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

\[ y(t) = Hs(t) + e(t) \]

\[ s(t) = f(s(t^-), u(t^-)) + v(t) \]

Tim Mullen
Swartz Center for Computational Neuroscience
Institute for Neural Computation
Dept. of Cognitive Science
UC San Diego
Review

<table>
<thead>
<tr>
<th>Theoretical Foundations I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Connectivity Measures (PLV, PAC, Coherence)</td>
</tr>
<tr>
<td>Effective Connectivity Measures and Granger Causality</td>
</tr>
<tr>
<td>Scalp versus Source</td>
</tr>
<tr>
<td>Adapting to Time-Varying Dynamics</td>
</tr>
</tbody>
</table>

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

\[ X(t) = \sum_{k=1}^{p} A^{(k)}(t)X(t-k) + E(t) \]
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 the same correlation structure as the real data
  - **Stability Test**: checking the stability/stationarity of the model.
Whiteness Tests

- We can regard the VAR[p] model coefficients $A^{(k)}$ as a filter which transforms random (white) noise $E(t)$, into observed, structured data $X(t)$:
  \[
  X(t) = f(L)E(t), \quad f(L) = \left( I - \sum_{k=1}^{p} A^{(k)} L^k \right)^{-1}, \quad L \text{ is a “lag operator”}
  \]
  \[
  L^k Z(t) = Z(t - k)
  \]
- For coefficient estimates $\hat{A}^{(k)}$, we can obtain the residuals
  \[
  \hat{E}(t) = X(t) - \sum_{k=1}^{p} \hat{A}^{(k)}(t)X(t - k)
  \]
- 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.
Whiteness Tests

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

\[ C_l = \langle \hat{\mathbf{E}}(t)\hat{\mathbf{E}}'(t-l) \rangle \]

\[ R_l = D^{-1}C_lD^{-1} \]

\[ D = \text{diag}\left(\sqrt{\text{diag}(C_0)}\right) \]

\[ R_h = (R_1, \ldots, R_h) \]

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
Whiteness Tests: ACF

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

$$\rho = \frac{\text{count}\left(|R_h| > 2/\sqrt{T}\right)}{\text{count}(R_h)} = \frac{\text{count}\left(|R_h| > 2/\sqrt{T}\right)}{M^2(h+1) - M}$$

If $\rho<0.05 (1-\rho > .95)$ then we cannot reject $H_0$ at the 5\% level and we accept that residuals $\hat{E}(t)$ 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
Whiteness Tests: Portmanteau

<table>
<thead>
<tr>
<th>Portmanteau Test</th>
<th>Formula (Test Statistic)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box-Pierce (BPP)</td>
<td>$Q_h := \hat{T} \sum_{l=1}^{h} \text{tr} \left( C_l' C_0^{-1} C_l C_0^{-1} \right)$</td>
<td>The original portmanteau test. Potentially overly-conservative. Poor small-sample properties.</td>
</tr>
<tr>
<td>Ljung-Box (LBP)</td>
<td>$Q_h := \hat{T} (\hat{T} + 2) \sum_{l=1}^{h} (\hat{T} - l)^{-1} \text{tr} \left( C_l' C_0^{-1} C_l C_0^{-1} \right)$</td>
<td>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.</td>
</tr>
<tr>
<td>Li-McLeod (LMP)</td>
<td>$Q_h := \hat{T} \sum_{l=1}^{h} \text{tr} \left( C_l' C_0^{-1} C_l C_0^{-1} \right) + \frac{M^2 h(h+1)}{2\hat{T}}$</td>
<td>Further modification of BPP to improve small-sample properties without variance inflation. Slightly more conservative than LBP. Probably the best choice in most conditions.</td>
</tr>
</tbody>
</table>

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

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.
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 ($R_s$) and real data ($R_r$).

The percent consistency (PC) is then given by

$$PC = \left(1 - \frac{\|R_s - R_r\|_2}{\|R_r\|_2}\right) \times 100$$

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 \( \leq 1 \)
  - 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*)
\[
\begin{align*}
X(t) &= \sum_{k=1}^{p} A^{(k)}(t)X(t-k) + E(t) \\
A(f,t) &= -\sum_{k=0}^{p} A^{(k)}(t)e^{-j2\pi fk}; \ A^{(0)} = I \\
X(f,t) &= A(f,t)^{-1}E(f,t) = H(f,t)E(f,t)
\end{align*}
\]

**Functional**

- **Spectrum (Brillinger, 2001)**
  \[ S(f) = X(f)X(f)^* = H(f)\Sigma H(f)^* \]

- **Coherence**
  \[ C_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)S_{jj}(f)}} \]  
  (Brillinger, 2001)

- **Partial Coherence**
  \[ P_{ij}(f) = \frac{S_{ij}^{-1}(f)}{\sqrt{S_{ii}^{-1}(f)S_{jj}^{-1}(f)}} \]  
  (Brillinger, 2001)

- **Multiple Coherence**
  \[ G_i(f) = \sqrt{1 - \frac{\det(S(f))}{S_{ii}(f)M_{ii}(f)}} \]  
  (Brillinger, 2001)

**Effective**

- **Granger-Geweke Causality**
  \[ F_{ij}(f) = \frac{\Sigma_{ij} - (\Sigma_{ij}^2 / \Sigma_{ii})\Sigma_{ij}(f)H_{ij}(f)^2}{S_{ii}(f)} \]  
  (Geweke, 1982; Bressler et al., 2007)

- **Directed Transfer Function (Kaminski and Blinowska, 1991)**
  \[ \eta_{ij}^2(f) = \frac{|H_{ij}(f)|^2}{\sum_f \sum_{k=1}^{M} |H_{ik}(f)|^2} \]

- **Directed Coherence (Baccalá and Sameshima, 2001)**
  \[ \pi_{ij}^2(f) = \frac{|A_{ij}(f)|^2}{\sum_{k=1}^{M} |A_{kj}(f)|^2} \]

- **Direct DTF (Korzeniewska, 2003)**
  \[ \delta_{ij}^2(f) = \eta_{ij}^2(f)P_{ij}^2(f) \]
<table>
<thead>
<tr>
<th>Estimator</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral Density Matrix</td>
<td>( S(f) = X(f)X(f)^* = H(f)\Sigma H(f)^* )</td>
</tr>
<tr>
<td>Coherency</td>
<td>( C_\alpha(f) = \frac{S_\alpha(f)}{\sqrt{S_a(f)S_\beta(f)}} ) ( 0 \leq</td>
</tr>
<tr>
<td>Imaginary Coherence (iCoh)</td>
<td>( i\text{Coh}<em>\beta(f) = \text{Im}(C</em>\beta(f)) )</td>
</tr>
<tr>
<td>Partial Coherence (pCoh)</td>
<td>( P_\alpha(f) = \frac{\hat{S}<em>\alpha(f)}{\sqrt{\hat{S}</em>\alpha(f)\hat{S}_\beta(f)}} ) ( \hat{S}(f) = S(f)^{-1} ) ( 0 \leq</td>
</tr>
<tr>
<td>Multiple Coherence (mCoh)</td>
<td>( G_\beta(f) = \sqrt{1 - \frac{\text{det}(S(f))}{S_\beta(f)M_\beta(f)}} ) ( M_\beta(f) ) is the minor of ( S(f) ) obtained by removing the ( i^{th} ) row and column of ( S(f) ) and returning the determinant.</td>
</tr>
<tr>
<td>Normalized Partial Directed Coherence (PDC)</td>
<td>( \pi_\alpha(f) = \frac{A_\alpha(f)}{\sqrt{\sum_k A_k(f)\Sigma_k}} ) ( 0 \leq</td>
</tr>
<tr>
<td>Generalized Partial Directed Coherence (GPDC)</td>
<td>( \bar{\pi}_\alpha(f) = \frac{1}{\sum_k A_k(f)} ) ( 0 \leq</td>
</tr>
<tr>
<td>Granger-Geweke Causality (GGC)</td>
<td>( \lambda_\alpha(f) = Q_\alpha(f)^*V_\alpha(f)^{-1}Q_\alpha(f) ) where ( Q_\alpha(f) = \begin{pmatrix} \text{Re}[A_\alpha(f)] \ \text{Im}[A_\alpha(f)] \end{pmatrix} ) and ( V_\alpha(f) = \sum_{k,p}R_{\alpha\beta}(k,p)\Sigma_{k,p}Z(2\pi f,k,p) ) ( Z(\alpha,k,l) = \begin{pmatrix} \cos(\alpha k)\cos(\alpha l) &amp; \cos(\alpha k)\sin(\alpha l) \ \sin(\alpha k)\cos(\alpha l) &amp; \sin(\alpha k)\sin(\alpha l) \end{pmatrix} ) ( R ) is the ([M_p]^2 \times [M_p]^2) covariance matrix of the VAR([p]) process (Lütkepohl, 2006)</td>
</tr>
<tr>
<td>Renormalized Partial Directed Coherence (rPDC)</td>
<td>( \lambda_\alpha(f) = Q_\alpha(f)^*V_\alpha(f)^{-1}Q_\alpha(f) ) where ( Q_\alpha(f) = \begin{pmatrix} \text{Re}[A_\alpha(f)] \ \text{Im}[A_\alpha(f)] \end{pmatrix} ) and ( V_\alpha(f) = \sum_{k,p}R_{\alpha\beta}(k,p)\Sigma_{k,p}Z(2\pi f,k,p) ) ( Z(\alpha,k,l) = \begin{pmatrix} \cos(\alpha k)\cos(\alpha l) &amp; \cos(\alpha k)\sin(\alpha l) \ \sin(\alpha k)\cos(\alpha l) &amp; \sin(\alpha k)\sin(\alpha l) \end{pmatrix} ) ( R ) is the ([M_p]^2 \times [M_p]^2) covariance matrix of the VAR([p]) process (Lütkepohl, 2006)</td>
</tr>
<tr>
<td>Directed Transfer Function (dDTF)</td>
<td>( \eta_\alpha(f) = \frac{</td>
</tr>
<tr>
<td>Full-Frequency Directed Transfer Function (FDTF)</td>
<td>( \eta_\alpha(f) = \frac{</td>
</tr>
<tr>
<td>Direct Transfer Function (DTF)</td>
<td>( \delta_\alpha(f) = \eta_\alpha(f)P_\alpha^2(f) )</td>
</tr>
</tbody>
</table>

\[ X(t) = \sum_{k=1}^{p} A^{(k)}(t)X(t-k) + E(t) \]
\[ A(f,t) = -\sum_{k=0}^{p} A^{(k)}(t)e^{-i2\pi ft}; \quad A^{(0)} = I \]
\[ X(f,t) = A(f,t)^{-1}E(f,t) = H(f,t)E(f,t) \]

**H(f)** Transfer Function

**A(f)** System Matrix

**\( \Sigma \)** Noise Covariance Matrix

**Variance Stabilization**

For additional details, see SIFT Handbook (scn.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.

\[
N = O(M^2p)
\]

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

Ensemble aggregation (\( T > 1 \))?

- \( M=20, p=10, T=50: 4000/50 \) samples/trial \( \rightarrow 20/50 = 0.4 \) 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, ...).
Multivariate Models: Constraints

\[ \hat{A} = \arg \max_A \left\{ p(A|D) \equiv p(D|A)p(A) \right\} \]

- **prior**
  - Unconstrained (all values equally probable). E.g. Uniform distribution
  \[ p(A) = U(a,b) \]

- **likelihood**
  - Smoothness constraints
    - large differences in values unlikely
    - small (non-zero) values most probable. E.g. Normal (gaussian) prior.
  \[ p(A) = N(0, \Sigma) \]

- **posterior**
  - Sparsity constraint
    - many values small or exactly zero with occasional large values e.g. Laplacian prior
  \[ p(A) = L(0, \beta) \]
Smoothness Constraints

- Standard least-squares solution

\[
A(t) = \arg\min_{\tilde{A}} \left\| Y - Z\tilde{A} \right\|_2^2
\]

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

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

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

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

\[
Y = X_0
\]

Rewrite VAR[p] as VAR[1]
Smoothness Constraints

- Ridge Regression
  (Tikhonov Regularization, Minimum-(L$_2$)-Norm Estimation, ...)

\[ A(t) = \arg\min_{\hat{A}} \left( \| Y - Z\hat{A} \|_2^2 + \lambda \| \hat{A} \|_2^2 \right) \]

- Equivalent to assuming a Gaussian prior with variance determined by $\lambda$
- 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.
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


Structural Connectivity
Standard least-squares solution

\[ A(t) = \arg \min_{\tilde{A}} \left( \| Y - Z\tilde{A} \|^2 \right) \]

\[ X(t) = \sum_{k=1}^{p} A^{(k)}(t)X(t - k) + E(t) \]
\[ \tilde{A} = [A^{(1)}(t), \ldots, A^{(p)}(t)]^T \]
\[ X_k = [X(p + 1 - k), \ldots, X(N - k)]^T \]
\[ Z = [X_1, \ldots, X_p] \]
\[ Y = X_0 \]

Rewrite VAR[p] as VAR[1]
Sparsity Constraints

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

\[
A(t) = \arg \min_{\hat{A}} \left( \|Y - Z\hat{A}\|_2^2 + \lambda \sum_{ij} \|\hat{A}_{ij}^{(1)}, \ldots, \hat{A}_{ij}^{(p)}\|_2 \right)
\]

- 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.
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?

- \( N = \text{number of samples required} \)
- \( M = \text{number of variables/sources, } p = \text{model order} \)

\[
N = O\left( K \log(\frac{M^2 p}{K}) \right) \approx O\left( \log M^2 p \right)
\]

\( N = O(M^2 p) \) (unconstrained)
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
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

- 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)

**Observation equation**
(e.g. noisy sensor observations)

\[ y(t) = Hs(t) + \epsilon(t) \]

**State transition equation**
(e.g. latent source and/or coupling dynamics)

\[ s(t) = f(s(t^-), u(t^-), \theta(t)) + v(t) \]

**Linear VAR[1]**

\[ s(t) = A(t)s(t-1) + v(t) \]
Kalman Filtering

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

\[ f(z,u) \text{ is linear in } \{z,u\} \text{ for classic Kalman Filtering} \]

\[ \hat{y}(t) \]

new observation

\[ y_t \quad u_t \]

Time Update (“Predict”)

\[ \hat{z}_t = f(\hat{z}_{t-1}, u_{t-1}) + v_t \]

\[ \hat{y}_t = H\hat{z}_t \]

\[ P_t^- = AP_{t-1}A^T + Q \]

Measurement Update (“Correct”)

\[ \epsilon = y_t - \hat{y}_t \]

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

\[ P_t = (I - G_t H) P_t^- \]

updated model

Initialize

\[ \hat{z}_0 = E(z_0) \]

\[ P_0 = \text{cov}(z_0) \]

\[ t \to t + 1 \]

unknown state vector at time \( t \)
e.g. delay-embedding of sources and/or coupling (VAR) parameters
Kalman Filtering

GPDC Causality From

To-Do

Causality To

Frequency (Hz)

Time (sec)
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

van-der-Pol coupled oscillators

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

Stochastic coupled Rössler oscillators

\[ \dot{\xi} = \begin{pmatrix} \dot{X}_1 \\ \dot{Y}_1 \\ \dot{Z}_1 \end{pmatrix} = \begin{pmatrix} -\alpha_1 Y_1 - Z_1 + \sum_{i \neq j} \epsilon_{ij}(X_i - X_j) \\ \omega_1 X_1 + \beta Y_1 \\ b + (X_1 - c)Z_1 \end{pmatrix} + \sigma_\eta \]

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

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 3 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.
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**

- 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
- Nonlinear Kalman Filter: est_fitMVARMVAR_DEKF.m
Statistics

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

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

$C(i,j)$ is the measured information flow from process $j \rightarrow i$.
$C_{null}$ is the expected measured information flow when there is no true information flow.
$C_{base}$ is the expected information flow in some baseline period.
## Statistical Approach

<table>
<thead>
<tr>
<th>Method</th>
<th>Tests</th>
<th>Parametric</th>
<th>Nonparam.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic analytic estimates of confidence intervals.</td>
<td>$H_{null}$, $H_{base}$, $H_{AB}$</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Applies to: PDC, nPDC, DTF, nDTF, rPDC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theiler phase randomization</td>
<td>$H_{null}$</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Applies to: all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bootstrap, Jacknife, Cross-Validation</td>
<td>$H_{AB}$, $H_{base}$</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Applies to: all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence intervals using Bayesian smoothing splines</td>
<td>$H_{base}$, $H_{AB}$</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Applies to: all</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$H_{null}: C_{ij} = 0$  $H_{base}: C_{ij} = C_{baseline}$  $H_{AB}: C^{A}_{ij} = C^{B}_{ij}$
Bootstrap Statistics

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

% self-awareness data, Wilcox, 2005, p58

Sample with replacement b times
compute estimate

Bootstrapped estimates

get PDF

Distribution of bootstrapped estimates of the mean

Sorting & get CI
Bootstrap Statistics

Bootstrap

sorted values (cdf)

thresholds (ci)

2.5%

97.5%
Suppose we have two conditions
A = \{a_1, \ldots, a_7\}
B = \{b_1, \ldots, b_6\}
We want to estimate the distributions of connectivity estimator
applied to A and B separately, as well as the difference
distribution (for testing H_0: A = B)
Bootstrap Differences

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

- 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

- 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)

\[
P_D
\]

- 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

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

Phase-Randomization Procedure:

1. Original data
2. Extract signal amplitude
3. $A\theta$
4. Phase-randomized data
5. Combine
6. iFFT
7. Random data
8. Extract (random) phase
9. FFT
10. FFT
11. iFFT
12. FFT
13. iFFT

Phase-Randomized data

Method for testing whether there is non-zero information flow ($H_{null}$)
Phase-Randomization

- Start with an n-trial sample: \( X = \{X_1, \ldots, X_n\} \)
- 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 
  \( P_{\text{null}} = \{C_1, \ldots, C_R\} \)
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
Causal Projection


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

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

Gaussian noise 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

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
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)

Thompson, Mullen, Makeig, 2011, ICONXI
Thompson, Mullen, Makeig, 2012, in prep
Subject $1$

$G = \{S, B\}$

$P(Z_1 | G)$

$P(Z_N | G)$

$P(S_1 | Z_1)$

$P(B_1 | Z_1)$

$P(S_N | Z_N)$

$P(B_N | Z_N)$

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

$C(f, t) = g(f, t, b) + \eta(f, t)$

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

$B = E\{b\}$
MCMC Estimation (Gibbs Sampling)

\[ D = \{\{S_i\}, \{B_i\}, \{\Sigma_s\}, \{\sigma_b^2\}\} \]

\[ G = \{\overline{S}, \overline{B}, \Sigma_{\overline{s}}, \sigma_b^2\} \]

**Target Distribution**

\[ P(G,Z \mid D) \]

**Likelihood**

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

\[ P(B_i \mid Z_i, G) = \prod_{j_1=1}^{M_i} \prod_{j_2=1}^{M} \prod_{k_1=1}^{M} \prod_{k_2=1}^{M} N(b_{ij_1j_2} \mid \overline{b}_{k_1k_2}, \sigma_{b,k_1k_2}^2) \]

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: \(P(G^*, Z^* \mid D) \rightarrow P(G, Z \mid D)\)
### Bayesian Multi-Subject Inference

#### Theta-band (4-8 Hz) event-related dDTF

- **Baseline:** [-750 -500] ms
- **Response-locked error trials**

<table>
<thead>
<tr>
<th>FROM</th>
<th>TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA19/SLO (+9670,22)</td>
<td>BA19/SLO (+9670,22)</td>
</tr>
<tr>
<td>BA31 (124134)</td>
<td>BA31 (124134)</td>
</tr>
<tr>
<td>BA10 (+527022)</td>
<td>BA10 (+527022)</td>
</tr>
<tr>
<td>BA32 (+16,13,34)</td>
<td>BA32 (+16,13,34)</td>
</tr>
<tr>
<td>BA32/ACC (123029)</td>
<td>BA32/ACC (123029)</td>
</tr>
<tr>
<td>BA24/MCC (3552,45)</td>
<td>BA24/MCC (3552,45)</td>
</tr>
<tr>
<td>BA29 (39819)</td>
<td>BA29 (39819)</td>
</tr>
<tr>
<td>BA2/Cuneus (373532)</td>
<td>BA2/Cuneus (373532)</td>
</tr>
<tr>
<td>Partial WM (275557)</td>
<td>Partial WM (275557)</td>
</tr>
</tbody>
</table>

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

---

**Thompson, Mullen, Makeig, 2011, ICONXI**

**Thompson, Mullen, Makeig, 2012, in prep**
Bayesian Multi-Subject Inference

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

Response-locked error trials
p<0.01 (N=24)

Theta-band (4-8 Hz) dDTF08

Clustered dipoles in the dorsal middle cingulate cortex (BA24/MCC) and bilateral prefrontal cortex (BA11)

Thompson, Mullen, Makeig, 2011, ICONXI
Thompson, Mullen, Makeig, 2012, in prep