ij}(f)\rvert^2}{\sum_f \sum_{k=1}^{M} \lvert H_{ik}(f)\rvert^2}$
	Direct DTF (dDTF)	$\delta_{ij}^2(f) = \eta_{ij}^2(f) P_{ij}^2(f)$

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

$$\mathbf{A}(f,t) = -\sum_{k=0}^{p} \mathbf{A}^{(k)}(t) e^{-i2pfk}; \quad \mathbf{A}^{(0)} = I$$

$$\mathbf{X}(f,t) = \mathbf{A}(f,t)^{-1}\mathbf{E}(f,t) = \mathbf{H}(f,t)\mathbf{E}(f,t)$$

$H(f)$ Transfer Function

$A(f)$ System Matrix

Σ Noise Covariance Matrix

🟨 Variance Stabilization

For additional details, see SIFT Handbook (sccn.ucsd.edu/wiki/SIFT)

Multivariate Models: Limitations

- However, multivariate methods come with a cost:
  - More parameters + limited data = higher risk of over-fitting or worse yet....
  - ...the problem becomes *ill-posed* or *under-determined*. There are insufficient observations to uniquely determine a solution to the system of equations defining our model.

Multivariate Models: Limitations

How many samples do we need?

- N = number of samples required
- M = number of variables/sources
- T = number of trials/realizations
- p = model order
- We have M^2p model coefficients to estimate. So our ordinary least-squares solution requires a *minimum* of M^2p samples.

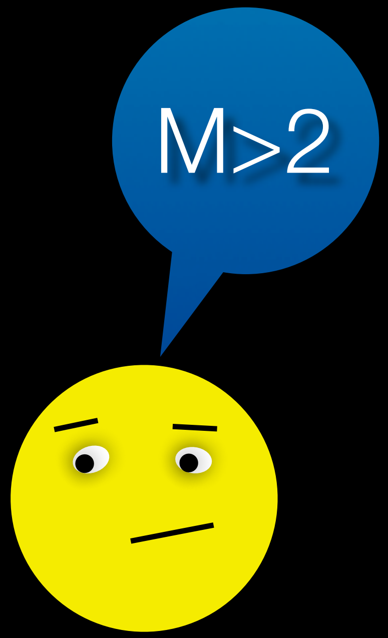
$$N \geq M^2 p$$

- Back-of-envelope: M=20, p=10, T=1 We need 20^2x10 = 4000 samples -- 20 second epoch at sampling rate of 200Hz!

Ensemble aggregation (T > 1)?

- M=20, p=10, T=50: 4000/50 samples/trial → 20/50 = 0.4 sec epoch

Multivariate Models: Constraints



Solutions?

Make assumptions (impose constraints)

We want to *a priori* restrict the range of allowable values for our parameters -- transforming the problem from one with infinite number of solutions in the original parameter space to one with a unique (“best”) solution in the new parameter space

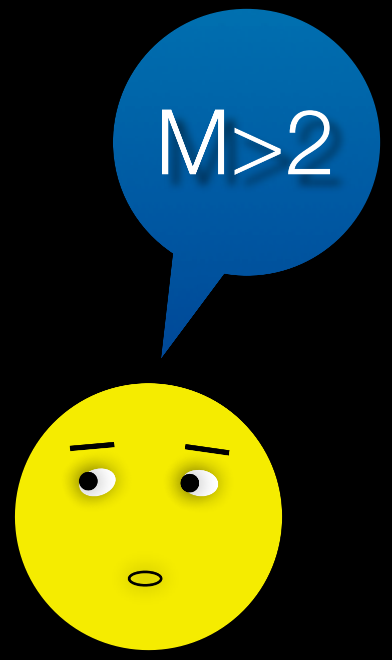
In a Bayesian context, this corresponds to making assumptions about the *prior distribution* of the parameters (Gaussian, Laplacian, ...)

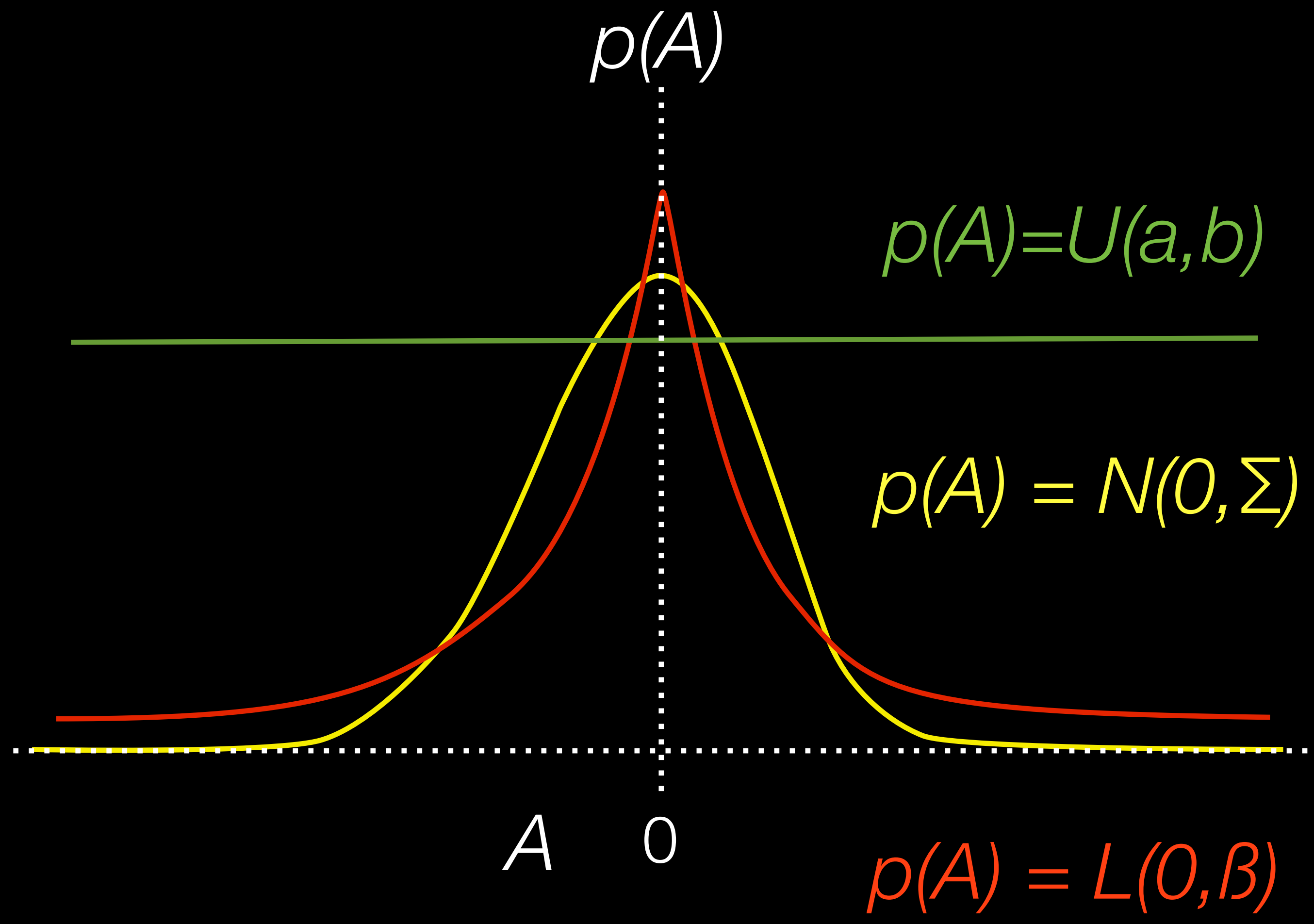
Tim Mullen

Multivariate Models: Constraints

posterior likelihood prior

$$\hat{A} = \underset{A}{\arg\max} \left\{ p(A|D) \cong p(D|A)\boxed{p(A)} \right\}$$





Unconstrained (all values equally probable). e.g. Uniform distribution

Smoothness constraints

- large differences in values unlikely
- small (non-zero) values most probable. e.g. Normal (gaussian) prior.

Sparsity constraint

- many values small or exactly zero with occasional large values e.g. Laplacian prior

Smoothness Constraints

Tim Mullen

- Standard least-squares solution

$$A(t) = \arg\min_{\hat{A}} \left(\overbrace{\left\| Y - Z\tilde{A} \right\|_2^2}^{\text{prediction error}} \right)$$

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

$$\tilde{A} = [A^{(1)}(t), \ldots, A^{(p)}(t)]^T$$

$$X_k = [X(p+1-k), \ldots, X(N-k)]^T$$

$$Z = [X_1, \ldots, X_p]$$

$$Y = X_0$$

Rewrite VAR[p] as VAR[1]

Smoothness Constraints

Tim Mullen

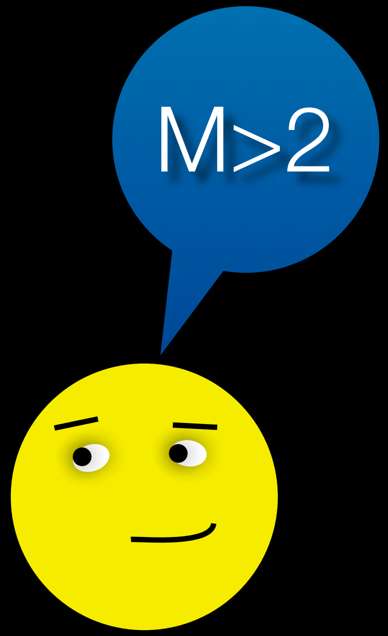
- Ridge Regression
  (Tikhonov Regularization, Minimum-(L_2)-Norm

prediction error

penalty term,
enforces smoothness

$$A(t) = \arg\min_{\hat{A}} \left(\overbrace{\left\| Y - Z\tilde{A} \right\|_2^2}^{\text{prediction error}} + \boxed{l} \left\| \tilde{A} \right\|_2^2 \right)$$

regularization



- Equivalent to assuming a Gaussian prior with variance determined by λ
- Large values of *A* are penalized. The range of allowable values for coefficients is restricted, reducing the *effective* degrees of freedom and allowing us to estimate VAR coefficients with fewer observations.

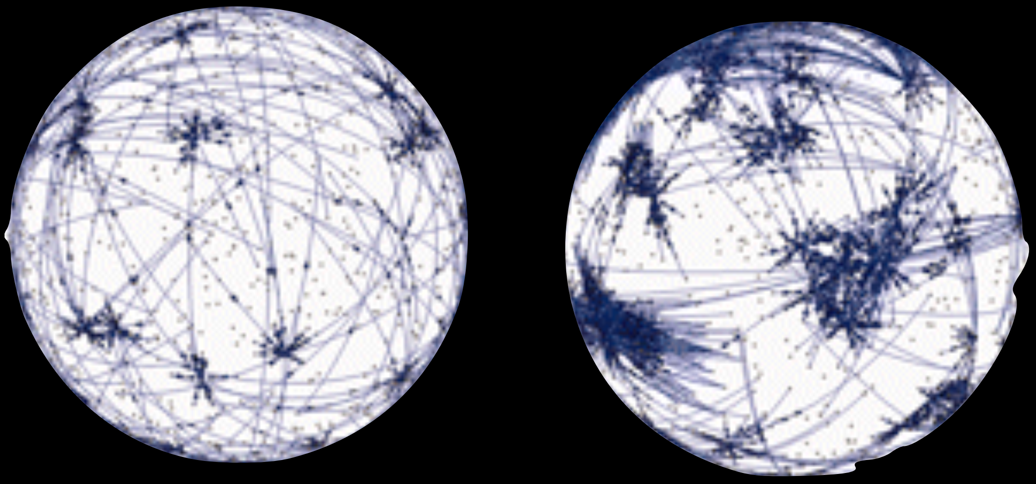
Tim Mullen

Sparsity Constraints

Sparsity

- Relatively low probability of a *direct* connection between any two anatomical functional units. This probability decreases with distance

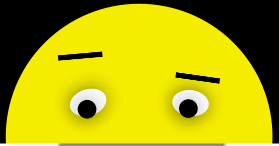
It's a small world...

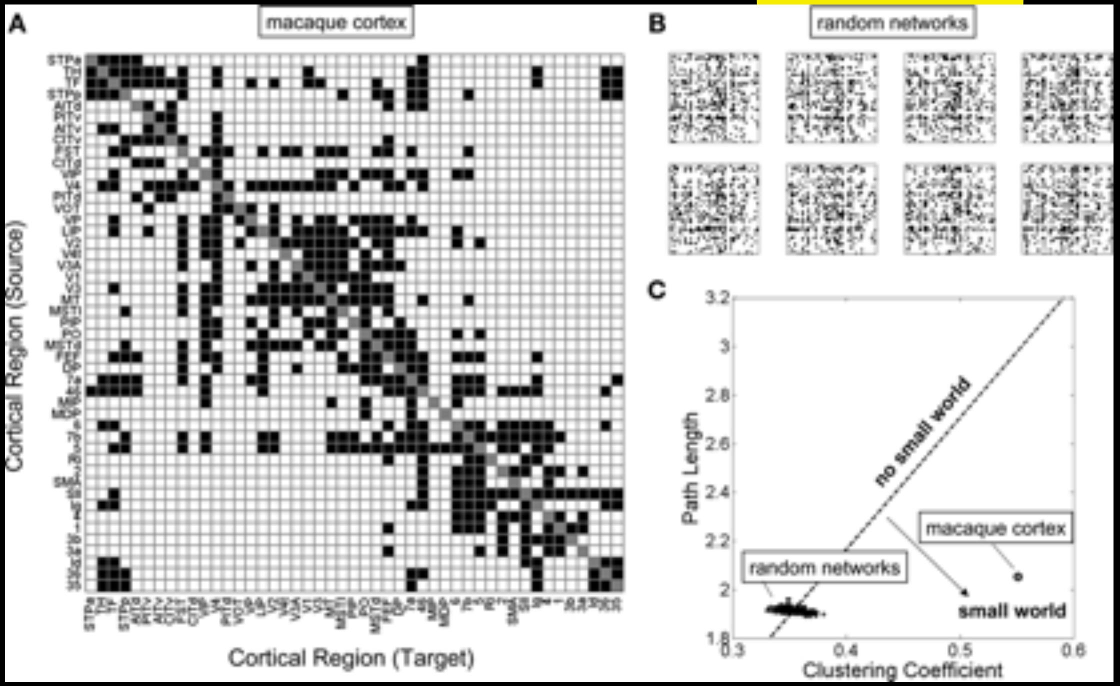


structural network

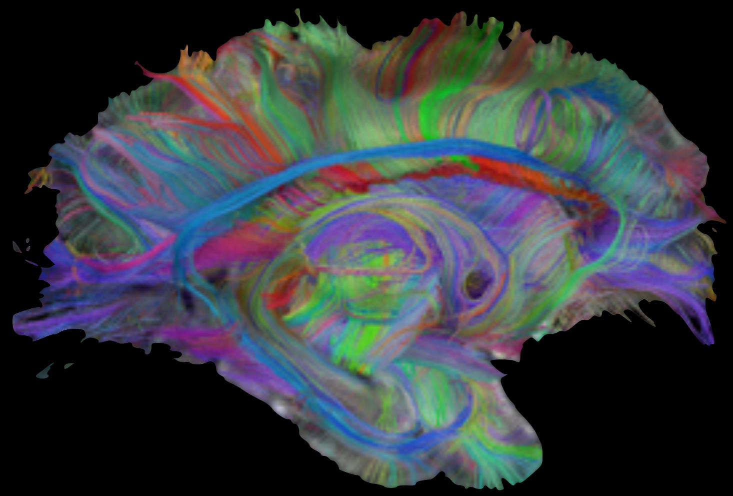
functional network

Sporns and Honey, *PNAS*, 2006





Sporns, *Frontiers in Computational Neuroscience*, 2011



Structural Connectivity

Tim Mullen

Sparsity Constraints

- Standard least-squares solution

$$A(t) = \arg\min_{\hat{A}} \left(\overbrace{\left\| Y - Z\tilde{A} \right\|_2^2}^{\text{prediction error}} \right)$$

$$\mathbf{X}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{X}(t-k) + \mathbf{E}(t)$$

$$\tilde{A} = [A^{(1)}(t),\ldots,A^{(p)}(t)]^T$$

$$X_k = [X(p+1-k),\ldots,X(N-k)]^T$$

$$Z = [X_1,\ldots,X_p]$$

$$Y = X_0$$

Rewrite VAR[p] as VAR[1]

Tim Mullen

Sparsity Constraints

- Group Lasso ($L_{1,2}$ norm)

$$A(t) = \arg\min_{\hat{A}} \left(\overbrace{\left\| Y - Z\tilde{A} \right\|_2^2}^{\text{prediction error}} + \underset{\text{regularization}}{\boxed{l}} \underbrace{\sum_{ij} \overbrace{\left\| \tilde{A}_{ij}^{(1)}, \dots, \tilde{A}_{ij}^{(p)} \right\|_2}^{\text{smoothness (L2) (preserves spectrum)}}}_{\text{group sparsity (L1)}} \right)$$

ADMM
DAL

- Equivalent to assuming a Gaussian prior over coefficients within groups and a Laplacian prior over the groups themselves
- Entire groups of coefficients are jointly pruned (set *exactly* to zero) while remaining groups assumed to have a Gaussian prior (ridge penalty). Allows us to estimate VAR coefficients with fewer observations.

Tim Mullen

Sparsity Constraints

Compressive Sensing

- The process of acquiring and reconstructing a quantity that is underdetermined but known to be sparse (compressible) in some basis

How many samples do we need?

- N = number of samples required
- *M = number of variables/sources, p = model order*

$$N = O\left(K \log(M^2 p / K)\right) \approx O\left(\log M^2 p\right) \le c \log\left(M^2 p\right)$$

$$N \ge M^2 p$$

(unconstrained)

Tim Mullen

Constraints Improve Estimation
(if prior assumptions are correct)

- Significant improvements using smoothness or sparsity assumptions
- (e.g. Haufe et al, 2009, Valdez-Sosa et al, 2009)

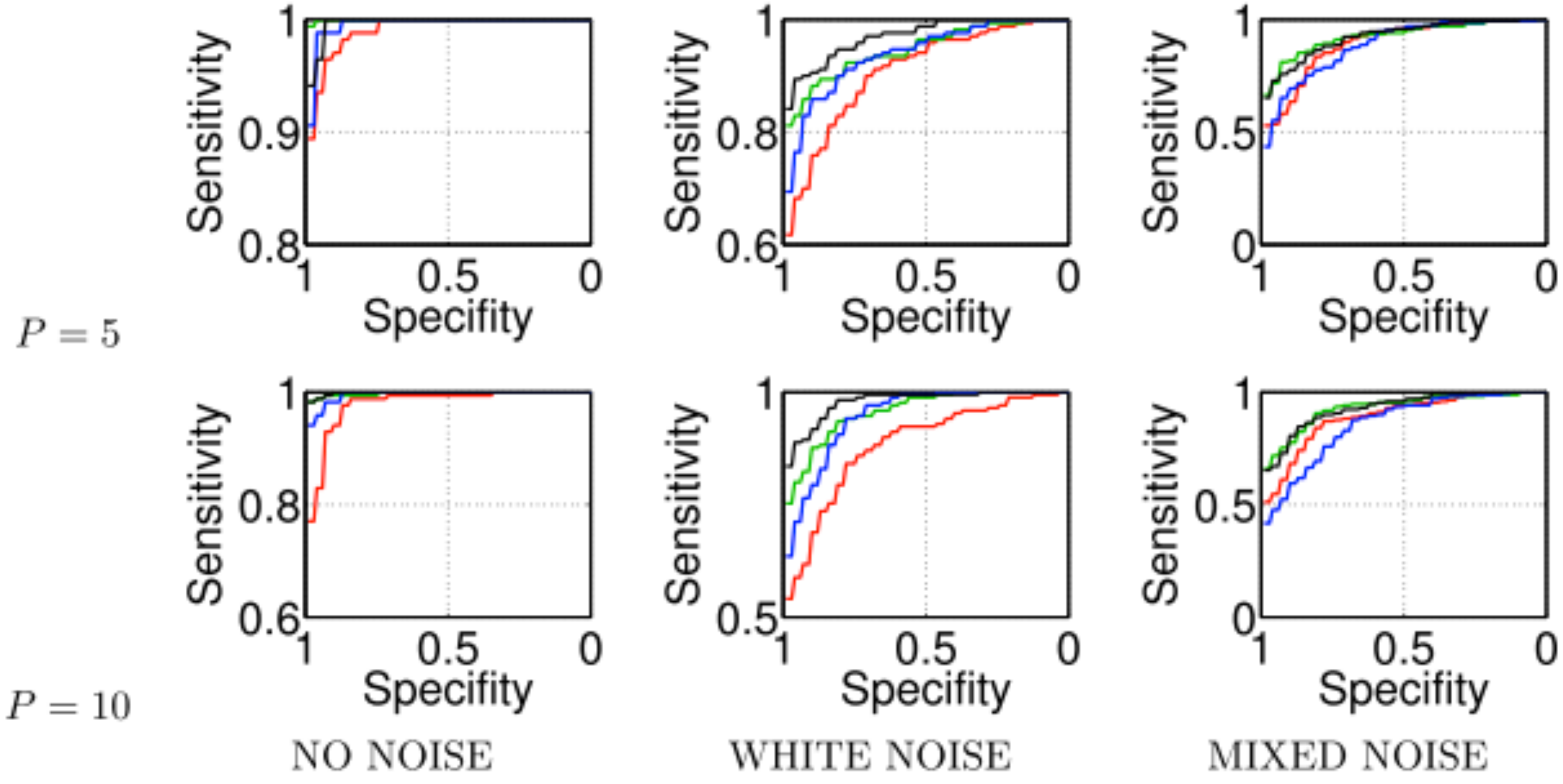


Figure 2: Average ROC curves of Granger Causality (red), Ridge Regression (green), Lasso (blue) and Group Lasso (black) in three different noise conditions and for two different model orders.

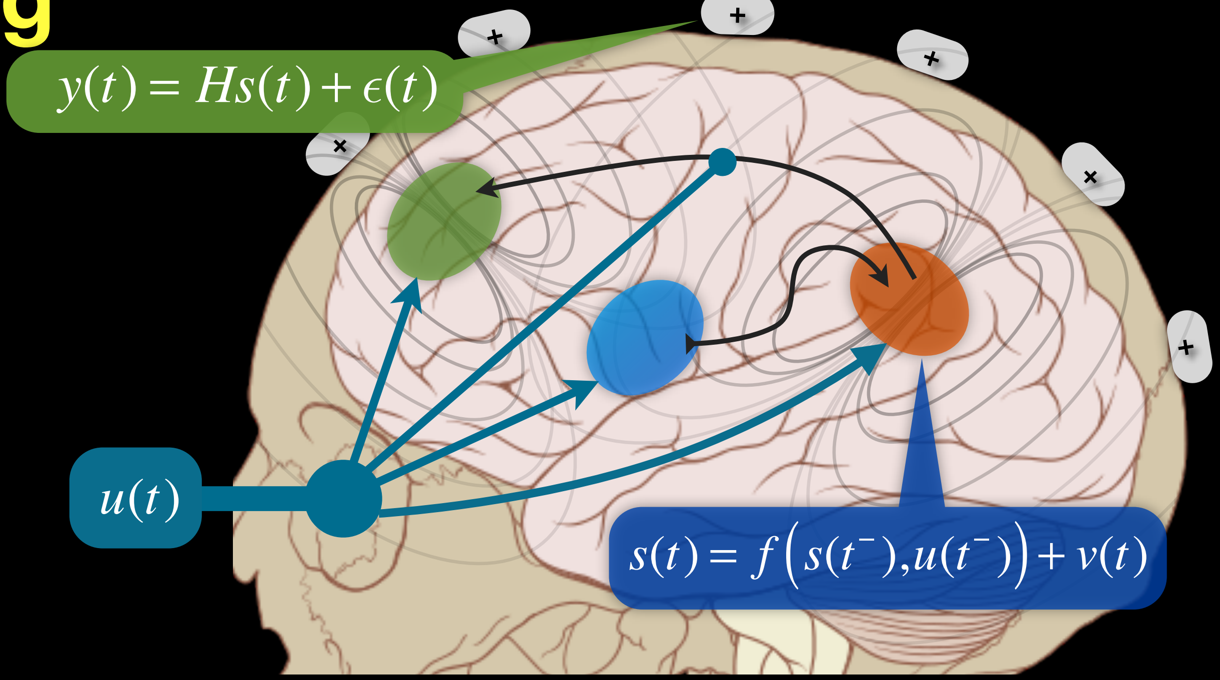
Haufe, 2009

Tim Mullen

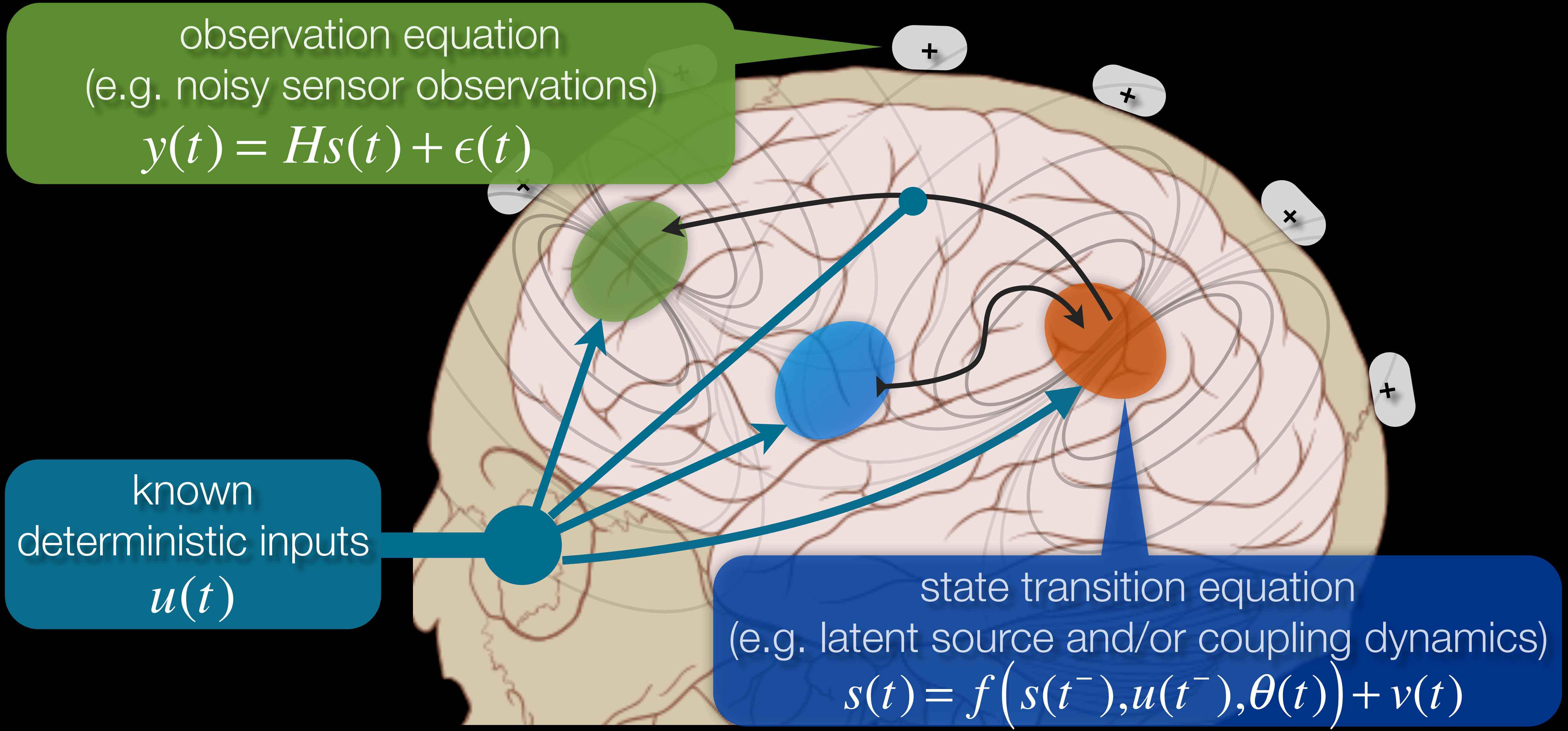
Adapting to Non-Stationarity

- **Many ways to do adaptive VAR estimation**
- Two popular approaches (adopted in SIFT):
  - Segmentation-based adaptive VAR estimation (assumes local stationarity)
  - **State-Space Modeling**

Kalman Filtering and extensions



Discrete State-Space Model (SSM) for Electrophysiological Dynamics



observation equation
(e.g. noisy sensor observations)
$y(t) = Hs(t) + \epsilon(t)$

known
deterministic inputs
$u(t)$

state transition equation
(e.g. latent source and/or coupling dynamics)
$s(t) = f\left(s(t^-), u(t^-), \theta(t)\right) + v(t)$

Linear VAR[1] $\quad s(t) = \mathbf{A}(t)s(t-1) + v(t)$

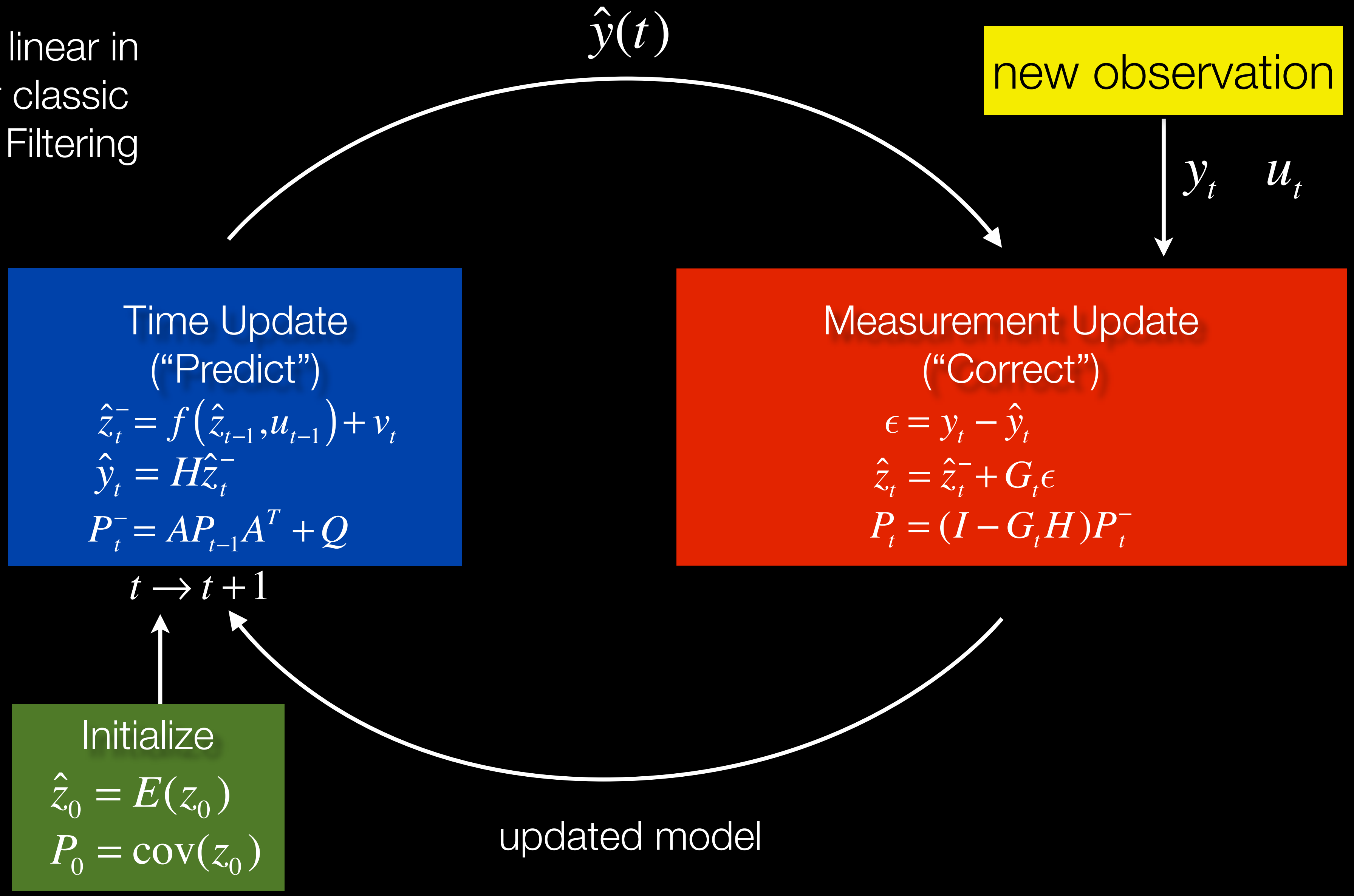
- Dynamical system may be linear or nonlinear, dense or sparse, non-stationary, high-dimensional, partially-observed, and stochastic
- Subsumes discrete Delay Differential Equation (DDE) and Vector Autoregressive (VAR) methods and closely related to Dynamic Causal Modeling (DCM)

Kalman Filtering

Tim Mullen

optimal estimator (in terms of minimum variance) for the state of a linear dynamical system

$f(\mathrm{z,u})$ is linear in $\{\mathrm{z,u}\}$ for classic Kalman Filtering



$\hat{y}(t)$

new observation

$y_t \quad u_t$

Time Update
("Predict")

$$\hat{z}_t^- = f\left(\hat{z}_{t-1}, u_{t-1}\right) + v_t$$

$$\hat{y}_t = H\hat{z}_t^-$$

$$P_t^- = AP_{t-1}A^T + Q$$

$$t \rightarrow t+1$$

Measurement Update
("Correct")

$$\epsilon = y_t - \hat{y}_t$$

$$\hat{z}_t = \hat{z}_t^- + G_t\epsilon$$

$$P_t = (I - G_t H)P_t^-$$

Initialize

$$\hat{z}_0 = E(z_0)$$

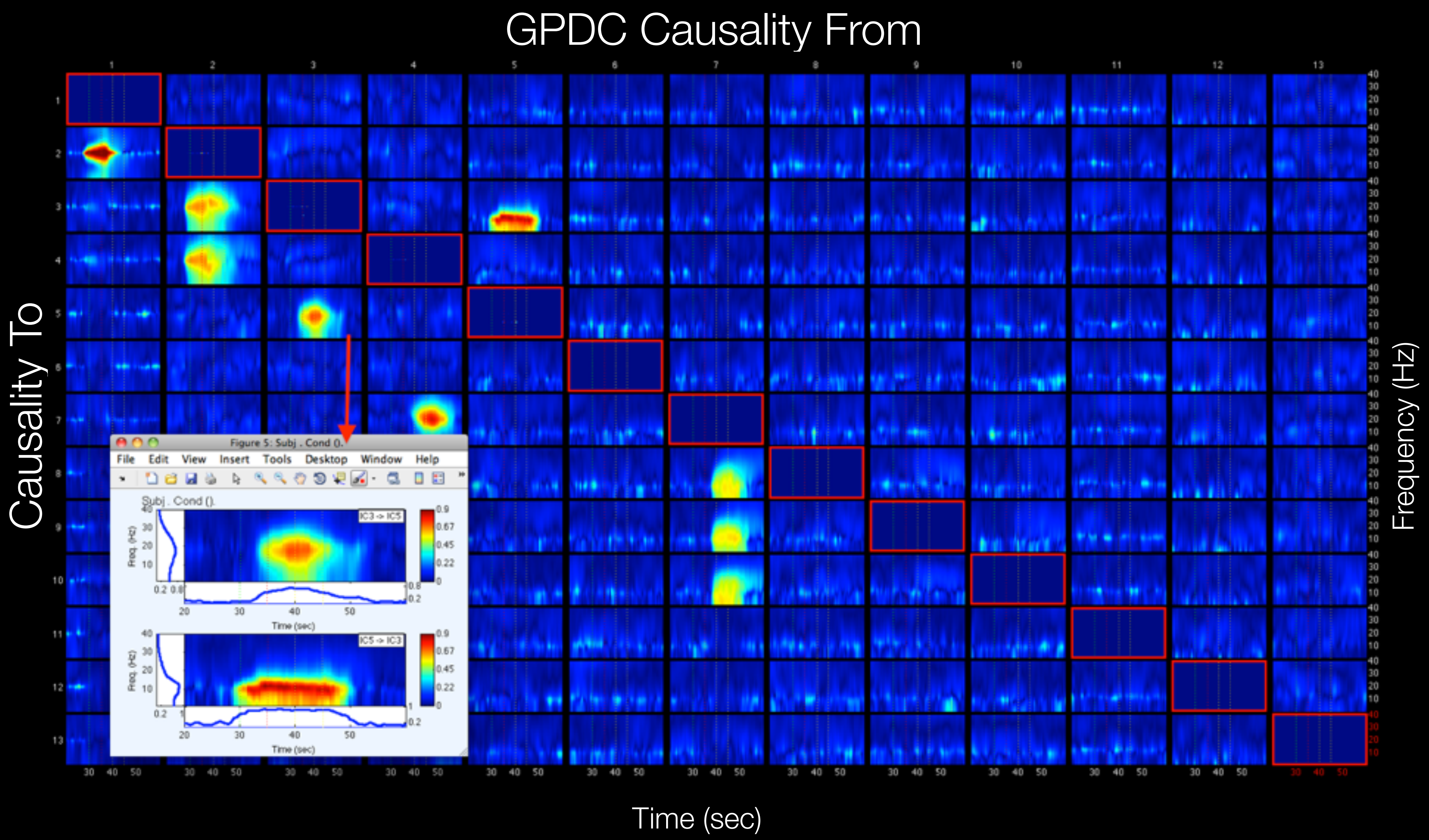
$$P_0 = \mathrm{cov}(z_0)$$

updated model

z_t := unknown state vector at time *t*
e.g. delay-embedding of sources and/or coupling (VAR) parameters

Kalman Filtering

Swartz Center for Computational Neuroscience
Tim Mullen



Nonlinear Modeling

- Interactions in brain are generally non-linear
- Purely linear models (e.g. high-order VAR models) can sometimes provide an approximation sufficient for correct detection of directed dependencies

Schelter B, Timmer J, Eichler M., *J. Neuro. Methods* 2009

van-der-Pol coupled oscillators

$$\ddot{x}_i = \mu(1 - x_i^2)\dot{x}_i + \omega_i^2 x_i + \sigma_i \eta_i + \sum_{j \neq i} \epsilon_{ij}(x_j - x_i)$$

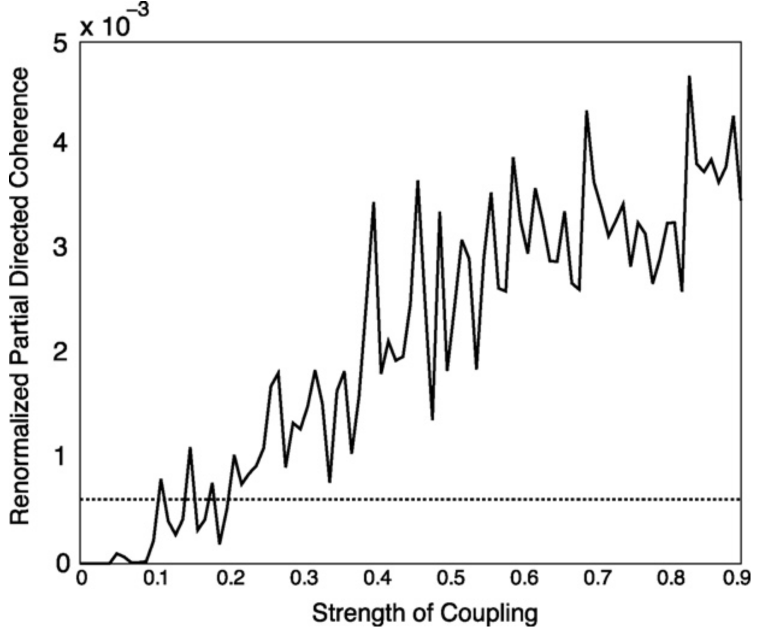


Fig. 8. Coupled van-der-Pol oscillators. Renormalized partial directed coherence for various coupling strengths ϵ_{12}. The dashed horizontal line marks the 5% significance level.

Stochastic coupled Rössler oscillators

$$\dot{\xi}_j = \begin{pmatrix} \dot{X}_j \\ \dot{Y}_j \\ \dot{Z}_j \end{pmatrix} = \begin{pmatrix} -\omega_j Y_j - Z_j + \left[\sum_{i,i \neq j} \varepsilon_{j,i} \left(X_i - X_j \right) \right] + \sigma_j \eta_j \\ \omega_j X_j + a Y_j \\ b + (X_j - c) Z_j \end{pmatrix}$$

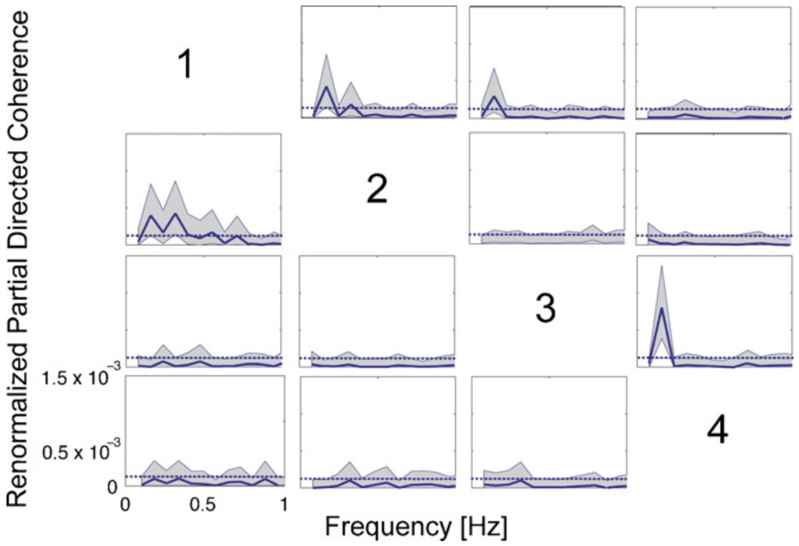
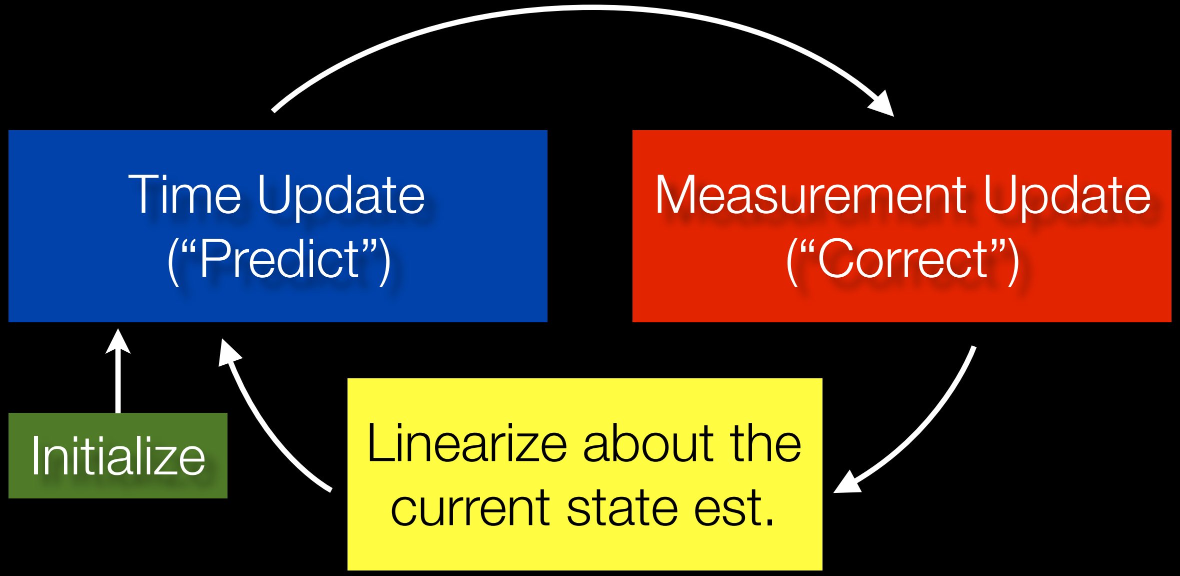


Fig. 9. Results of the renormalized partial directed coherence analysis for a network coupled Rössler oscillators. The coupling is bidirectional between oscillators 1 and 2 and unidirectional from oscillator 2 to oscillator 4 and from oscillator 4 to oscillator 3. The dashed horizontal line marks the 5% significance level, while the gray area represents the 95% confidence intervals.

Nonlinear Modeling

- A more general approach is to transform the non-linear system to a linear representation and apply the linear model there.
- e.g. **local-linearization**, kernel methods, etc.
- Nonlinear extensions of Kalman filtering provide efficient ways to model the time-evolving states and parameters of nonlinear processes
  - (Dual) Extended Kalman Filtering



Time Update ("Predict")

Measurement Update ("Correct")

Initialize

Linearize about the current state est.

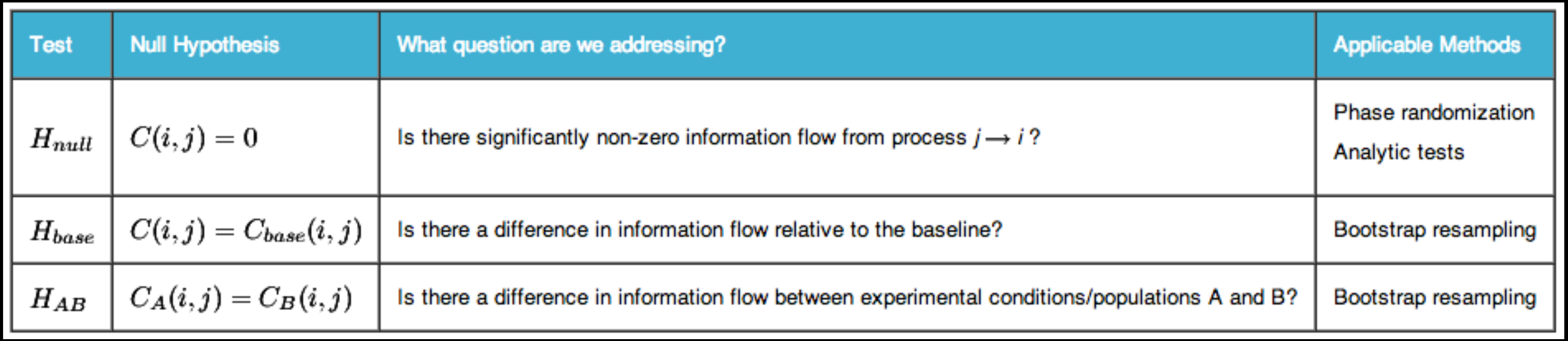
- Linearize about the current state (first-order Taylor approximation).
- Apply the Kalman Filter update rules using the linearized model
- Apply Granger-Geweke Causality to the linearized coefficient matrices

Kalman Filtering in SIFT

- Linear Kalman Filter: est_fitMVARKalman.m
- Nonliear Kalman Filter: est_fitMVAR_DEKF.m

Statistics

- Different ways to do statistics in SIFT
  - Phase Randomization
  - Bootstrapping
  - Analytic Tests



Test	Null Hypothesis	What question are we addressing?	Applicable Methods
H_{null}	$C(i,j) = 0$	Is there significantly non-zero information flow from process $j \longrightarrow i$?	Phase randomization Analytic tests
H_{base}	$C(i,j) = C_{base}(i,j)$	Is there a difference in information flow relative to the baseline?	Bootstrap resampling
H_{AB}	$C_A(i,j) = C_B(i,j)$	Is there a difference in information flow between experimental conditions/populations A and B?	Bootstrap resampling

C(i,j) is the measured information flow from process j --> i.
C_{null} is the expected measured information flow when there is no true information flow.
C_{base} is the expected information flow in some baseline period.

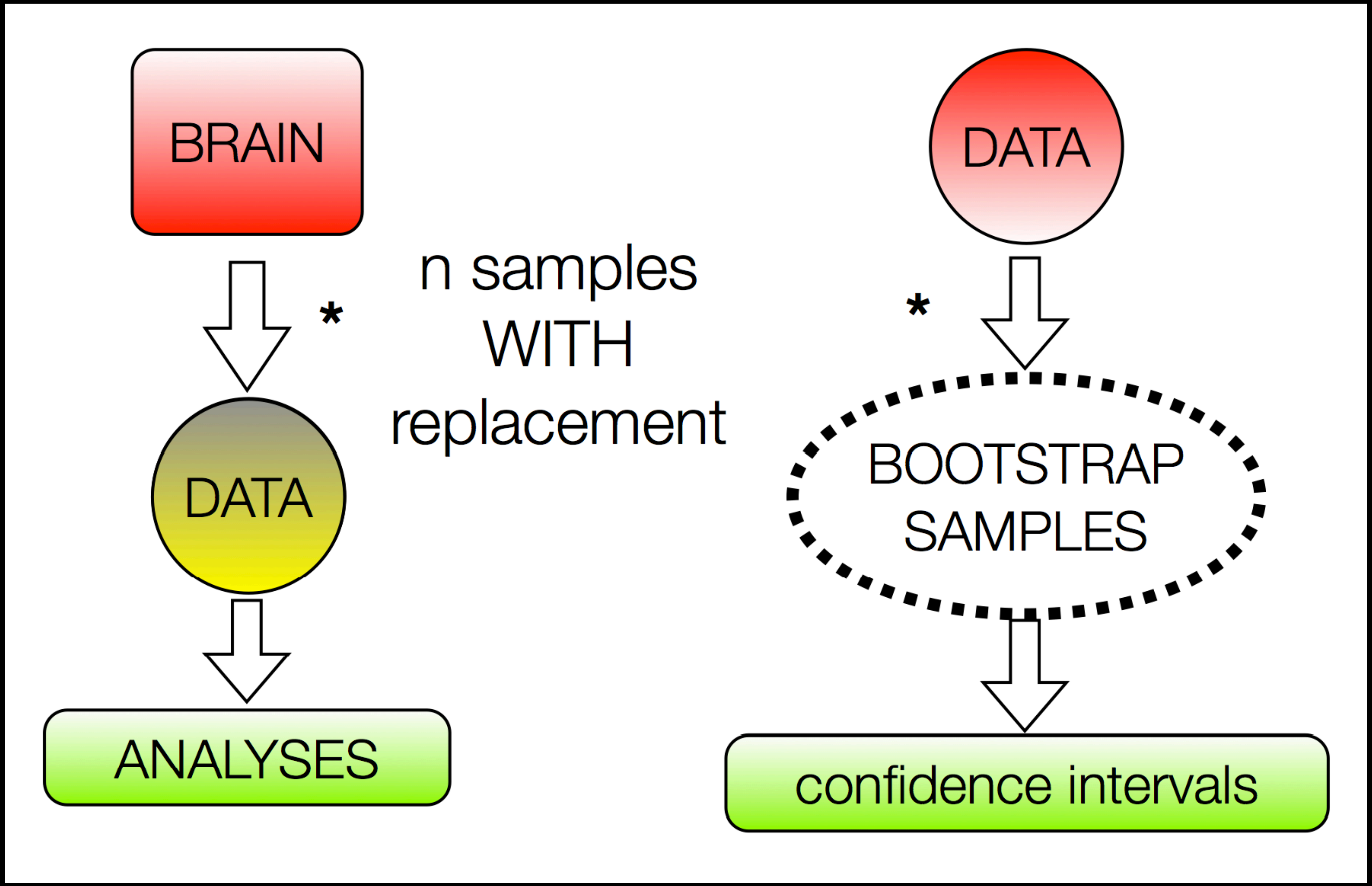
Tim Mullen

Statistics

Statistical Approach	Test	Parametric	Nonparam.
Asymptotic analytic estimates of confidence intervals. Applies to: PDC, nPDC, DTF, nDTF, rPDC	H_{null}, H_{base}, H_{AB}	☑	
Theiler phase randomization Applies to: all	H_{null}		☑
Bootstrap, Jacknife, Cross-Validation Applies to: all	H_{AB}, H_{base}		☑
Confidence intervals using Bayesian smoothing splines Applies to: all	H_{base}, H_{AB}	☑	☑

$H_{null} : \mathbf{C}_{ij} = 0$ $\quad H_{base}: \mathbf{C}_{ij} = \mathbf{C}_{baseline}$ $\quad H_{AB}: \mathbf{C}^{A}_{ij} = \mathbf{C}^{B}_{ij}$

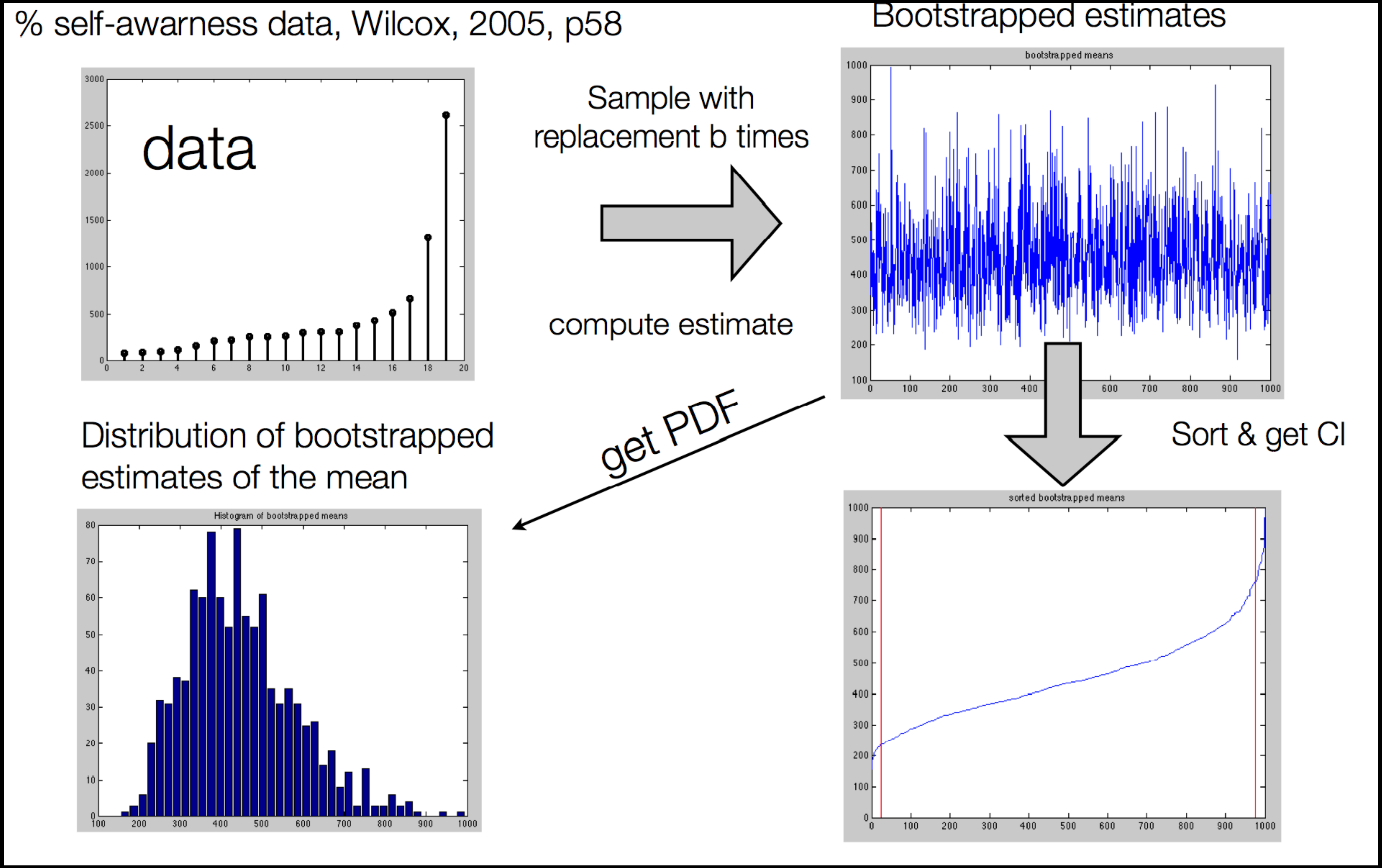
Bootstrap Statistics



- sample = X1, ..., Xn
- for k=1:R (number of bootstrap resamples/iterations)
  - resample n observations (trials) with replacement $X^* = \{X^*_1, \dots X^*_n\}$
  - compute estimator E_k (fit model, obtain connectivity) based on X^*
  - repeat
- with R large enough $P_E = \{E_1, \dots, E_R\}$ provides a good approximation to the true distribution of the estimator (connectivity, power, etc)

Bootstrap Statistics

Tim Mullen



% self-awarness data, Wilcox, 2005, p58

Sample with replacement b times

compute estimate

Bootstrapped estimates

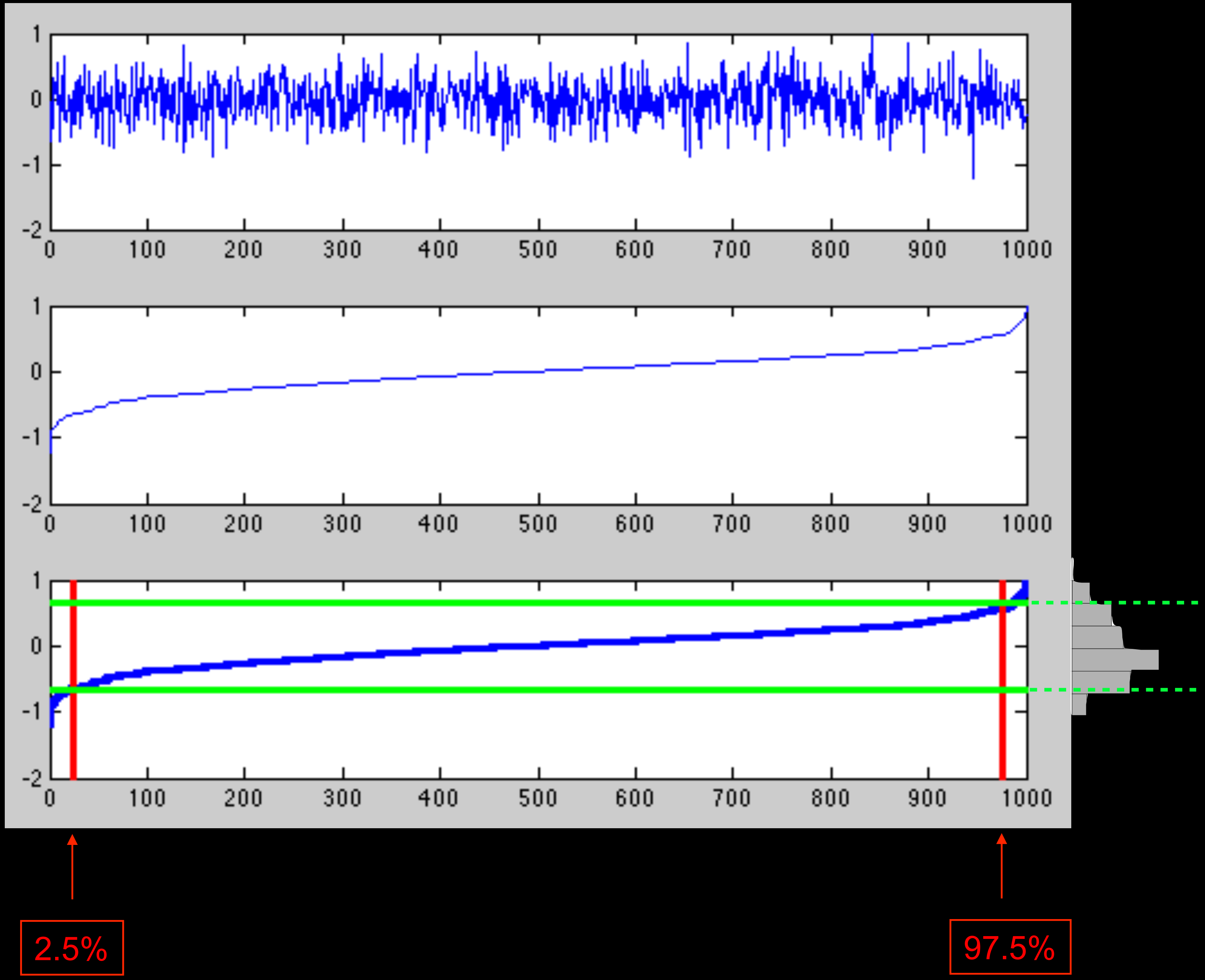
Sort & get CI

get PDF

Distribution of bootstrapped estimates of the mean

Bootstrap Statistics

Swartz Center for Computational Neuroscience
Tim Mullen



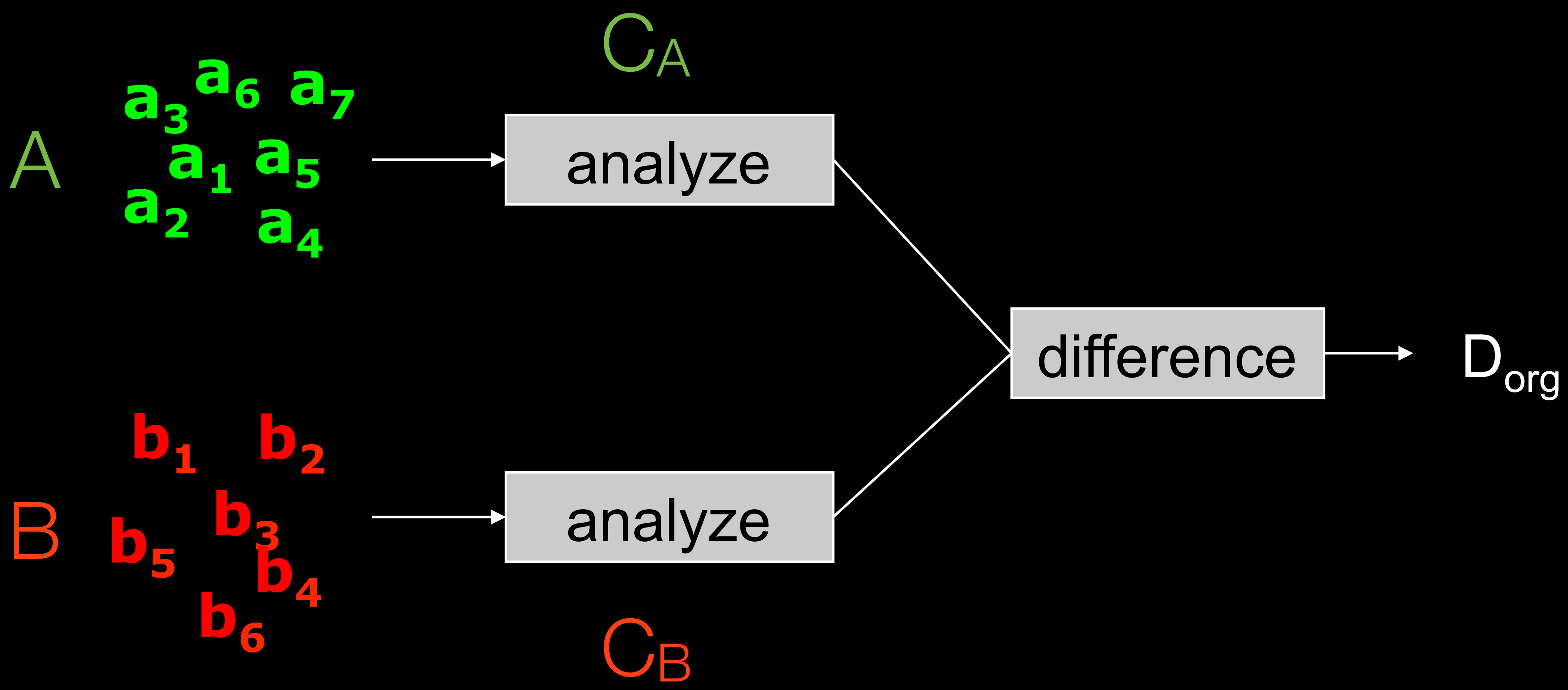
bootstrap

sorted values
(cdf)

thresholds
(ci)

Bootstrap Differences

- Suppose we have two conditions
  A = {a1,...a7}
- B = {b1,...,b6}
- We want to estimate the distributions of connectivity estimator applied to A and B separately, as well as the difference distribution (for testing H0: A=B)



C_A

A a_3 a_6 a_7 a_1 a_5 a_2 a_4

analyze

difference

D_{org}

B b_1 b_2 b_3 b_5 b_4 b_6

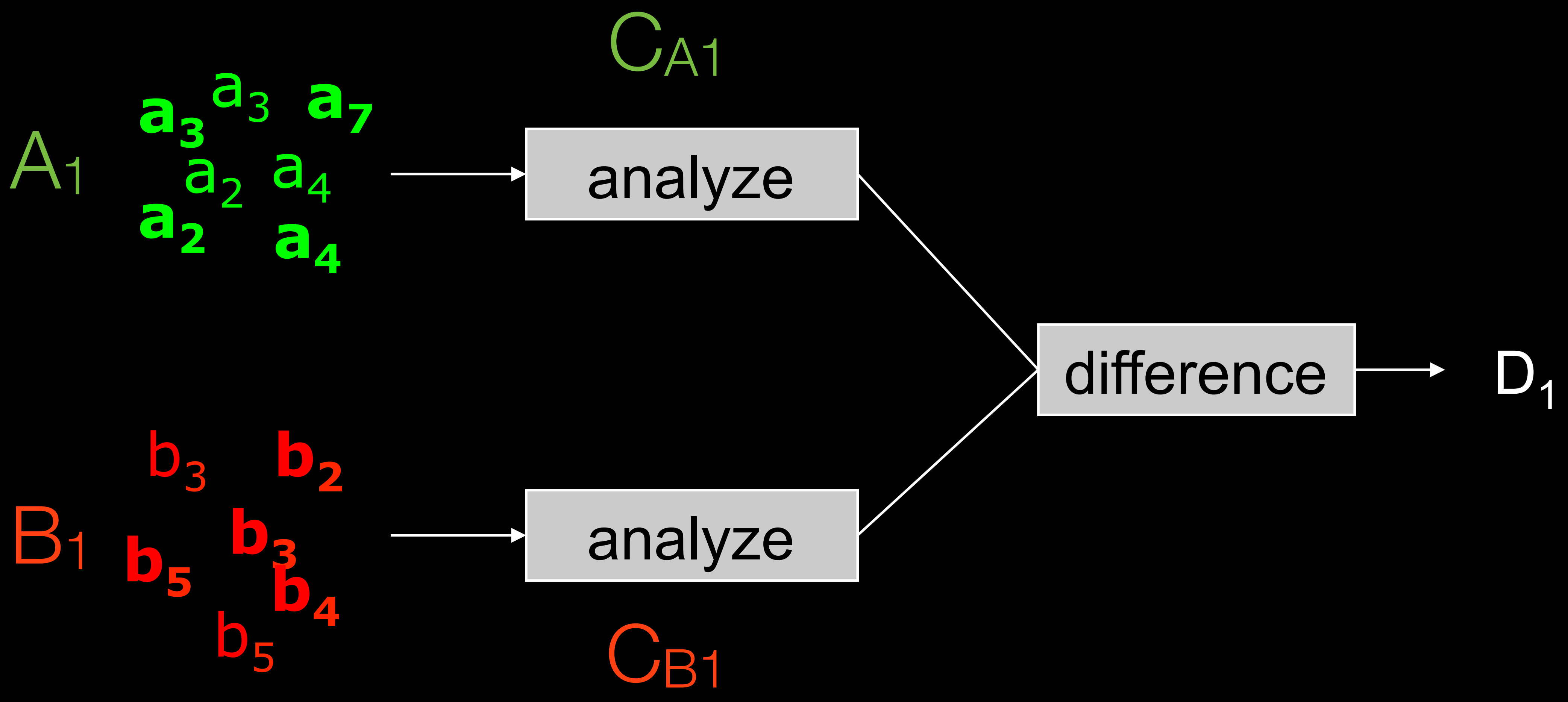
analyze

C_B

Bootstrap Differences

Tim Mullen

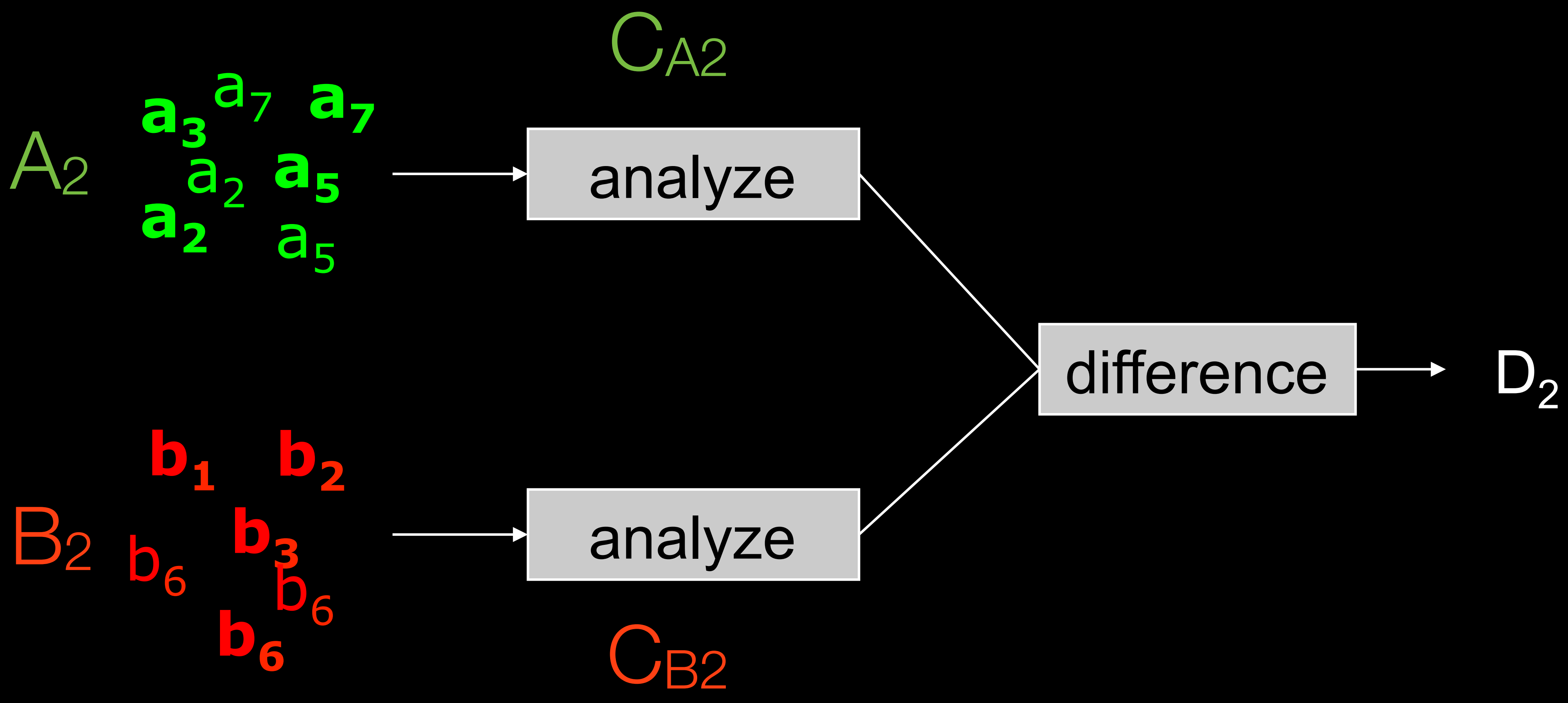
- For k=1:R (number of bootstrap iterations)
  - Resample with replacement from both groups to get A_k and B_k
  - Fit models and obtain connectivity C_{Ak}, C_{Bk}
  - Compute difference $D_k = C_{Ak} - C_{Bk}$
- Repeat



Bootstrap Differences

Tim Mullen

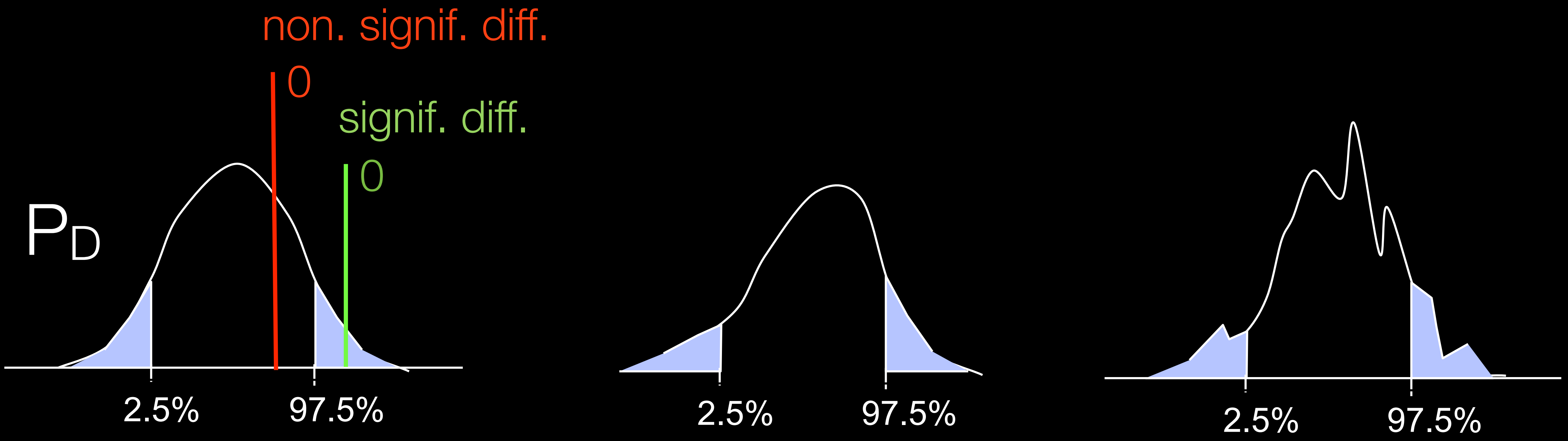
- For k=1:R (number of bootstrap iterations)
    - Resample with replacement from both groups to get A_k and B_k
    - Fit models and obtain connectivity C_{Ak}, C_{Bk}
    - Compute difference $D_k = C_{Ak}-C_{Bk}$
- Repeat



Bootstrap Statistics

Tim Mullen

- The procedure yields a distribution $P_D = \{D_1, \ldots. D_R\}$
- If 0 lies in the right (or left) tail of this "difference distribution", then we **reject** the null hypothesis that A=B at the chosen confidence level (below: alpha=0.05 for a two-sided test)

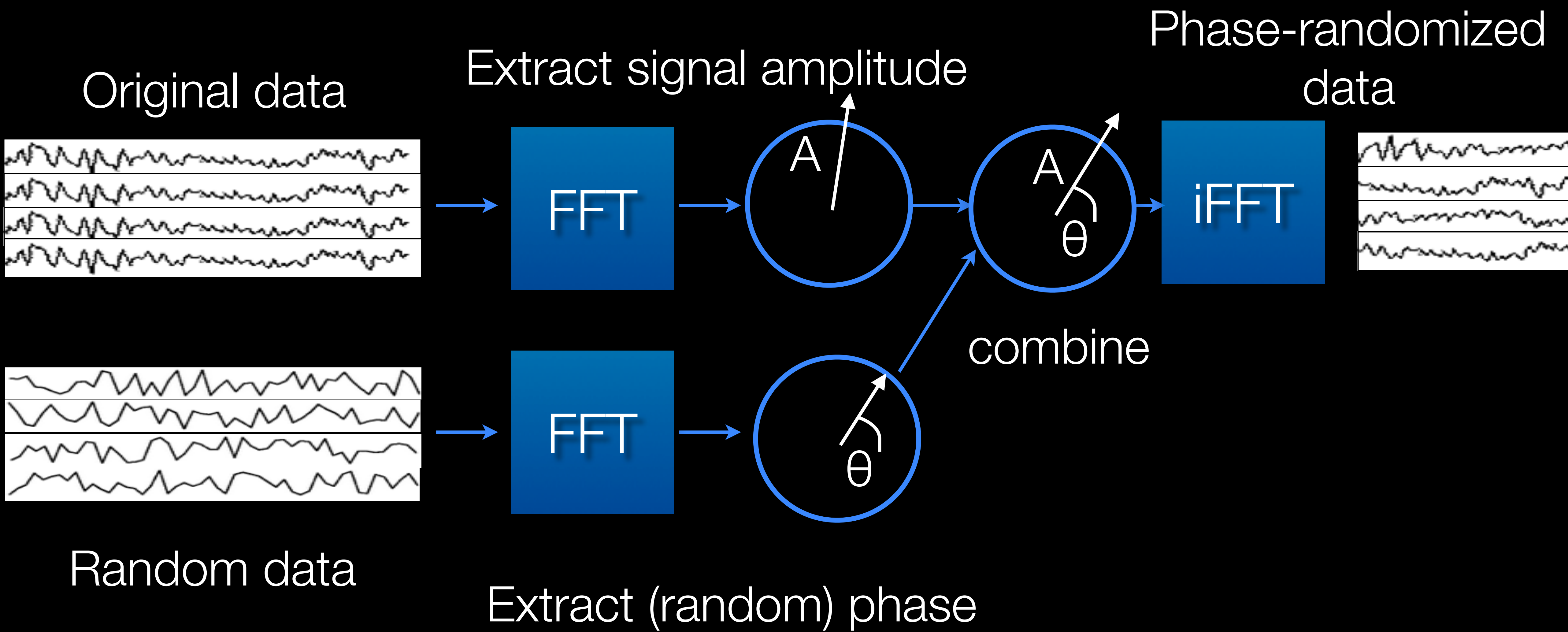


- Difference distribution can take any shape
- The procedure above also provides estimates of the individual distributions of C_A and C_B yielding confidence intervals for H1

Phase-Randomization

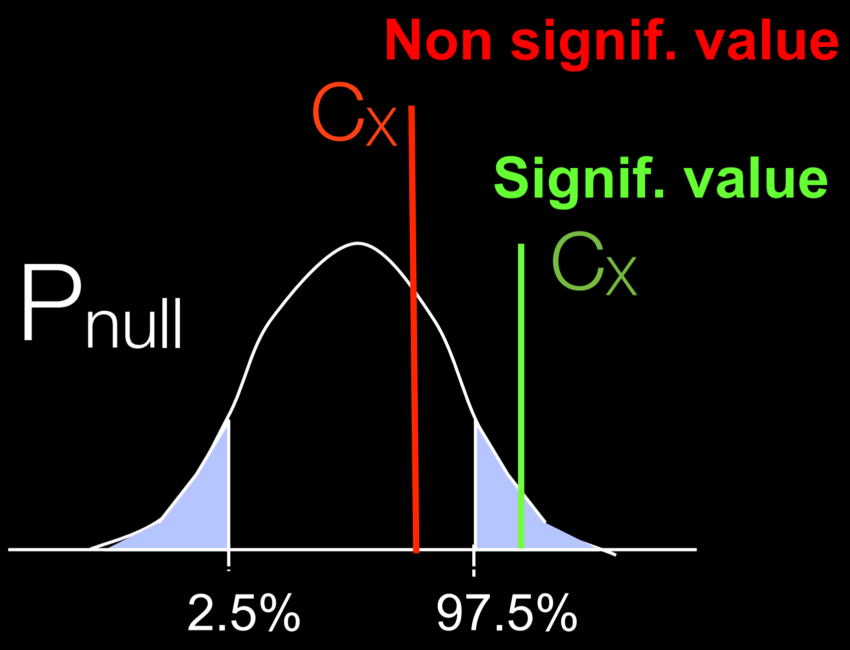
Tim Mullen

- Phase Randomization Procedure (Theiler, 1992)
  - Method for testing whether there is non-zero information flow (H_{null})



Phase-Randomization

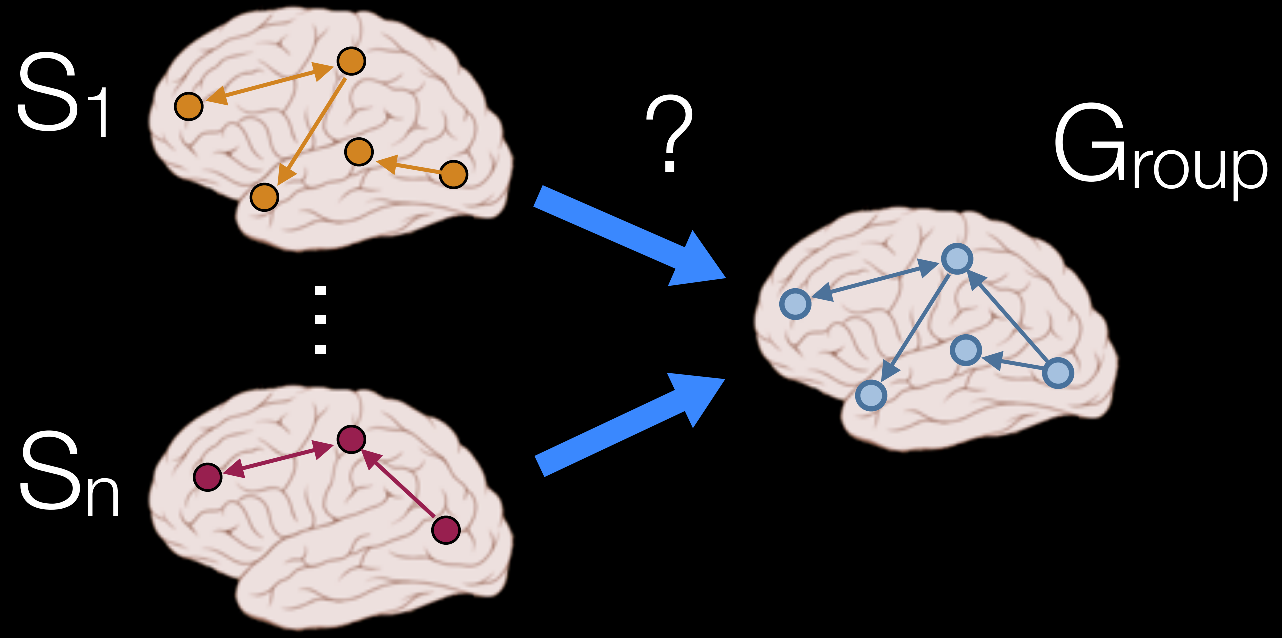
- Start with an n-trial sample: X = {X1, ..., Xn}
- for k=1:R (number of resamples)
  - randomize phases for all trials
  - compute connectivity estimate C_k
  - repeat
- With B large enough the B estimates provide a good approximation of the **null** distribution of the connectivity estimator
- Compare connectivity C_X from original (non-randomized) samples X to quantiles of $P_{null} = \{C_1, ..., C_R\}$



Multi-Subject Inference

- In many cases of source analysis involving focal/point sources (e.g. BSS + dipole fitting, sparse patch-based estimation) we encounter two key problems w.r.t. multi-subject inference:

1. Identification/Co-registration
2. Missing Data



- Conventional approaches utilizing disjoint clustering do not suitably address Issue 2 and generally lack means for rigorously quantifying statistical uncertainty in addressing Issue 1

- Issue 2 is exacerbated in connectivity analysis due to the combinatorial explosion of variables associated with a given source

Tim Mullen

Causal Projection

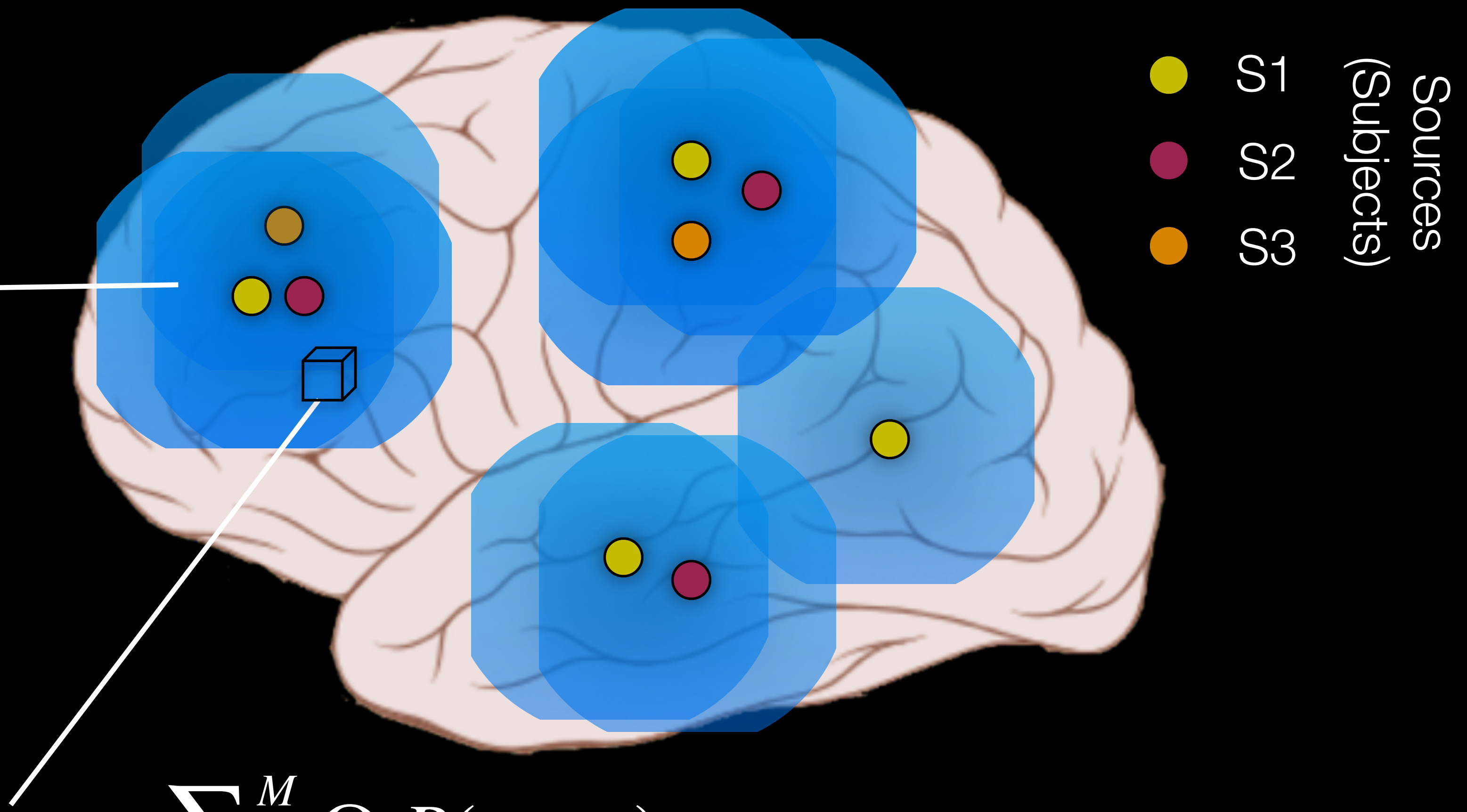
Mullen, et al (2010) *Human Brain Mapping*, Barcelona

- Based on dipole density concept (Delorme and Makeig, 2003)

r_i := 3D coordinate of i^{th} source
r := 3D coordinate of a candidate source location (e.g. reference voxel)
Ω_i := univariate measure (outflow, ERSP, etc) associated with i^{th} source
$\Omega(r)$:= estimated measure value at 3D location r

Gaussian likelihood model

$$P(r = r_i) = N(r \mid r_i, \Sigma_i)$$



$$E[\Omega(r)] = \frac{\sum_{i=1}^{M} \Omega_i P(r = r_i)}{Z}$$

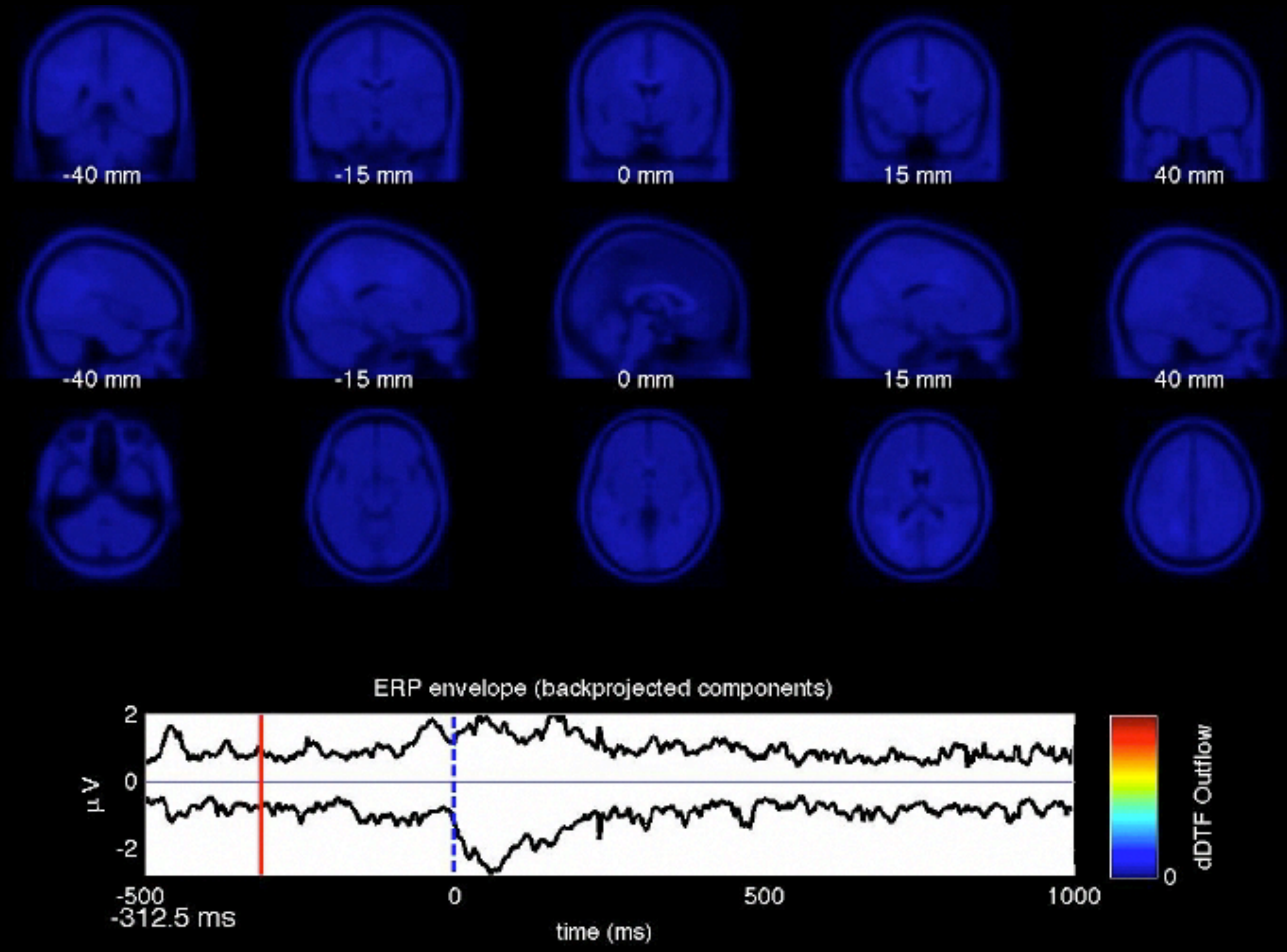
Causal Projection

Tim Mullen

Error > Correct ($p < 0.05$, N=24)

dDTF

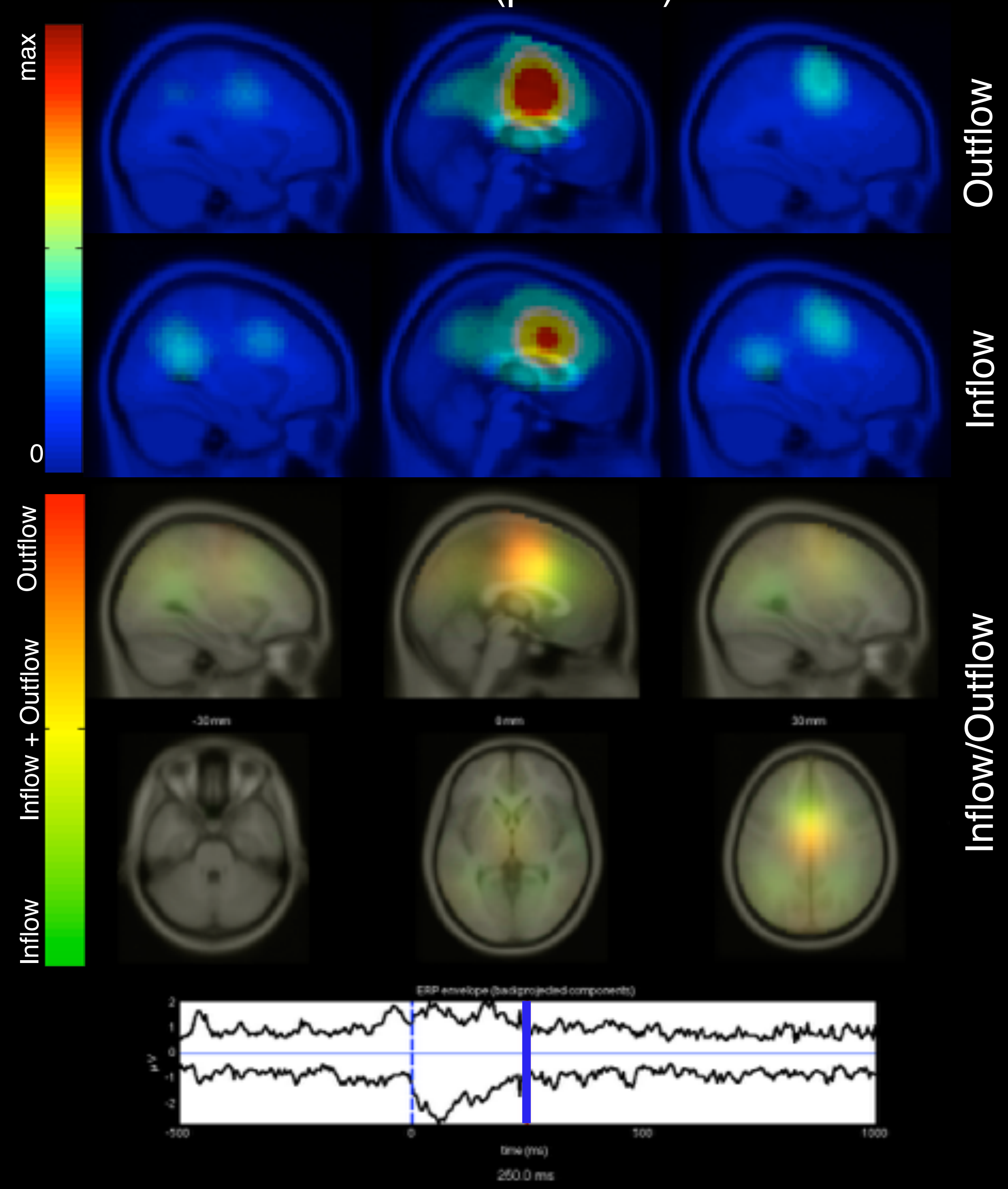
3-7 Hz



Mullen, et al, 2010, *HBM*, Barcelona

Causal Projection

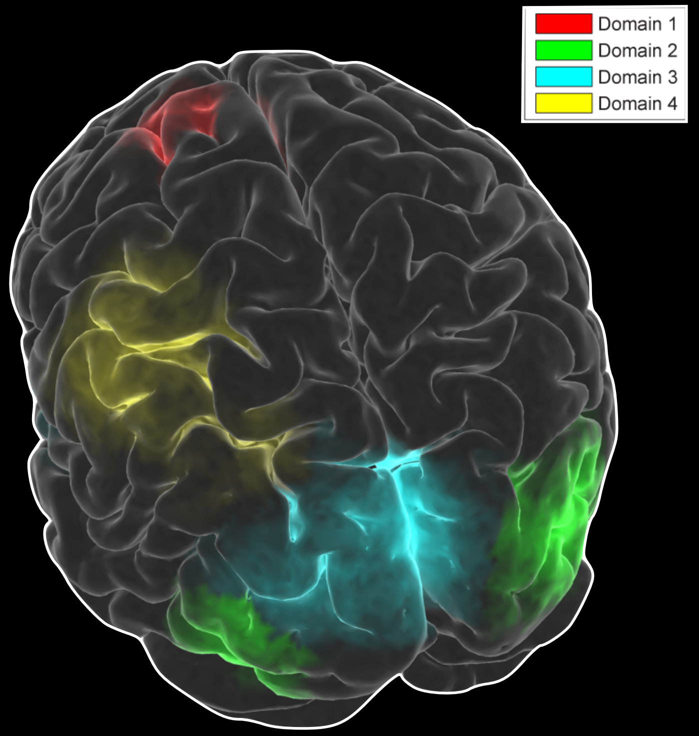
Error > Correct ($p < 0.05$) 3-7 Hz

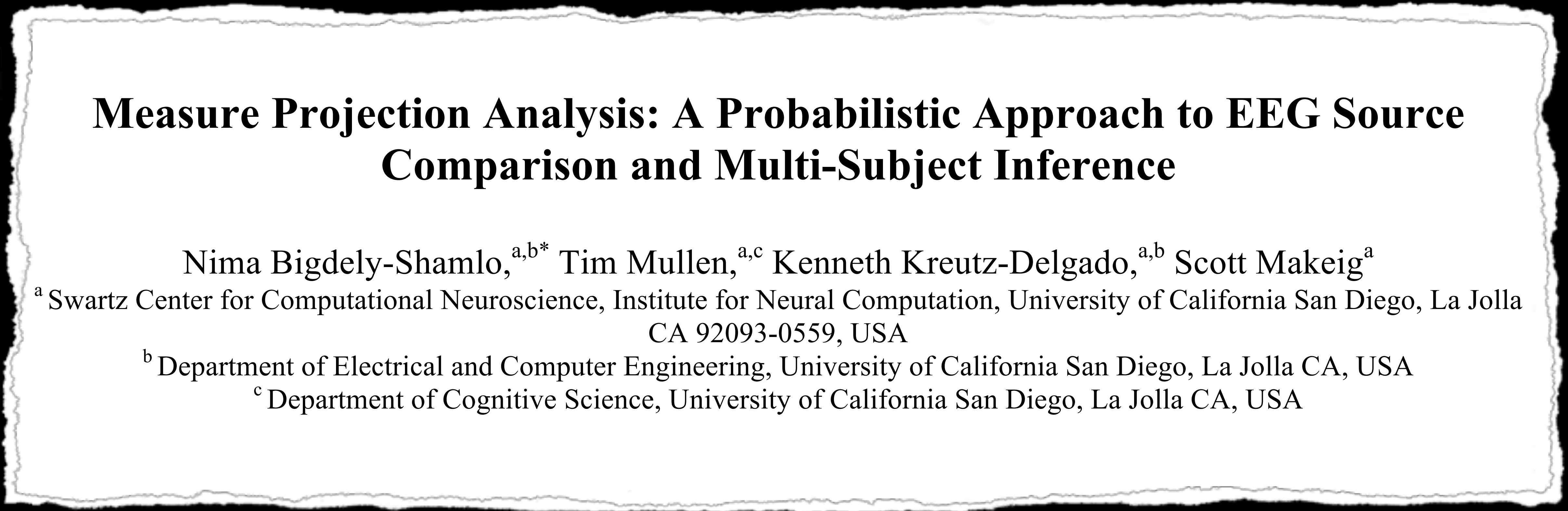


Mullen, et al, 2010, *HBM*, Barcelona

Measure Projection Toolbox

- Similar to dipole density / causal projection, but uses a standardized measure of “convergence” to reduce dimensionality and discount unknown scaling and offsets. Adds clustering on projected measures to find “domains.”
- EEGLAB Toolbox developed by **Nima Bigdely-Shamlo**



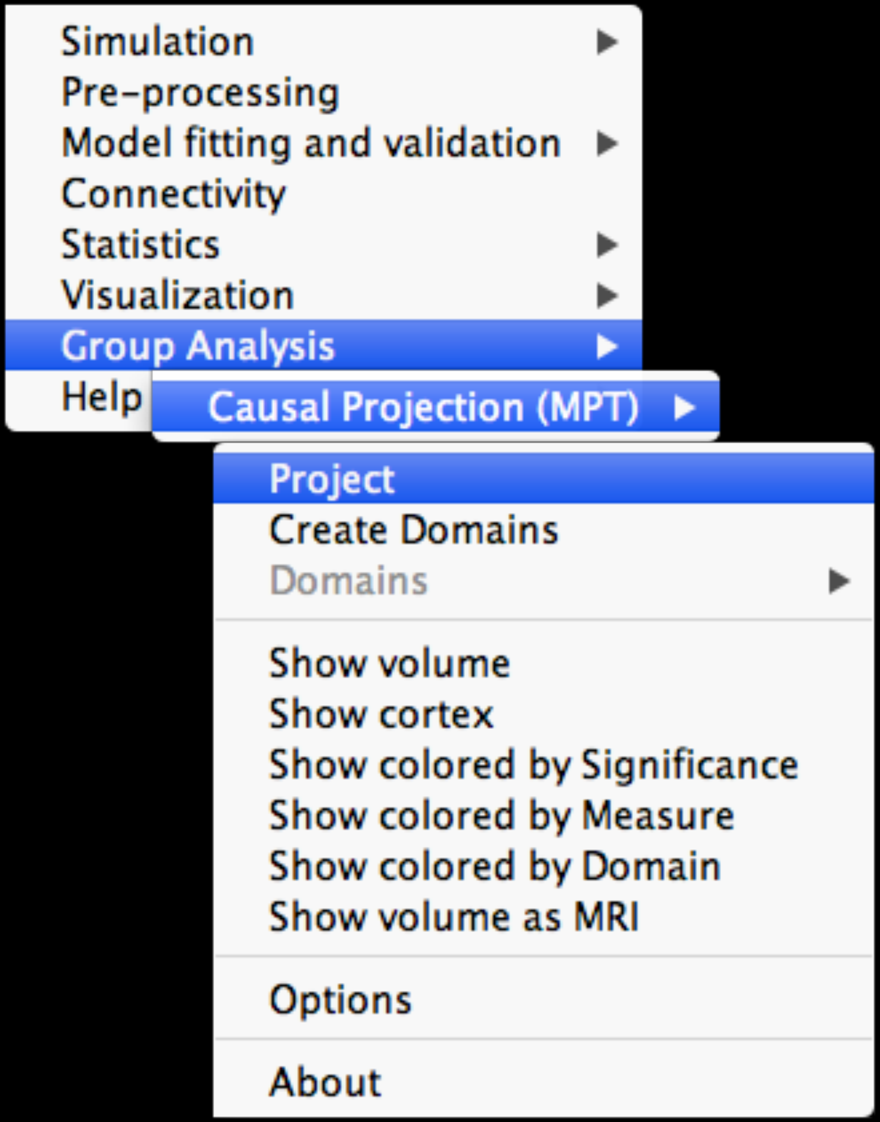


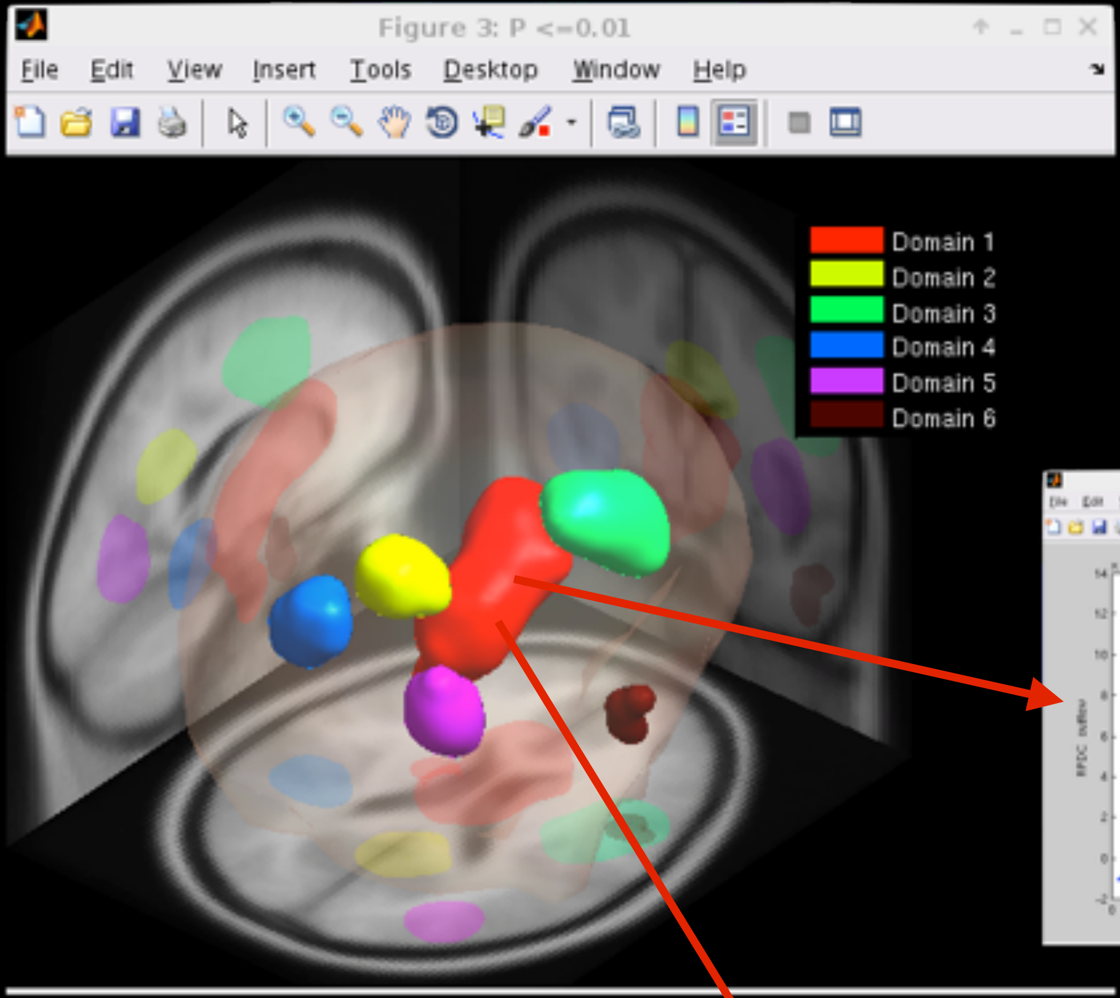
Measure Projection Analysis: A Probabilistic Approach to EEG Source Comparison and Multi-Subject Inference

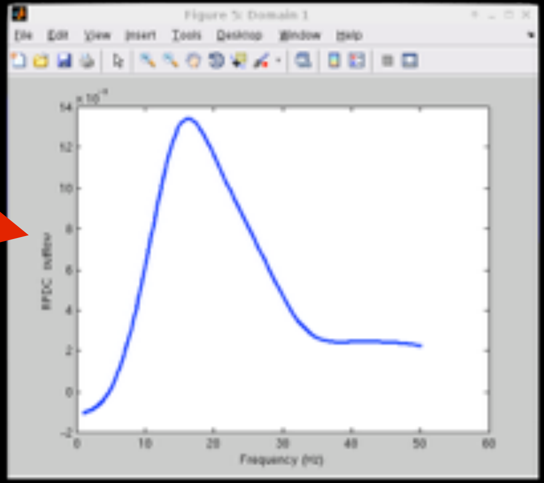
Nima Bigdely-Shamlo,[a,b*] Tim Mullen,[a,c] Kenneth Kreutz-Delgado,[a,b] Scott Makeig[a]
[a] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla CA 92093-0559, USA
[b] Department of Electrical and Computer Engineering, University of California San Diego, La Jolla CA, USA
[c] Department of Cognitive Science, University of California San Diego, La Jolla CA, USA

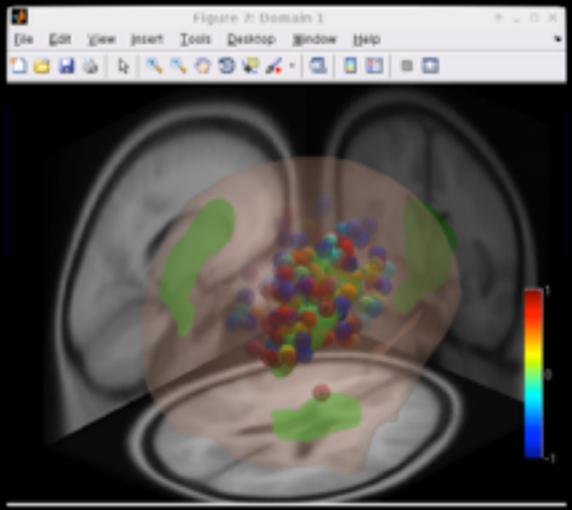
NeuroImage, 2013

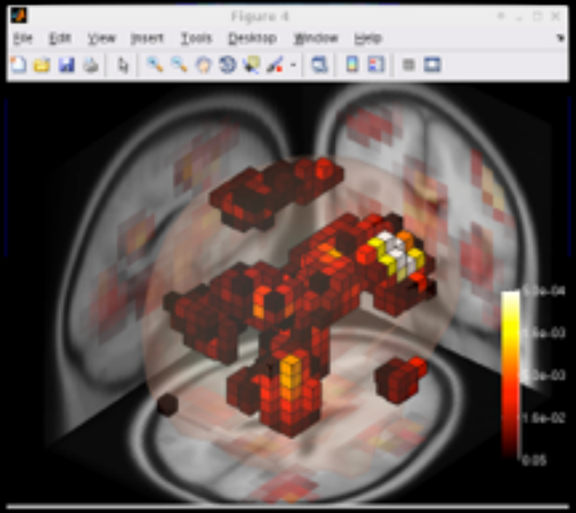
Measure Projection

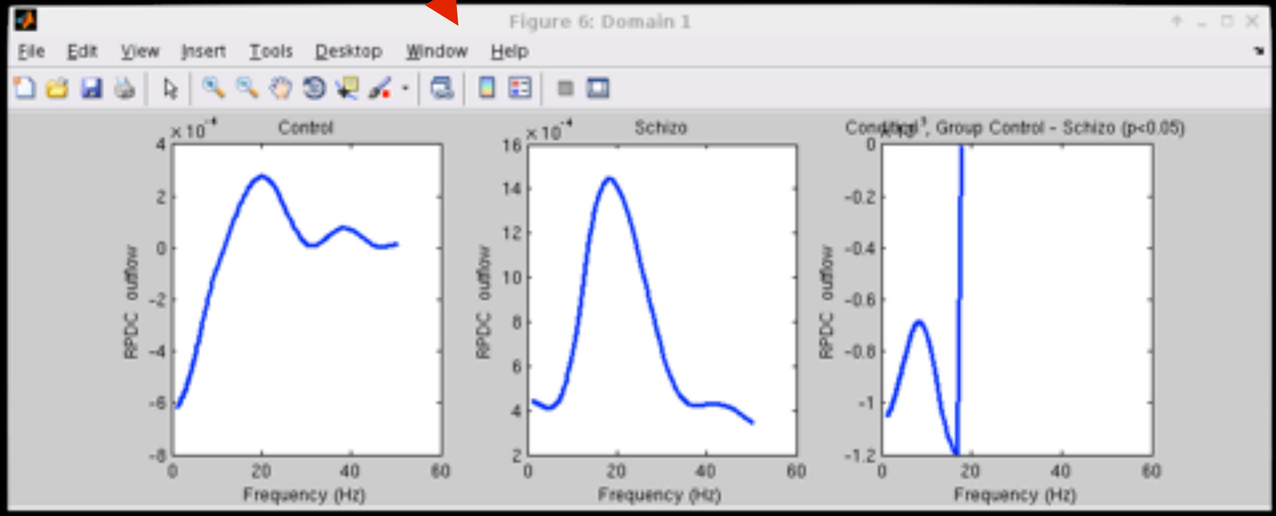








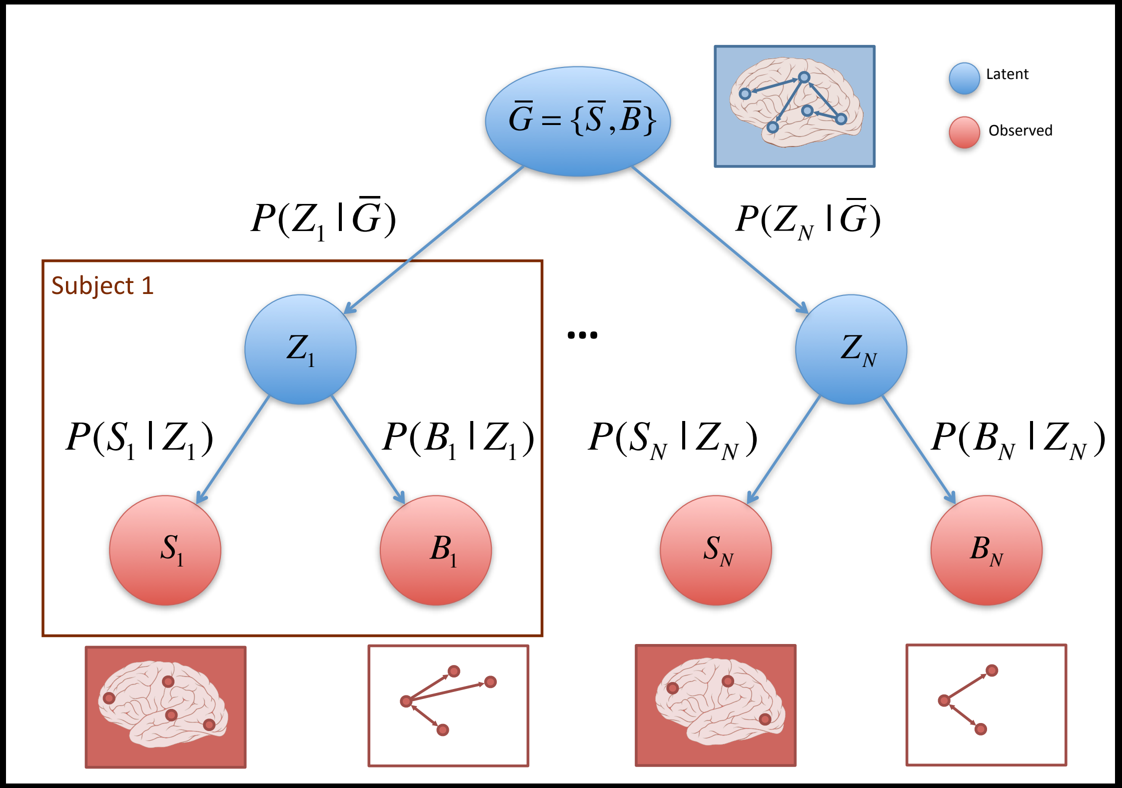




Tim Mullen

Bayesian Hierarchical Model

- Perform multi-subject (second-level) inference via hierarchical (mixture) model, approximating the posterior distribution of source locations and connectivity surfaces.
- Advantages:
  - Handles multivariate measures such as connectivity
  - Yields posterior distributions allowing robust statistics and increased range of hypothesis testing (incl. analysis of individual variability)
  - Hierarchical structure can be adapted to obtain conditional probability distribution w.r.t. other metrics (ERSP, genetics, morphometry, behavior, etc)

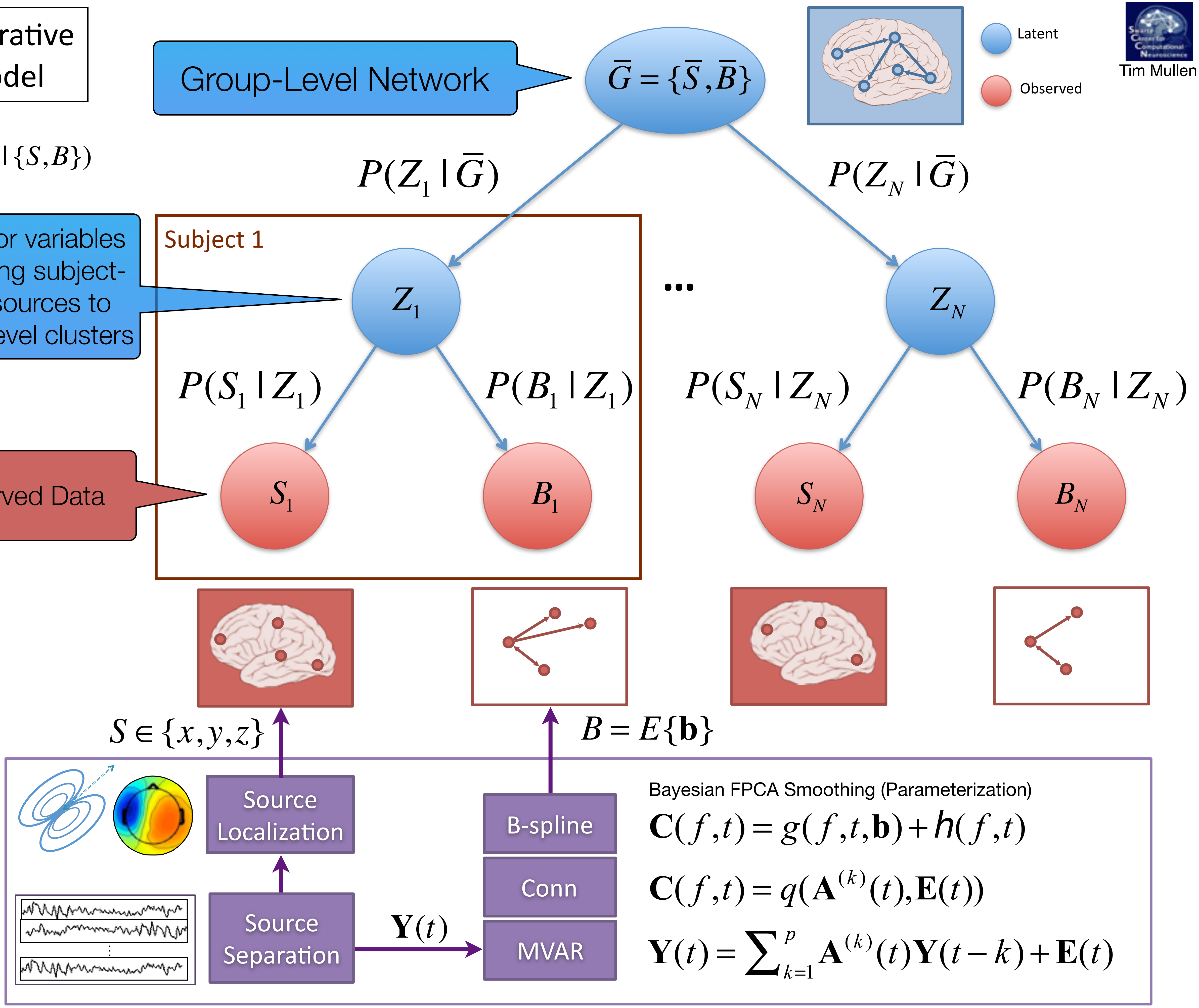


See: Mullen, T. "The Dynamic Brain: Modeling Neural Dynamics and Interactions From Human Electrophysiological Recordings" Chapter 3.

Intro
Theory
SIFT
Apps
To-Do
Fin

Generative Model

Goal:
$P(G,Z \mid \{S,B\})$



Group-Level Network

$\bar{G} = \{\bar{S}, \bar{B}\}$

● Latent
● Observed

Tim Mullen

$P(Z_1 \mid \bar{G})$

$P(Z_N \mid \bar{G})$

Indicator variables assigning subject-level sources to group-level clusters

Subject 1

Z_1

…

Z_N

$P(S_1 \mid Z_1)$

$P(B_1 \mid Z_1)$

$P(S_N \mid Z_N)$

$P(B_N \mid Z_N)$

Observed Data

S_1

B_1

S_N

B_N

$S \in \{x,y,z\}$

$B = E\{\mathbf{b}\}$

Source Localization

Source Separation

$\mathbf{Y}(t)$

B-spline

Conn

MVAR

Bayesian FPCA Smoothing (Parameterization)

$$\mathbf{C}(f,t) = g(f,t,\mathbf{b}) + h(f,t)$$

$$\mathbf{C}(f,t) = q(\mathbf{A}^{(k)}(t), \mathbf{E}(t))$$

$$\mathbf{Y}(t) = \sum_{k=1}^{p} \mathbf{A}^{(k)}(t)\mathbf{Y}(t-k) + \mathbf{E}(t)$$

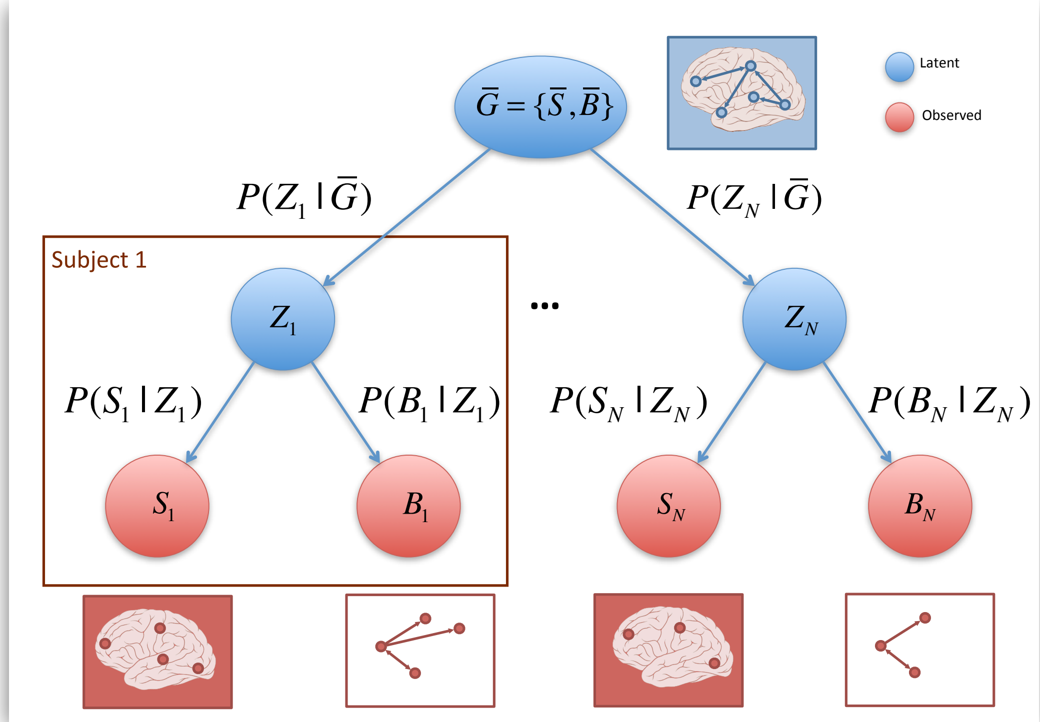
Tim Mullen

MCMC Estimation (Gibbs Sampling)

$$D = \{\{S_i\},\{B_i\},\{\Sigma_s\},\{\boldsymbol{s}_b^2\}\}$$

$$G = \{\bar{S},\bar{B},\Sigma_{\bar{s}},\boldsymbol{s}_{\bar{b}}^2\}$$

$P(G,Z \mid D)$ Target Distribution



Likelihood

$$P(S_i \mid Z_i,G) = \prod_{j=1}^{M_i}\prod_{k=1}^{M}\left[N(s_{ij} \mid \bar{s}_k,\Sigma_{s,k})\right]^{z_{ijk}}$$

$$P(B_i \mid Z_i,G) = \prod_{j_1=1}^{M_i}\prod_{j_2=1}^{M_i}\prod_{k_1=1}^{M}\prod_{k_2=1}^{M}\left[N(b_{ij_1j_2} \mid \bar{b}_{k_1k_2},\boldsymbol{s}_{\bar{b},k_1k_2}^2 I)\right]^{z_{ij_1k_1}z_{ij_2k_2}}$$

Gibbs Sampling Scheme:

1) Initialize G^*, Z^* from initial clustering (e.g. k-means) solution

2) Sample G^* from posterior $P(G^* \mid Z^*,D) \propto P(D \mid Z^*,G^*)P(G^*)$

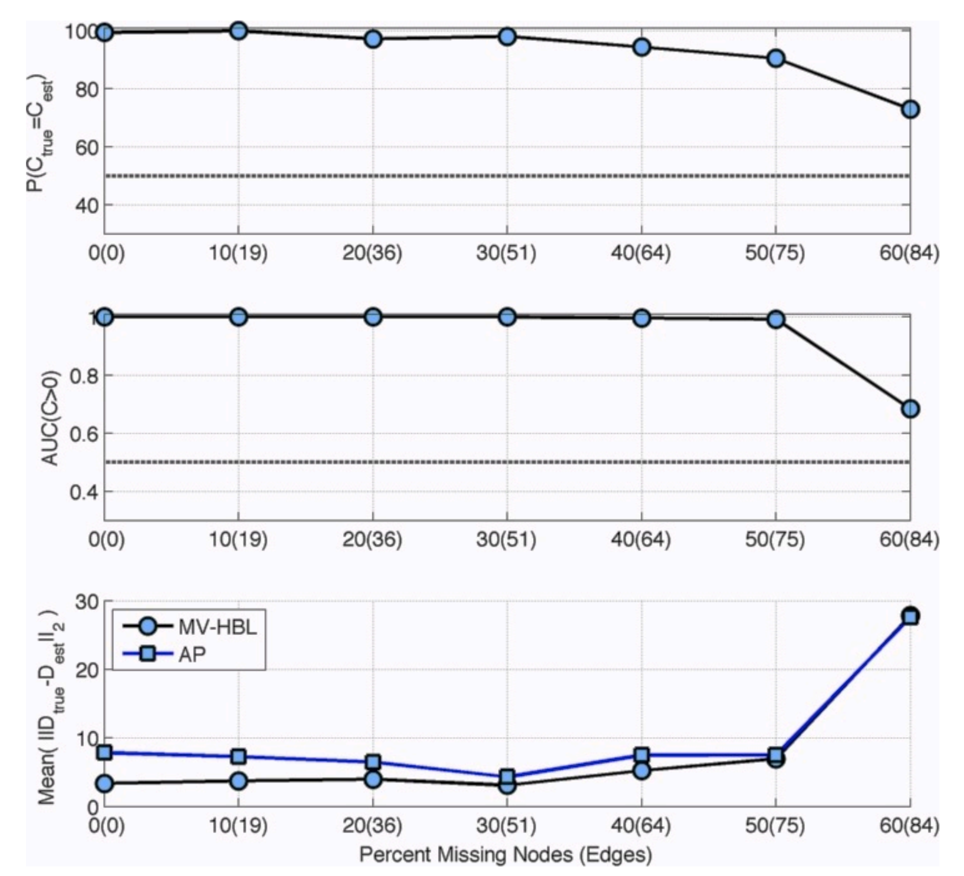
3) Sample Z^* from posterior $P(Z^* \mid G^*,D) \propto P(D \mid Z^*,G^*)P(Z^*)$

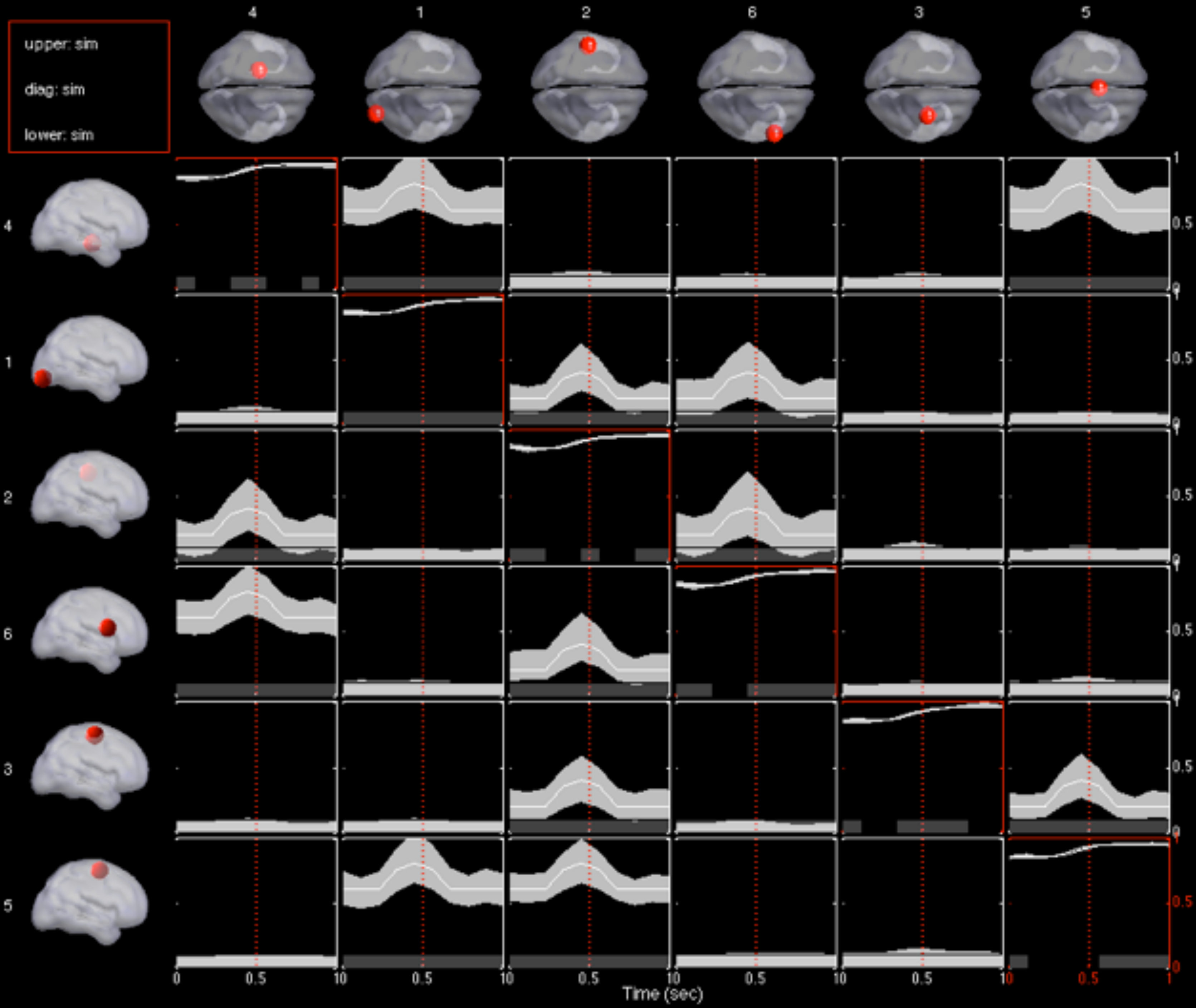
4) Repeat (2-3) many times: $P(G^*,Z^* \mid D)$ → $P(G,Z \mid D)$

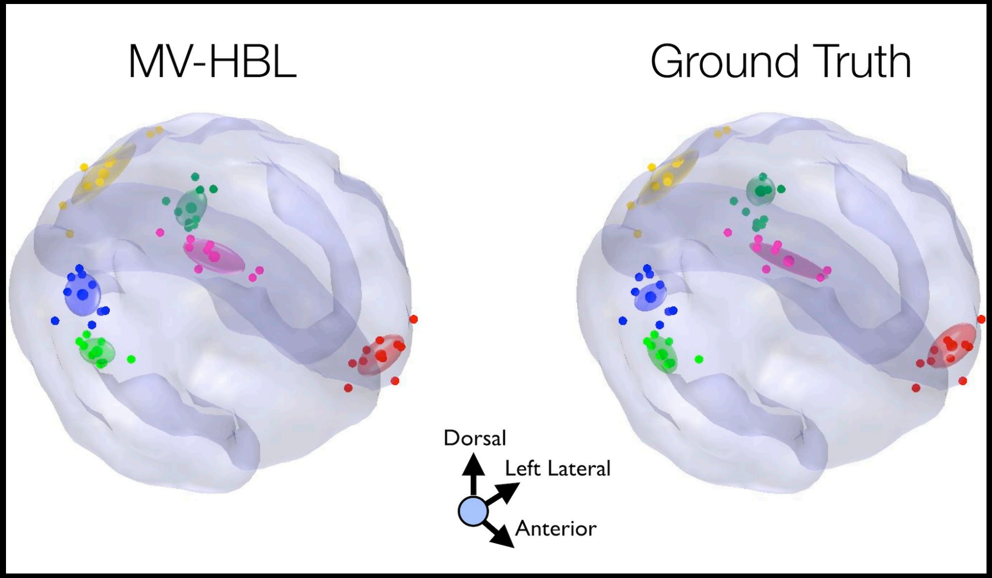
MV-HBL

Sim: 10 Subjects, 30% missing dipoles

Estimated 95% CI (gray) with Ground Truth Superimposed (white)







■ Est = True ($p < 0.05$)

□ Est. 95% CI

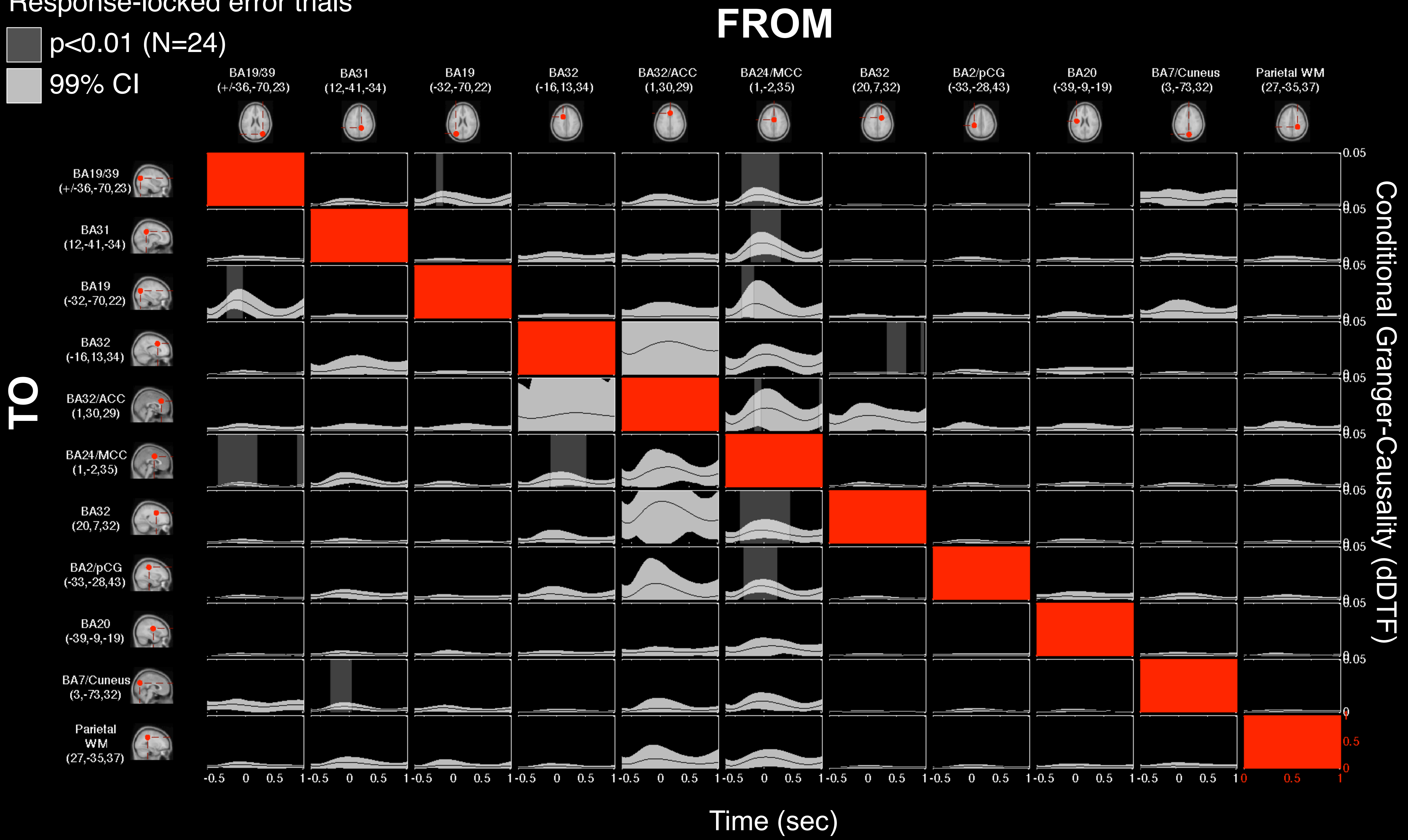
Mullen, T., 2014 "The Dynamic Brain"; Mullen & Thompson, in prep

Bayesian Multi-Subject Inference

Tim Mullen

Theta-band (4-8 Hz) event-related dDTF
Baseline: [-750 -500] ms
Response-locked error trials

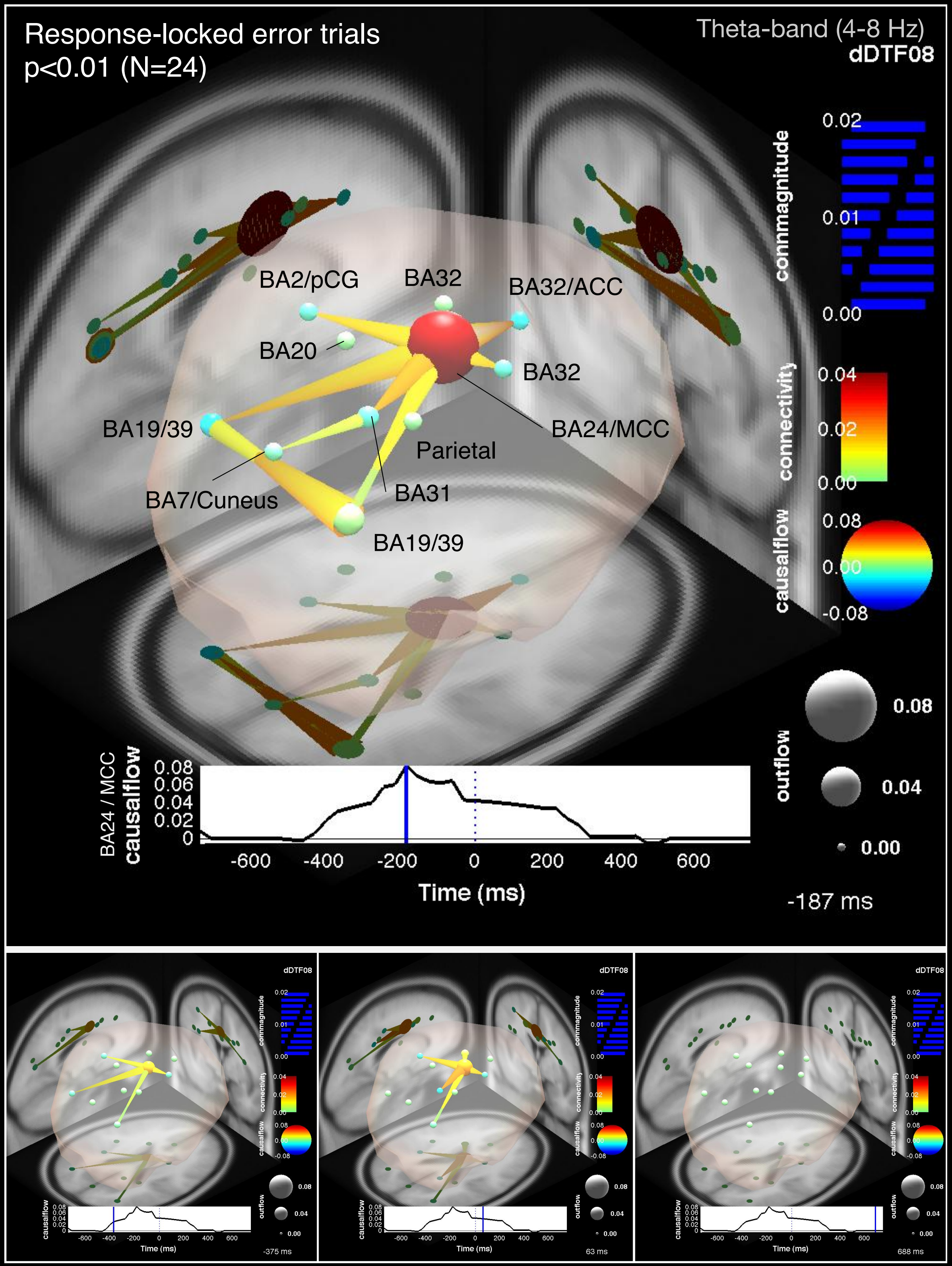
■ p<0.01 (N=24)
□ 99% CI



Cluster retained if more than 33% of subjects have greater than 50% probability of cluster membership

Bayesian Multi-Subject Inference

Tim Mullen



Cluster retained if more than 33% of subjects have greater than 50% probability of cluster membership

Conventional Statistics

- An alternative approach:
  For each subject…

  1. Perform distributed source localization (possibly after separating a subspace of brain components using ICA)

  2. Select *M* regions of interest (ROIs) e.g. from a standardized anatomical atlas (e.g. Desikan-Killiany, Destrieux, etc) and integrate current density within each ROI. This yields *M* source time-series for each subject

  3. Store results in EEG.srcpot

  4. Obtain connectivity estimates for sources using SIFT with the ‘Sources’ option set in pre-processing. Resulting [M x M x N_{freq} x N_{times}] connectivity matrices are stored in EEG.CAT.Conn.

5. Apply your favorite mass-univariate or multivariate statistical approach (e.g. GLM, t-test, (M)ANOVA, etc) to the collection of connectivity estimates from all subjects to obtain desired statistics. See LIMO-EEG Toolbox and EEGLAB’s statcond(). Beware of multiple comparisons issues! FDR may not be suitable.