Bayesian Population Modeling of Effective Connectivity*

Eric R. Cosman, Jr.[1,3] and William M. Wells III[1,2,3]

[1] Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, Cambridge, MA, USA
ercosman@mit.edu, sw@csail.mit.edu
[2] Harvard Medical School, Brigham and Woman's Hospital,
Department of Radiology, Boston, MA, USA
[3] The Harvard Center for Neurodegeneration and Repair, Boston, MA, USA

Abstract. A hierarchical model based on the Multivariate Autoregessive (MAR) process is proposed to jointly model neurological time-series collected from multiple subjects, and to characterize the distribution of MAR coefficients across the population from which those subjects were drawn. Thus, inference about effective connectivity between brain regions may be generalized beyond those subjects studied. The posterior on population- and subject-level connectivity parameters are estimated in a Variational Bayesian (VB) framework, and structural model parameters are chosen by the corresponding evidence criteria. The significance of resulting connectivity statistics are evaluated by permutation-based approximations to the null distribution. The method is demonstrated on simulated data and on actual multi-subject neurological time-series.

1 Introduction

Neuroimaging studies are regularly conducted in which measurements of brain activity are collected simultaneously from multiple, analogous brain regions in multiple subjects. These measurement may be taken by EEG, MEG, fMRI or intracranial electrical monitoring. Such studies are often motivated by hypotheses about the interaction among gross brain regions under particular experimental conditions or due to some neurological disorder. Numerous techniques have been proposed which use such functional data to characterize *Effective Connectivity*, defined as the influence that one brain region exerts on another under a given interaction model. These include Structural Equation Modeling (SEM) [14], Multivariate Autoregressive Modeling (MAR) [9, 16, 7] and Dynamic Causal Modeling (DCM) [4]. The MAR and DCM approaches characterize effective connectivity by modeling particular brain regions as variables in a causal, dynamical system. Model parameters may thereby inform about the influence that each region exerts on the others, either directly or indirectly through other regions.

It is of interest to determine which interactions between modeled brain regions are characteristic under experimental conditions within the population

* This research was supported by grants NIH 5 P41 RR13218 and FIRST BURN.

from which studied subjects are drawn. However, to our knowledge, models for effective connectivity have been applied only in subject-independent manner. As such, statistical inference about connectivity is limited to the specific subjects included in a study, except by *post hoc* analysis of the variation in subject-specific connectivity parameters. We propose a hierarchical, or Random Effects (RFX), approach to the analysis of multi-subject functional time-series such that the effective connectivity parameters of all subjects are estimated jointly along with a density describing the variation in those parameters across the population from which studied subjects were drawn. Such joint estimation of connectivity under SEM, MAR, DCM or other models, would be valuable for neuroimaging studies. We have chosen the MAR process as a starting point for investigation of the applicability of such hierarchical modeling to generalize inference about effective connectivity to the population level.

In particular, we present the RFX-MAR model which parameterizes the interactions of specified brain regions for each subject using the MAR process, and describes the variability in those subject-level models by the mean and variance of their MAR coefficients. We estimate population- and subject-specific parameters in a Variational Bayesian (VB) framework, and characterize effective connectivity at the population level by inference on the posterior of the MAR coefficients' means, which we refer to as the population-level MAR coefficients. Specifically, we compute statistics related to evidence of non-zero directional influence between each pair of monitored brain regions under the MAR model, across the sampled population. Structural parameters of the RFX-MAR model are selected by an approximate maximum evidence criteria.

Though Gaussian population models have been used extensively in neuroimaging analysis for the purpose of localizing protocol-related neural activity [5], this work is the first to investigate their utility in modeling of neural connectivity. We describe the results of our analysis on synthetic and EEG time-series collected from multiple subjects.

2 The Multivariate Autoregressive Process

The Multivariate Autoregressive (MAR) process models the temporal dynamics of multivariate systems causally and without hidden state variables, such that multivariate measurements at the present time are a linear function of measurements in the past. This kind of parametric process has been used to identify the linear, time-invariant (LTI) system dynamics and spectra of multichannel time-series data in a number of contexts, including geophysics, economics and neuroimaging [9, 16, 7]. In neuroimaging, the MAR process has been used to model and test effective connectivity based on neurological data collected by fMRI, EEG and direct electrical recording. In general, the MAR process may be used when specialization of a dynamical model, possibly through hidden state variables, is difficult or unnecessary. Since the MAR process is strictly causal, it can model directional influence between channels, and elucidate causal chains and loops.

A MAR(p) process is defined as follows, where $\mathbf{Y}_{n\bullet} \in \Re^d$ is a sample from d channels at time n arranged in a row vector, $\mathbf{A}(l) \in \Re^{d\times d}$, $l = 1, \ldots, p$, is a series of matrices comprising the coefficients of the MAR model, and $\mathbf{E}_{n\bullet} \in \Re^d$ is a temporally-white innovation with stationary distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Lambda}^{-1})$. Note that we replace a matrix or sequence index with a large dot $\bullet$ to refer collectively to elements corresponding to all values of that index.

$$\mathbf{Y}_{n\bullet} = \sum_{l=1}^{p} \mathbf{Y}_{(n-l)\bullet}\, \mathbf{A}(l) + \mathbf{E}_{n\bullet} = \underbrace{[\mathbf{Y}_{(n-1)\bullet} \,|\, \ldots \,|\, \mathbf{Y}_{(n-p)\bullet}]}_{\equiv \mathbf{X}_{n\bullet}} \underbrace{\begin{bmatrix} \mathbf{A}(1) \\ \hline \vdots \\ \hline \mathbf{A}(p) \end{bmatrix}}_{\equiv \mathbf{W}} + \mathbf{E}_{n\bullet} \quad (1)$$

We denote the p coefficients by which channel i directly influences channel j as $A_{ij}(\bullet) \in \Re^p$, and refer to this as the *Direct Influence Function* (DIF) from channel i to j. Stacking these equations for each time sample n, we get the matrix equation $\mathbf{Y} = \mathbf{XW} + \mathbf{E}$. To highlight that this is a specialized linear regression model, we will henceforth use its vectorized form, where $\mathbf{y} \equiv \mathrm{vec}(\mathbf{Y}) \in \Re^{Nd}$, $\mathbf{w} \equiv \mathrm{vec}(\mathbf{W}) \in \Re^{d^2 p}$, and $\otimes$ denotes the Kronecker product (as defined in [13]):

$$p(\mathbf{y} \mid \mathbf{w}, \mathbf{\Lambda}) = \mathcal{N}(\mathbf{y}\,;\, (\mathbf{I}_d \otimes \mathbf{X})\mathbf{w}\,,\, \mathbf{\Lambda}^{-1} \otimes \mathbf{I}_N) \quad (2)$$

The utility of maximum likelihood (ML) estimates of MAR parameters, $\tilde{\mathbf{w}} = \mathrm{vec}(\mathbf{X}^{+}\mathbf{Y})$ and $\tilde{\mathbf{\Lambda}} = (\mathbf{Y} - \mathbf{X}\tilde{\mathbf{W}})'(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{W}})/(N - d^2 p)$, is limited by the large amount of data required to fit these $\Theta(d^2)$ parameters reliably. A Variational Bayesian framework for MAR estimation has been proposed which relieves this data requirement to some degree, by means of a prior that regularizes coefficient magnitudes [16].

Neuroimaging experiments are commonly associated with time-varying stimuli or tasks. Encodings of such information can be added as linear terms in Equation 1 without affecting analysis. As such, they account for bias in the innovations, and do not influence system dynamics *per se*, which are assumed to be stationary in the time-series $\mathbf{y}$. Nonlinear coupling between variables can be approximated in the MAR framework by adding new variables which are nonlinear functions of data from other variables (e.g. product terms) [7].

3 The RFX-MAR Model

To generalize inference about effective connectivity to the greater population from which studied subjects were drawn, we propose to model the variation of subject-specific MAR coefficients across that greater population. As is standard in population inference, we approximate the population density as Gaussian, so that its mean characterizes dynamical structure common to the population, and its covariance characterizes the degree of variability in that structure found within that population. Hence, we construct a hierarchical, or *Random Effects* (RFX), model which describes both the subject-specific MAR parameters and

the inter-subject variation of those parameters.

$$p(\mathbf{y}_k \mid \mathbf{w}_k,\, \mathbf{\Lambda}_k) = \mathcal{N}\left(\mathbf{y}_k\,;\, (\mathbf{I}_d \otimes \mathbf{X}_k)\mathbf{w}_k\,,\, \mathbf{\Lambda}_k^{-1} \otimes \mathbf{I}_n\right),\;\; k = 1, \ldots, S \tag{3}$$

$$p(\mathbf{w}_k \mid \mathbf{w}_0,\, \boldsymbol{\gamma}) = \mathcal{N}\Big(\mathbf{w}_k;\; \mathbf{w}_0\,,\, \sum_{h=1}^{H} \gamma_h^{-1}\mathbf{Q}_{\gamma_h}\Big) \tag{4}$$

$$p(\mathbf{w}_0 \mid \boldsymbol{\alpha}) = \mathcal{N}\Big(\mathbf{w}_0;\; \mathbf{0},\, \sum_{g=1}^{G} \alpha_g^{-1}\mathbf{Q}_{\alpha_g}\Big) \tag{5}$$

In particular, we model the multivariate time-series $\mathbf{y}_k \in \Re^{N_k d}$ from each subject k as MAR(p) process with coefficients $\mathbf{w}_k$ and innovations precision (inverse covariance) $\mathbf{\Lambda}_k$ (Equation (3)). The data from subject k comprise N_k time samples of dimension d. Each subject-specific MAR coefficient $[\mathbf{w}_k]_i$ is drawn independently about its population mean $[\mathbf{w}_0]_i$ with some precision γ_h (Equation (4)). We group coefficients to reflect similarity in their inter-subject variation, rather than assuming a single variance for all coefficients a priori. Each of these random-effects variance groups are associated with a precision $\gamma_h,\; h = 1, \ldots, H$. This "structuring" of inter-subject variation is defined as follows:

$$\text{group}_\gamma(i) \equiv \text{the RFX group of } [\mathbf{w_k}]_i \qquad [\mathbf{Q}_{\gamma_h}]_{ij} \equiv \delta_{ij}\delta(\text{group}_\gamma(i) = h) \tag{6}$$

Additionally, we assume that each population-level coefficient $[\mathbf{w}_0]_i$ is drawn independently from a zero-mean Gaussian, and partitioned into one of G groups with other coefficients with similar magnitudes (Equation (5)). This kind of "structured" prior was used to reduce the effective degrees of freedom in single-subject MAR modeling [16, 7], and has been referred to as an *Automatic Relevance Determination* (ARD) prior [11]. This ARD structuring is defined as follows, where $g = 1, \ldots, G$ indexes the groups:

$$\text{group}_\alpha(i) \equiv \text{the ARD group of } [\mathbf{w}_0]_i \qquad [\mathbf{Q}_{\alpha_g}]_{ij} \equiv \delta_{ij}\delta(\text{group}_\alpha(i) = g) \tag{7}$$

In summary, we have a three-level linear Gaussian model. Its first level describes subject-specific variation with the MAR process. Its second level describes variation in subject-specific coefficients across the sampled population by means of a Gaussian density with diagonal precision matrix. The third level regularizes the magnitude of the population-level coefficients. Naturally, care should be taken to look for inconsistency between this model and the qualities of data to which it is applied. Since the data has zero mean under the MAR model, the sample mean is removed from each data channel as a pre-processing step. Furthermore, each channel's signal is individually normalized by its sample variance (we expect to model stable systems). By doing so, the cross-coefficients of $\mathbf{w}_k$, i.e. $[\mathbf{A}_k]_{ij}(\bullet)$ for $i \neq j$, are comparable across subjects, normalized by the ratio between the amplitudes of the "to" and "from" signals $\big\|[\mathbf{Y}_k]_{\bullet j}\big\| / \big\|[\mathbf{Y}_k]_{\bullet i}\big\|$.

For notational simplicity, we will henceforth refer to the $N \equiv \sum_{k=1}^{S} N_k$ samples of multi-subject data collectively as $\mathbf{y} \equiv [\mathbf{y}_1' \ldots \mathbf{y}_S']'$; the subject-level precisions as $\mathbf{\Lambda} \equiv \{\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_S\}$; and the population and subject-specific coefficients as $\mathbf{w} \equiv [\mathbf{w}_0'\;\; \mathbf{w}_1' \ldots \mathbf{w}_S']'$.

3.1 Precision Priors

Since we intend to estimate the RFX-MAR model in a Bayesian framework, we set the following "noninformative" priors for the precision parameters in (3)–(5) to represent an absence of prior information.

$$p(\mathbf{\Lambda}) \propto \prod_{k=1}^{s} \left|\mathbf{\Lambda}_k\right|^{-\frac{d+1}{2}} \qquad p(\boldsymbol{\alpha}) = \prod_{g=1}^{G} \mathrm{Ga}(\alpha_g\,;\, a_p\,,\, b_p), \;\; a_p,\, b_p \equiv 10^{-3}$$
$$p(\boldsymbol{\gamma}) = \prod_{h=1}^{H} \left(2u\gamma_h^{\frac{3}{2}}\right)^{-1}, \;\; \gamma_h \geq u^{-2}, \;\; u \equiv 10^3 \tag{8}$$

We subscribe to the view in [6] that "any noninformative prior distribution [is] inherently provisional— after the model has been fit, one should look at the posterior distribution to see if it makes sense." We follow [16] in the form of the priors on precisions $\mathbf{\Lambda}$ and $\boldsymbol{\alpha}$ due to their success in the single-subject MAR modeling and in our experiments on synthetic multi-subject data. The prior on $\boldsymbol{\alpha}$ follows a Gamma density [1], and the prior on $\mathbf{\Lambda}$ is improper. Both were motivated by Jeffreys' Rule [2]; however, they do not follow from its strict application, but rather from its application to the first and third levels of the model in isolation. The prior on $\boldsymbol{\gamma}$ is equivalent to a locally uniform prior on the standard deviation of inter-subject variation $\sigma_h \equiv \gamma_h^{-0.5}$ for each RFX variance group h. This prior is suggested for RFX models as preferable to those of the family $\mathrm{Ga}(\gamma_h; \epsilon, \epsilon)$, under which inference is sensitive to ϵ when σ_h is near zero [6]. We observed this sensitivity in our experiments.

4 Variational Posterior Estimation

In this section, we describe a Variational Bayesian (VB) algorithm [8, 10] for estimating the posterior of the RFX-MAR model's real-valued parameters, including the MAR coefficients $\mathbf{w}$ and precision parameters $\mathbf{\Lambda}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$. Our choice of this framework for posterior estimation was motivated by its suitability to estimation of a single MAR process [16, 7]. In Sect. 5, we address selection of the model's discrete-valued, structural parameters, which include the MAR model order and the RFX and ARD structuring functions.

The VB algorithm proceeds as follows: For a generic statistical model with parameters $\boldsymbol{\theta}$ and observed data $\mathbf{D}$, an approximation to the true posterior $q(\boldsymbol{\theta}) \equiv \hat{p}(\boldsymbol{\theta} \mid \mathbf{D})$ is produced by maximizing a lower bound on the model's log evidence

$$F \;\equiv\; \log p(\mathbf{D}) - D(q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta} \mid \mathbf{D})) \leq \log p(\mathbf{D}) \tag{9}$$

with the simplifying assumption that the posterior approximation factorizes $q(\boldsymbol{\theta}) \equiv \prod_i q(\boldsymbol{\theta}_i)$ in some way. The quantity F is referred to as the *negative variational free energy*. Maximization proceeds by fixed-point iteration whereby the posterior for each subset of parameters $\boldsymbol{\theta}_i$ is updated sequentially, while holding the posterior of remaining parameters constant. If priors are set to be conditionally-conjugate under such posterior independence assumptions, each

VB update step may reduce to a closed-form update of the induced sufficient statistics of $q(\boldsymbol{\theta}_i)$. This is referred to as *Free-Form Variational Bayes*. Alternatively, *Fixed-Form Variational Bayes* refers to a VB update step which follows from assuming additionally that the posterior factor has a particular parametric form. For instance, the *Expectation Maximization* (EM) algorithm [3] is a special case in which the second of two parameter groups is assumed to have a singular posterior $q(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \equiv q(\boldsymbol{\theta}_1)\delta(\boldsymbol{\theta}_2 - \widehat{\boldsymbol{\theta}}_2)$ with parameter $\widehat{\boldsymbol{\theta}}_2$.

For the RFX-MAR model, we assume the posterior independence of the precisions parameters and MAR coefficients. With this assumption, it can be shown that precisions' posterior further factorizes, due to the graphical structure of the model and their prior independence:

$$q(\mathbf{w}, \boldsymbol{\Lambda}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) \equiv q(\mathbf{w})q(\boldsymbol{\Lambda}, \boldsymbol{\gamma}, \boldsymbol{\alpha}) = q(\mathbf{w}) \prod_{k=1}^{S} q(\boldsymbol{\Lambda}_k) \prod_{h=1}^{H} q(\gamma_h) \prod_{g=1}^{G} q(\alpha_g) \quad (10)$$

Furthermore, with priors on the precision parameters of the form given in (8), it can be shown that the posterior factors follow Normal, Wishart, Gamma and incomplete Gamma[4] densities (denoted $\mathcal{N}$, $\mathcal{W}$, Ga, and IGa, respectively) [1]. Thus the free-form VB algorithm proceeds by sequential update of the sufficient statistics of each posterior factor until convergence to a fixed point. These updates are given below in (11)–(14). For clarity in these equations, we let $\sum_{i_g}$ denote summation over the κ_g coefficient indices i which are part of ARD prior group g, i.e. $\sum_{i\,:\,\mathrm{group}_\alpha(i)=g}$. Similarly, $\sum_{i_h}$ denotes $\sum_{i\,:\,\mathrm{group}_\gamma(i)=h}$, where the RFX group h contains ν_h coefficients. We also define $\tilde{\boldsymbol{\Lambda}}_k \equiv \widehat{\boldsymbol{\Lambda}}_k \otimes \mathbf{X}_k'\mathbf{X}_k$ and $\tilde{\mathbf{w}}_k \equiv \mathrm{vec}(\,\mathbf{X}_k^{+}\mathbf{Y}_k\,)$, and let $\widehat{\boldsymbol{\Gamma}} \equiv \sum_{h=1}^{H} \widehat{\gamma}_h \mathbf{Q}_{\gamma_h}$ and $\widehat{\boldsymbol{\Xi}} \equiv \sum_{g=1}^{G} \widehat{\alpha}_g \mathbf{Q}_{\alpha_g}$ denote the estimated precision matrices for the second and third levels of the RFX-MAR model.

Update for $q(\mathbf{w}) \leftarrow \mathcal{N}(\mathbf{w}\,;\,\widehat{\mathbf{w}}\,,\,\widehat{\boldsymbol{\Sigma}})$: The $(k,l)^{th}$ block of $\widehat{\boldsymbol{\Sigma}}$ of size $d^2p \times d^2p$ is denoted $\widehat{\boldsymbol{\Sigma}}^{(kl)}$, $k, l = 0, \ldots, S$

$$\begin{aligned}
\widehat{\boldsymbol{\Sigma}}^{(00)} &= \left(\widehat{\boldsymbol{\Xi}} + S\widehat{\boldsymbol{\Gamma}} - \widehat{\boldsymbol{\Gamma}} \left[\sum_{k=1}^{S} (\widehat{\boldsymbol{\Gamma}} + \tilde{\boldsymbol{\Lambda}}_k)^{-1} \right] \widehat{\boldsymbol{\Gamma}} \right)^{-1} \\
\widehat{\boldsymbol{\Sigma}}^{(0k)} &= \widehat{\boldsymbol{\Sigma}}^{(k0)\prime} = \widehat{\boldsymbol{\Sigma}}^{(00)} \widehat{\boldsymbol{\Gamma}} (\widehat{\boldsymbol{\Gamma}} + \tilde{\boldsymbol{\Lambda}}_k)^{-1}, \quad k = 1, \ldots, S \\
\widehat{\boldsymbol{\Sigma}}^{(kl)} &= \delta_{kl} (\widehat{\boldsymbol{\Gamma}} + \tilde{\boldsymbol{\Lambda}}_k)^{-1} + (\widehat{\boldsymbol{\Gamma}} + \tilde{\boldsymbol{\Lambda}}_k)^{-1} \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Sigma}}^{(00)} \widehat{\boldsymbol{\Gamma}} (\widehat{\boldsymbol{\Gamma}} + \tilde{\boldsymbol{\Lambda}}_l)^{-1}, \quad k, l \geq 1 \\
\widehat{\mathbf{w}}' &= \left[\begin{array}{cccc} \mathbf{0}' & (\tilde{\boldsymbol{\Lambda}}_1 \tilde{\mathbf{w}}_1)' & \ldots & (\tilde{\boldsymbol{\Lambda}}_S \tilde{\mathbf{w}}_S)' \end{array} \right] \widehat{\boldsymbol{\Sigma}}
\end{aligned} \quad (11)$$

Update for $q(\boldsymbol{\Lambda}) \leftarrow \prod_{k=1}^{S} \mathcal{W}_d(\boldsymbol{\Lambda}_k\,;\,a_k\,,\,\mathbf{B}_k)$**:**

$$\begin{aligned}
a_k &= N_k \qquad\qquad \boldsymbol{\Omega}_k \equiv \sum_{n=1}^{N_k} (\mathbf{I}_d \otimes [\mathbf{X}_k]_{n\bullet}) \widehat{\boldsymbol{\Sigma}}^{(kk)} (\mathbf{I}_d \otimes [\mathbf{X}_k]_{n\bullet})' \\
\mathbf{B}_k &= (\mathbf{Y}_k - \mathbf{X}_k \widehat{\mathbf{W}}_k)'(\mathbf{Y}_k - \mathbf{X}_k \widehat{\mathbf{W}}_k) + \boldsymbol{\Omega}_k \qquad \widehat{\boldsymbol{\Lambda}}_k \equiv \mathcal{E}_{q(\boldsymbol{\Lambda}_k)}\{\boldsymbol{\Lambda}_k\} = a_k \mathbf{B}_k^{-1}
\end{aligned} \quad (12)$$

[4] We define the incomplete Gamma density $\mathrm{IGa}(x; a, b, x_{min})$ to be proportional to the Gamma density $\mathrm{Ga}(x; a, b)$ for positive values $x \geq x_{min}$, and zero otherwise.

Update for $q(\boldsymbol{\gamma}) \leftarrow \prod_{h=1}^{H} \mathrm{IGa}(\gamma_h\,;\, a_{\gamma_h},\, b_{\gamma_h},\, u^{-2})$ **:**

$$a_{\gamma_h} = \frac{\nu_h S - 1}{2} \qquad b_{\gamma_h} = \frac{1}{2}\sum_{k=1}^{S}\sum_{i_h}\left[[\widehat{\boldsymbol{\Sigma}}^{(00)} - 2\widehat{\boldsymbol{\Sigma}}^{(k0)} + \widehat{\boldsymbol{\Sigma}}^{(kk)}]_{i_h i_h} + [\widehat{\mathbf{w}}_k - \widehat{\mathbf{w}}_0]^2_{i_h}\right] \tag{13}$$

$$\widehat{\gamma}_h \equiv \mathcal{E}_{q(\gamma_h)}\{\gamma_h\} = \frac{1}{b_{\gamma_h}}\left[a_{\gamma_h} + \frac{(b_{\gamma_h}u^{-2})^{a_{\gamma_h}}\exp\{-b_{\gamma_h}u^{-2}\}}{\Gamma(a_{\gamma_h},\, b_{\gamma_h}u^{-2})}\right]$$

Update for $q(\boldsymbol{\alpha}) \leftarrow \prod_{g=1}^{G} \mathrm{Ga}(\alpha_g\,;\, a_{\alpha_g}, b_{\alpha_g})$**:**

$$a_{\alpha_g} = a_p + \frac{\kappa_g}{2} \qquad b_{\alpha_g} = b_p + \frac{1}{2}\sum_{i_g}\left[[\widehat{\boldsymbol{\Sigma}}^{(00)}]_{i_g i_g} + [\widehat{\mathbf{w}}_0]^2_{i_g}\right] \tag{14}$$

$$\widehat{\alpha}_g \equiv \mathcal{E}_{q(\alpha_g)}\{\alpha_g\} = a_{\alpha_g} b_{\alpha_g}^{-1}$$

In our experiments with this algorithm, we found that the sufficient statistics for $q(\boldsymbol{\Lambda})$ and $q(\boldsymbol{\alpha})$ converge quite rapidly, but that convergence of those for $q(\boldsymbol{\gamma})$ is extremely slow for data in which any RFX group has a large precision γ_h. This observation is consistent with the literature on EM estimation of generic RFX models [15]. However, we found that posterior optimization could be made quite rapid by using Powell's direction set method [17] to estimate the otherwise slow-converging sufficient statistics of $q(\boldsymbol{\gamma})$. In this approach, the optimal negative variational free energy F at each setting of $q(\boldsymbol{\gamma})$ is computed using the relatively rapid VB fixed-point iteration for the remaining parameters (Equations (11), (12) and (14)):

$$\underbrace{\max_{\widehat{\boldsymbol{\gamma}}}\overbrace{\max_{\widehat{\mathbf{w}},\widehat{\boldsymbol{\Sigma}},\widehat{\boldsymbol{\Lambda}},\widehat{\boldsymbol{\alpha}}} F(\widehat{\mathbf{w}},\widehat{\boldsymbol{\Sigma}},\widehat{\boldsymbol{\Lambda}},\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\alpha}})}^{\text{VB iteration}}}_{\text{Powell's Method}} \tag{15}$$

Above, we are able write F above as a function solely of the precision means since only one of the two sufficient statistics of their respective posteriors varies during VB optimization.

We initialize this hybrid algorithm by computing the sample mean and precision of source-independent ML estimates of the source-level coefficients and noise precisions, and then by running the full VB iteration until the change in $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\boldsymbol{\alpha}}$ becomes less than 10^{-4}. This quickly produces a posterior estimate $q(\boldsymbol{\Lambda}, \boldsymbol{\gamma}, \boldsymbol{\alpha})$ which is nearly optimal. We cache the optimal $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\boldsymbol{\alpha}}$ for recent evaluations of $F(\widehat{\boldsymbol{\gamma}})$ to initialize the VB iteration of subsequent evaluations, so that as Powell's method converges, the VB iteration is started very close to its fixed-point. We terminate the hybrid optimization when F is maximized to precision $\pm 10^{-10}$.

The following is an expression for the negative variational free energy F under the RFX-MAR model. Note that we have dropped an infinite constant due to the improper prior on $\boldsymbol{\Lambda}$, and have canceled a number of terms by assuming that F is evaluated after the update steps for $q(\boldsymbol{\Lambda})$ and $q(\boldsymbol{\alpha})$, but before the next update for $q(\mathbf{w})$. If $q(\boldsymbol{\gamma})$ is being updated, then the last line also cancels.

The block structure of $\widehat{\mathbf{\Sigma}}$ can be used to make evaluation of its determinant $|\widehat{\mathbf{\Sigma}}| = |\widehat{\mathbf{\Sigma}}^{(00)}| \prod_{k=1}^{S} |\widehat{\mathbf{\Gamma}} + \tilde{\mathbf{\Lambda}}_k|^{-1}$ computationally manageable.

$$
\begin{aligned}
F = \frac{d}{2}\left(\frac{S(d-1)}{2} - N\right)\log\pi + \frac{(S+1)d^2p}{2} + \frac{1}{2}\log|\widehat{\mathbf{\Sigma}}| + \sum_{g=1}^{G}\log\frac{b_{\alpha_g}^{a_{\alpha_g}}\ \Gamma(\widehat{a}_{\alpha_g})}{\widehat{b}_{\alpha_g}^{\widehat{a}_{\alpha_g}}\ \Gamma(a_{\alpha_g})} \\
+ \sum_{h=1}^{H}\log\frac{\Gamma(\widehat{a}_{\gamma_h}, \widehat{b}_{\gamma_h}u^{-2})}{2u\widehat{b}_{\gamma_h}^{\widehat{a}_{\gamma_h}}} + \sum_{k=1}^{S}\left[\frac{N_k}{2}\log\left(|\widehat{\mathbf{\Lambda}}_k|N_k^{-d}\right) + \sum_{i=0}^{d-1}\log\Gamma\left(\frac{N_k - i}{2}\right)\right] \\
+ \sum_{h=1}^{H}\widehat{\gamma}_h\left(\widehat{b}_{\gamma_h} - \frac{\widehat{\gamma}_h}{2}\sum_{k=1}^{S}\sum_{i_h}\left[[\widehat{\mathbf{\Sigma}}^{(00)} - 2\widehat{\mathbf{\Sigma}}^{(k0)} + \widehat{\mathbf{\Sigma}}^{(kk)}]_{i_h i_h} + [\widehat{\mathbf{w}}_k - \widehat{\mathbf{w}}]_{i_h}^2\right]\right)
\end{aligned}
$$

5 Model Structure Selection

Since the VB algorithm maximizes a lower bound on the log evidence $p(\mathbf{D}|H)$ of a generic model H, the optimized negative variational free energy $F(H)$ can be used to approximate the posterior on model structure $p(H|\mathbf{D})$ by search over competing models [16, 12]. For the RFX-MAR model, H is parameterized by its discrete-valued structural parameters: the MAR model order p, the RFX structuring function $\text{group}_\gamma()$, and the ARD structuring function $\text{group}_\alpha()$. Thus, inference can proceed either by model averaging (for which $\exp(F(H))$ weights the posterior $q(\boldsymbol{\theta}|H)$ under each H), or by model selection (for which inference proceeds based only on the posterior $q(\boldsymbol{\theta}|\widehat{H})$ under the model $\widehat{H} = \text{argmax}\, F(H)$ with maximum approximate evidence). The latter VB model selection criteria is equivalent to the Bayesian Information Criteria (BIC) in the large sample limit, and has been shown superior to BIC for model order selection in single-subject MAR modeling [16]. We choose a model selection framework since in our experience with the RFX-MAR on real and synthetic data, $F(H)$ is typically strongly peaked and the coefficient posterior is very similar under all explored model structures with non-vanishing evidence.

We search the MAR order parameter p exhaustively over positive integer values, up to some maximum. Since exhaustive search is not possible for the RFX and ARD group functions, we follow [16] which suggests a semi-automatic, heuristic search method for finding a structuring function likely to group coefficients appropriately. These may include hand-tailored structurings, or generic ones such as a "global" function which puts all coefficients into the same group. Another example is an "interaction" function which groups all coefficients corresponding to interactions between channels, $A_{ij}(\cdot)$ for all $i \neq j$, and places the remaining coefficients into a second group. Such functions may also be selected semi-automatically. For instance, an "auto" function for the ARD structuring can be produced by k-means clustering of the MAP coefficient estimates under the "global" structuring. An "auto" function for the RFX variance structuring can be produced by k-means clustering of the sample variances of MAP coefficient estimates from subject-independent MAR modeling.

6 Connectivity Inference

We are principally interested in inferring which direct interactions between modeled brain regions are non-zero within a population under experimental conditions. Inference of this kind will be based on (marginalization of) the posterior of the population-level coefficients $q(\mathbf{w}_0)$. Having selected a model structure by the VB maximum evidence criteria, we can report on the population-level effective connectivity between variables i and j by computing a statistic that relates to the posterior "plausibility" of $A^0_{ij}(\cdot) = \mathbf{0}$ under $q(\mathbf{w}_0)$, where $A^0_{ij}(\cdot)$ are the coefficients in $\mathbf{w}_0$ related to the direct influence of variable i on j. Generally speaking, one can report on the posterior plausibility of a specific parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ by computing the complementary probability content of the smallest *Highest Probability Density* (H.P.D.) region of the posterior $q(\boldsymbol{\theta})$ that contains the value in question [2]. For a Gaussian posterior $q(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta};\ \boldsymbol{\mu}, \boldsymbol{\Omega})$ with full-rank covariance $\boldsymbol{\Omega}$, it is readily shown that this connectivity statistic s_0 is a monotonic function of the Mahalanobis distance from the posterior mean $\boldsymbol{\mu} \in \Re^r$ to the specified value $\boldsymbol{\theta}_0$. Note that s_0 decreases as $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ becomes less plausible a posteriori.

$$s_0 = 1 - \chi^2_r\left((\boldsymbol{\theta}_0 - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0 - \boldsymbol{\mu})\right) \tag{16}$$

To assess the specificity of a test which rejects the null hypothesis (that $A^0_{ij}(\cdot) = \mathbf{0}$) by thresholding such connectivity statistics, we simulate a null distribution by permutations of the time-series data and re-estimation of the these statistics [9]. In particular, we randomly sample circular translations of each of the d time-series for each subject such that every pair of univariate time-series are shifted by more than $2p_{max}$ lags. Here, p_{max} is the maximum MAR model order entertained in our search. This produces multivariate time-series whose univariate statistics are roughly unaffected, while removing causal interactions between variables.

7 Experiments

The ability of the MAR process to capture inter-regional neural dynamics has been shown on synthetic data from biologically plausible models [9]. We validated our estimation algorithm in terms of its ability to estimate all real-valued and structural RFX-MAR parameters using multivariate time-series sampled from known RFX-MAR models. The scope of our investigation was quite broad and involved varying the number of subjects S, number of samples per subject N_k, and degree of inter-subject variance of MAR coefficients. We considered the cases where the subject-specific coefficients were drawn about a single mean, about two means, and drawn independently. We also compared performance of the RFX-MAR model to the subject-independent MAR estimation of [16], and to a "fixed-effects" model in which all subject-level coefficients were assumed to be identical. In summary, we found parameter estimation and model selection to be robust when the number of subjects $S \geq 10$, and the number of samples per subject N_k was greater than approximately three times the number of parameters

per subject at the optimal model order $d^2(p+1)$. Classification error for non-zero direct influences between pairs of variable was less than 5% in general. When the amount of data was more limited, the RFX-MAR model tended to overestimate the model order p, but with little impact on connectivity inference. Subject-independent estimation of p was generally more robust.

RFX-MAR analysis of a small number of fMRI and EEG multi-subject datasets have produced promising results. Here, we detail our analysis of multi-subject EEG time-series from the UCI Knowledge Discovery Database[5]. This dataset contains measurements from 64 electrodes placed on the scalps of healthy subjects and sampled at 256 Hz for 1 second during a picture presentation task. The protocol and data are described in [18]. We performed RFX-MAR modeling of 1 second of data from each of S=20 healthy subjects, having selected d=6 channels from the frontal (F5, F6), temporal (T7, T8), and parietal (P3, P4) regions in each hemisphere. We searched all combinations of the following structural parameter settings: $p = \{1, \ldots, 5\}$, $\text{group}_\gamma = \{\text{global, interaction, auto (H=2, 3, 4)}\}$ and $\text{group}_\alpha = \{\text{global, interaction, auto (G=2, 3, 4)}\}$. The MAP model structure was $p = 5$, $\text{group}_\gamma = \text{group}_\alpha = \text{interaction}$. This model order was consistent with the most common model order estimated by subject-independent analysis of the data. The MAP inter-subject standard deviation of coefficients involved with within-channel predictions was 0.027, whereas that for coefficients involved in cross-channel prediction was 0.007. The posterior of the population-level coefficients is shown in Fig. 1. We computed connectivity statistics using (16) to assess evidence of a non-zero causal interaction between each pair of channels, and estimated the null distribution of these statistics by repeating analysis on 100 randomly, circularly shifted versions of each channel, in each subject. This yielded $100\,d(d-1) = 3000$ samples of connectivity statistics between channels for which there is unlikely to be a causal connection. Figure 2 shows the inferred population-level effective connectivity pattern, the associated connectivity statistics, and their estimated p-values. We note that few variables were found to be interacting when the same data was modeled in a subject-independent manner, except when a larger sample of the data was included for each subject. This points to the cross-subject regularizing effect that joint modeling of multiple subjects can have to elucidate more subtle connectivity patterns in the data. We repeated this analysis for S=20 different healthy subjects from the same study, and found the population-level posterior and effective connectivity pattern to be very similar.

8 Discussion

We have presented the initial development of a method for population modeling of effective connectivity among brain regions based on neurological time-series. Numerous avenues exist for further elaboration, validation and utilization of this kind of model. We are particularly interested in different ways two populations

[5] *http://kdd.ics.uci.edu* These data were provided by Henri Begleiter at the Neurodynamics Laboratory at the State University of New York Health Center at Brooklyn.

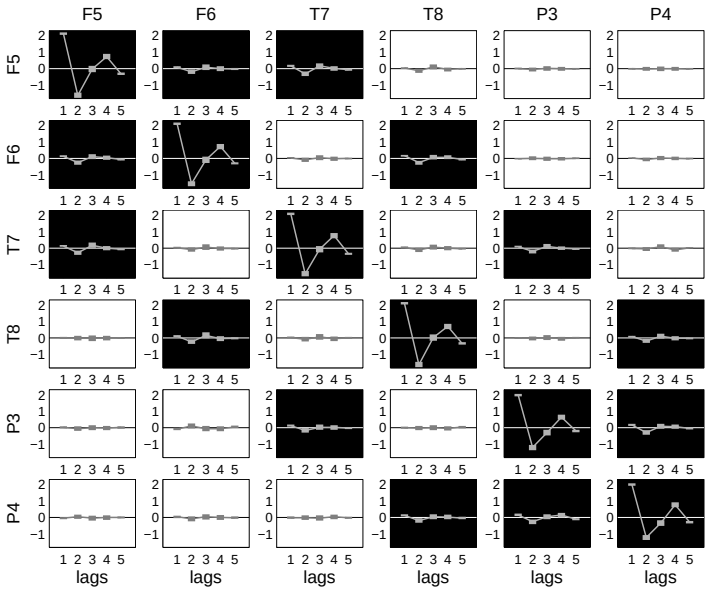

Fig. 1. The posterior mean of the population-level directed influence functions $A^0_{ij}(\cdot)$ (plotted with 99% H.P.D.-content error bars) given the RFX-MAR structural parameters with maximum F for the EEG dataset. The background of plot (i, j) is colored black when the connectivity statistic s_0 for $A^0_{ij}(\cdot) = \mathbf{0}$ is less than 10^{-6}.

(distinguished, for example, by the presence of disease) might be compared on the basis of effective connectivity. Certainly, one can perform inference on the population-level coefficients produced by independent RFX-MAR modeling of the two populations. However, the success of classifiers produced from such population models would lend credence to what is necessarily a model-dependent characterization of the interaction among brain regions. It would also be valuable to construct similar population models for more elaborate systems models, such as those not limited to time-invariant connectivity patterns. Finally, we note that the RFX-MAR model can also be used to characterize other types of variation in connectivity parameters, such as that arising from repeated trials under similar experimental conditions.

References

1. J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* John Wiley & Sons, 1994.
2. G. E. P. Box and G. C. Tao. *Bayesian Inference in Statistical Analysis.* John Wiley & Sons, 1992.
3. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *J. Royal Statist. Soc. Ser. B*, 39:1–38, 1977.
4. K. J. Friston, L. Harrison, and W. Penny. Dynamic causal modeling. *NeuroImage*, 19, 2003.

To Channel j

From Channel i	F5	F6	T7	T8	P3	P4
F5	<1e-16 0	6e-14 0	<1e-16 0	2e-5 0	8e-2 9e-3	1e-3 6e-2
F6	<1e-16 0	<1e-16 0	2e-2 2e-3	<1e-16 0	1e-1 1e-2	1e-3 3e-4
T7	<1e-16 0	5e-4 0	<1e-16 0	4e-5 0	<1e-16 0	8e-4 0
T8	4e-1 1e-1	2e-14 0	4e-3 6e-4	<1e-16 0	3e-1 5e-2	3e-16 0
P3	2e-4 0	2e-6 0	<1e-16 0	4e-6 0	<1e-16 0	<1e-16 0
P4	2e-1 2e-2	1e-1 1e-2	3e-2 2e-3	<1e-16 0	<1e-16 0	<1e-16 0

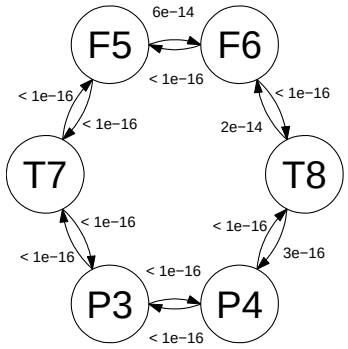

Fig. 2. A graphical representation of the population-level effective connectivity inferred from the EEG data. We only show edges between channels whose connectivity statistic is $\leq 10^{-6}$, and omit self interactions for clarity. EEG channels with odd (even) index are in the left (right) hemisphere. This pattern suggests that regions which are spatially closer interact to a greater degree under the MAR model. The table gives the connectivity statistic (top) and the estimated p-value (bottom) for each interaction.

5. K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and bayesian inference in neuroimaging: Theory. *Neuroimage*, 16:465–483, 2002.
6. A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, June 2004.
7. L. Harrison, W. D. Penny, and K. J. Friston. Multivariate autoregressive modeling of fmri time series. *NeuroImage*, 19:1477–1491, 2003.
8. G. E. Hinton and D. V. Camp. Keeping neural networks simple by minimizing the description length of the weights. *Proceedings of the COLT'93*, pages 5–13, 1993.
9. M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer functions and statistical assessment of significance. *Biological Cybernetics*, 85:145–157, 2001.
10. H. Lappalainen and J. W. Miskin. Ensemble learning. *Advances in Independent Components Analysis*, 2000.
11. D. J. C. MacKay. Bayesian non-linear modeling for the energy prediction competition. *ASHRAE Transactions*, 100:1053–10062, 1994.
12. D. J. C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
13. J. R. Mangus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics (Revised Edition)*. John Wiley & Sons, 1999.
14. A. R. McIntosh and F. Gonzalez-Lima. The application of structural equation modeling to metabolic mapping of functional neural systems. *HBM*, 2:2–22, 1994.
15. X.-L. Meng and D. van Dyk. Fast em-type implementations for mixed effects models. *J. Royal Statist. Soc. Ser. B*, 60(3):559–578, 1998.
16. W. D. Penny and S. J. Roberts. Bayesian multivariate autoregressive models with structured priors. *IEE Proc.-Vis. Image Signal Process.*, 149(1), February 2002.
17. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.
18. X. L. Zhang, H. Begleiter, B. Porjesz, W. Wang, and A. Litke. Event related potentials during object recognition taks. *Brain Research Bulletin*, 38(6), 1995.