Graphical models for functional connectivity

Vicente Malave

October 6, 2008

1 Introduction

A lack of functional integration has been proposed as a signature of many degenerative and developmental disorders, including schizophrenia and autism (Just et al., 2004). These connectivity patterns are often assessed with ad-hoc techniques based on pairwise correlations. The procedures may lead to a misleading characterization of the underlying neural connectivity pattern, because these techniques cannot distinguish direct connections from mediated connections. More powerful statistical techniques exist which can sometimes uncover a model reflecting these distinctions. Although only a handful of papers to date have applied graphical model techniques to fMRI, they are often used in bioinformatics, which shares some statistical issues with fMRI data, including a small number of observations (i.e. hundreds) with a large number of variables (thousands). The proposed project will examine existing techniques for learning functional structure from fMRI time series, focusing on methods from the graphical models framework. The major part of the project will compromise a review of existing techniques, evaluation using real and simulated fMRI datasets, and work on algorithms which can scale up to a large number of regions than usually considered in functional connectivity analysis.

The remainder of this proposal is organized as follows: Section 2 explains the problem with correlation-based approaches. Section 3 contains a brief overview of graphical models and algorithms for structure learning. Section 4 summarizes previous literature in structure learning for fMRI. Section 5 summarizes work already done on this project and Section 6 outlines proposed extensions and comparisons to existing methods. Section 7 contains predictions and open questions to be addressed during this project.

2 Problems with pairwise correlations

Many papers employ the following technique: first all pairwise connections are computed, and then subjected to some subsequent analysis such as MDS (Welchew et al., 2005, 2002), graph-theoretic approaches (Achard et al., 2006),or factor analysis.

Correlation, however, is a poor source of information about statistical dependence between regions, because it neglects the influence of the other variables. An alternative measure is the partial correlation coefficient, which instead encodes *conditional* independence relationships. Partial correlation is the correlation between the residuals of two variables, after regressing out all the other variables. This measure was proposed for fMRI in Marrelec et al. (2007, 2006), for MEG by Langheim et al. (2006) and related models were presented in Salvador et al. (2005); Battle et al. (2007).

3 Graphical Modeling Methods

When the prior knowledge can identify the existence and nature of connections between regions, techniques such as structural equation modeling (SEM) or dynamic causal modeling DCM (Friston et al., 2003) can be used to assess the strength of a connection.

Many times there is insufficient prior knowledge to completely specify a model. Data-driven approaches for exploratory modeling of fMRI include Granger Causality (Roebroeck et al., 2005), Structural Equation Modeling (Storkey et al., 2007; Bullmore et al., 2000; Stein et al., 2007), Bayesian Networks (Zheng and Rajapakse, 2006) and Dynamic Bayesian Networks (Zhang et al., 2006; Li et al., 2008b; Rajapakse and Zhou, 2007). Working from different assumptions, all of these approaches lead to a best-fit model identifying the existence and direction of connections between regions of interest. Probabilistic graphical models form a general framework which includes structural equation models, Bayesian networks, and many other measures of connectivity previously applied to fMRI, including factor analysis, principal components analysis, independent components analysis and clustering (Roweis and Ghahramani, 1999).

These models define a probability distribution on directed graph. Significant interactions between variables are expressed in the model as edges in the graph. Edges encode information about the existence of a probabilistic dependence between the variables in the model, and for directed models also encode the direction of the causal interaction.

3.1 Structure Learning Algorithms

3.1.1 Search-and-score algorithms

One approach to structure learning is to exhaustively test every possible model, which is feasible only for small numbers of variables. Because an overparametrised model may fit better, the "best" model is chosen based on some penalized weighing of model fit and model complexity. Exhaustive searches have been applied to fMIR models in several studies (Hanson et al., 2007; Bullmore et al., 2000; Zhuang et al., 2005). Because the space of possible models grows exponentially with the number of variables, for larger models it is necessary to use approximate search procedures (Storkey et al., 2007; Zheng and Rajapakse, 2006; Zhang et al., 2006; Li et al., 2008b; Rajapakse and Zhou, 2007).

3.1.2 Constraint-based approaches

Another approach to structure learning is the use of constraint-based models,yielding the closely related PC and IC algorithms Spirtes et al. (2000); Pearl (2000) . These models conduct a number of statistical tests and then apply a set of structural rules to determine the existence and number of connections. Recent work has shown this algorithm to be feasible even in high-dimensional spaces (Kalisch and Bühlmann, 2007).

3.1.3 LiNGAM Algorithm

Gaussian distributed data leads to several models which can fit equally well due to identifiability problems. Howver, if all the underlying variables are non-Gaussian, it is possible to fully recover the model from the data. The LINGAM algorithm by Shimizu et al. (2006) uses Independent Component Analysis (Bell and Sejnowski, 1995; Comon et al., 1994), and some algebra to uncover a structural equation model from the data.

3.1.4 Graphical Gaussian Model

The graphical Gaussian model (Dempster, 1972; Whittaker, 1990; Lauritzen, 1996; Edwards, 2000) models the data as multivariate Gaussian, but constrains the inverse of the covariance matrix to have a zero for all pairs of variables which are conditionally independent.

The inverse of the covariance matrix $D = \Sigma^{-1}$ is known as the concentration or precision matrix.

The partial correlation between two variables Π_{ij} can be calculated from the precision matrix as

$$\Pi_{ij} = -\frac{d_{ij}}{\sqrt{d_{ii}d_{jj}}}$$

A zero in the off-diagonal entries of D, d_{ij} corresponds to conditional independence, ie.

$$X_i \perp\!\!\!\perp X_j | X_{\setminus ij}$$

Recently there has been some work on fast estimation of larger version of these models, following an influential paper by Meinshausen and Buhlmann (2006).

Undirected models may be a better model for fMRI time series data as the causal relationships between regions occur at a much faster temporal rate than we can hope to observe using the BOLD response. Undirected connections may also be used as a seed for further analysis with structural equation modeling or other techniques. The relationship between partial correlation, graphical Gaussian models and structural equation models is discussed in (Kiiveri and Speed, 1982) and (Marrelec et al., 2005).

4 fMRI Applications

Several data models have been applied to functional MRI. Here I summarize those which can be described as graphical models (structural equation models and Bayesian networks), and describe attempts to learn the structure of the underlying model. Granger causality is not discussed here. The Dynamic Causal Modeling framework (Friston et al., 2003) can be seen as a special case of a Dynamic Bayesian network, and structure learning techniques should be applicable to this model although this has not been done to date.

4.1 SEM Learning

The earliest paper I have found to date learning structure from fMRI time series is Bullmore et al. (2000). The method first sets all SEM coefficients to zero, and then gradually allows them to become unconstrained. Several measures of model fit are considered, including Akaike Information criterion, Bollen parsimonious fit index, as well as a chi-square test. A more recent study with a similar method Stein et al. (2007) use simulated annealing to learn SEMs. This one again uses parsimonious fit index, but the space of networks searched is also constrained by using (macaque) anatomical information. This is the only paper I have seen to date using anatomical information as a prior. For a small network (5 regions), Zhuang et al. (2005) exhaustively tested every possible model. Goodness of fit was tested using Adjusted Goodness of Fit Index and comparing residuals.Storkey et al. (2007) present a Bayesian approach to SEM learning for fMRI. In this case a Markov Chain Monte Carlo (MCMC) method is used to sample the space of possible networks. The authors note that no checks for cyclic structure are necessary in this approach, contrasting to searches for DAGs.

4.2 Learning Bayesian Networks

Zhang et al. (2006) present two algorithms for learning dynamic Bayesian networks, loosely characterized as a hill-climbing and an expectation maximization (EM) approach. Model fit was assessed using Bayesian Information Criterion, but the main evaluation is ability of learned networks to distinguish cocaine-addicted from control subjects.

Zheng and Rajapakse (2006) learn Bayesian networks using an MCMC approach and evaluating fit using Bayesian Information Criterion. This paper is a bit sparse on methodological details. The authors note that they prefer search-and-score methods is preferred to constraint-based approaches because constraint-based approaches conduct multiple independence tests and lose statistical power. A follow-up by this group extends to method to dynamic Bayesian networks Rajapakse and Zhou (2007). Again an MCMC/BIC search is used, and there is a (brief) comparison to granger causality.

(Li et al., 2007, 2006) have a series of papers applying dynamic bayesian networks, leading up to a Neuroimage paper considering how to best model

individual differences Li et al. (2008b). The models considered include "virtual-typical subject" (pooling all data and learning a single model ignoring individual variability), "individual subjects" (learning a different model for each subject) and "common structure" (same structure for each model, but parameters are allowed to vary between subjects). The most recent publication from this group applies false discovery rate to detection of connections Li et al. (2008a).

5 Work To date

To date, three methods have been investigated using an unpublished dataset from Deborah Harrington's laboratory. These methods included correlations and MDS, the LiNGAM algorithm, and the graphical Gaussian model.

5.1 Subjects

There were 19 control subjects, and 21 subjects with early-onset Parkinson's disease. The Parkinson's group were scanned twice, on and off medication.

5.2 Experimental Task

The task involves perceiving either a visual or auditory stimulus and then a second stimulus after 5 seconds, and the participant needs to make a response as to whether the second stimulus is longer or shorter than the first. Each pair of events (encode and decide) is always constrained to the same modality, and there are 60 visual and 60 auditory blocks. After extraction, each ROI timecourse was linearly detrended within blocks and scaled to have mean zero and variance one.

5.3 Correlations and Multidimensional Scaling

For exploratory analysis, we computed all cross-correlations between ROIs. The cross correlation matrix is then reshaped into a vector and submitted to an MDS analysis, which yields a clear group difference. Similar analyses were performed in (Welchew et al., 2005, 2002). To measure significance of connection differences, correlations were converted to z-scores using Fisher's r-to-z transformation. A t-test between groups was then performed. Connections shown in the graph indicate a significant difference in connection strengths between groups at a p-value of 0.01, and using the Bonferroni correction for multiple comparisons, yielding a total of 1198 significantly different interactions.

5.4 LiNGAM

MATLAB code is available for the LiNGAM algorithm from the authors. This was applied to the Parkinson's data as in the MDS section, learning a model from each individual subject. To compare LiNGAM models, I applied MDS to the (pruned) matrix of connection strengths returned by the LiNGAM algorithm.

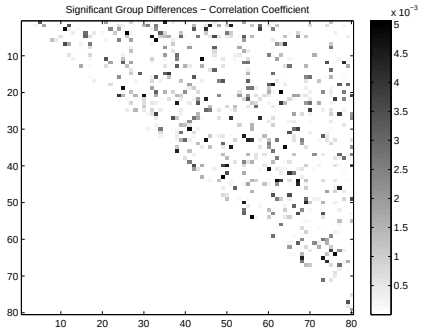

Figure 1: Significant Group Differences in Correlation

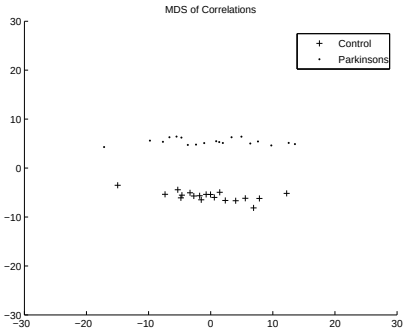

Figure 2: Multidimensional Scaling of Group Correlations

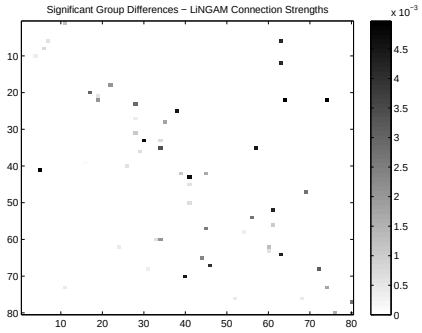

Figure 3: Significant Group Differences in Connection Strength

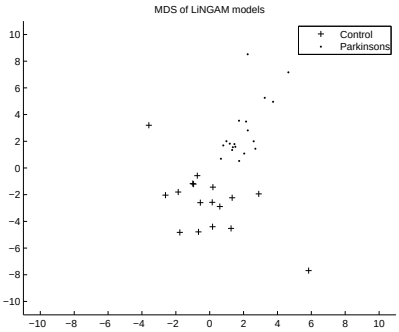

Figure 4: Multidimensional Scaling of LiNGAM models

This is similar to using the IS (individual subjects) method of Li et al. (2008b). The resulting models had 57 parameters which significantly differed between controls and the Parkinson's group. (Using a t-test on the pruned connection strength matrices).

5.5 Graphical Gaussian Model

In order to fit a graphical Gaussian model to our data, we need to consider model selection and fitting of a model. Model selection in this context is choosing which elements of the precision matrix to set to zero, and model fitting is estimating the covariance matrix, respecting the conditional independencies.

Correlations are tranformed to normally distributed variables using Fisher's r to z-transform:

$$z(i,j) = \frac{1}{2}\log\left(\frac{1+\Pi_{ij}}{1-\Pi_{ij}}\right)$$

The test then is to see if the partial correlation significantly differs from zero:

$$\sqrt{N-k-3}|z(i,j)| \leq \Phi^{-1}(1-\alpha/2)$$

Φ is the cdf of the standard normal distribution, N is the sample size, k is the order of the correlation. After computing all partial correlations the graph is thresholded to yield a conditional independence graph. To correct for multiple comparisons, we use the False Discovery Rate Genovese et al. (2002); Benjamini and Hochberg (1995).

This yields an "individual subjects" model of functional connectivity. For group analysis, we conduct several additional hypothesis tests.

A binomial test was used to detect edges which exist in a majority of subjects within each group. These edges are shown in the figures. To find edges which reliably distinguish subjects, the distribution of edges is compared to a binomial distribution. Edges which are significant (corrected with FDR) are shown in Figure. These edges indicate those which reliably distinguish groups, but they are only a subset of the entire conditional independence graph. The subset of edges which pass both binomial tests are colored in the figure: this indicates that they are significantly present in the group show, and significantly different in this group from the number of edges present in the other group.

To facilitate comparison with correlation based approaches, we also conducted one-sided t-tests of the differences in correlations and partial correlations between groups.

Once the model is chosen, the covariance matrix has to be fit to this data, respecting the conditional independence relationships encoded in the inverse covariance. An iterative algorithm is given by Speed and Kiiveri (1986) and we make use of the implementation in the R package ggm.

6 Results

Here we present the results of three comparisons: Patients on and off medication, Patients off medication vs controls, and patients on medication vs controls.

All of the methods were able to detect differences between the control and Parkinson's groups. Although both the correlation and partial correlation models are simple to compute and explain, the interpretation differs. One issue with the partial correlation analysis is that the direction of the correlations can reverse in the partial correlations. The LiNGAM algorithm runs for a long time, which may indicate an inability to scale to a larger number of regions. This is largely due to the large number of significance tests, which were performed using the bootstrap method implemented in the software. It may be possible to improve this algorithm by imposing a sparsity constraint on the inverse of the mixing matrix returned by the ICA algorithm.

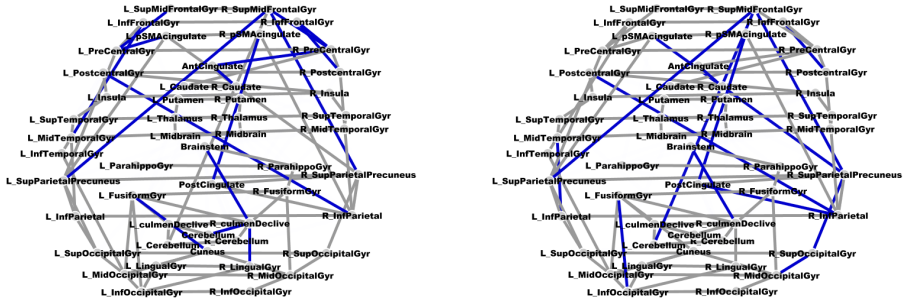

Figure 5: Control vs PD-Off and Control vs PD-On Decide Phase

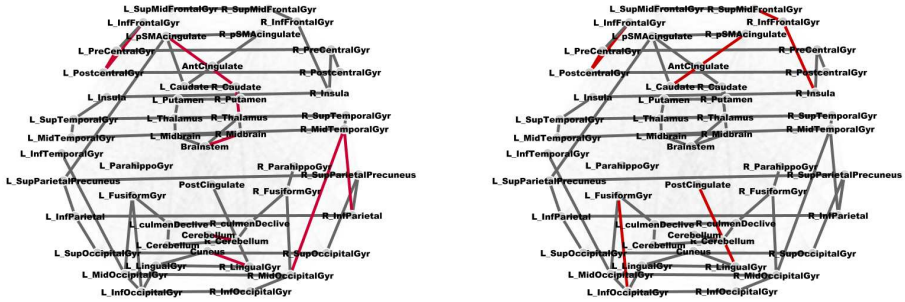

Figure 6: PD-Off vs Control and PD-Off vs PD-On Decide Phase

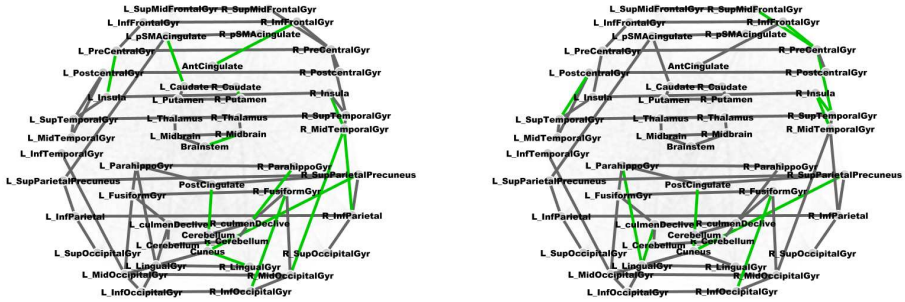

Figure 7: PD-On vs Control and PD-On vs PD-Off Decode Phase

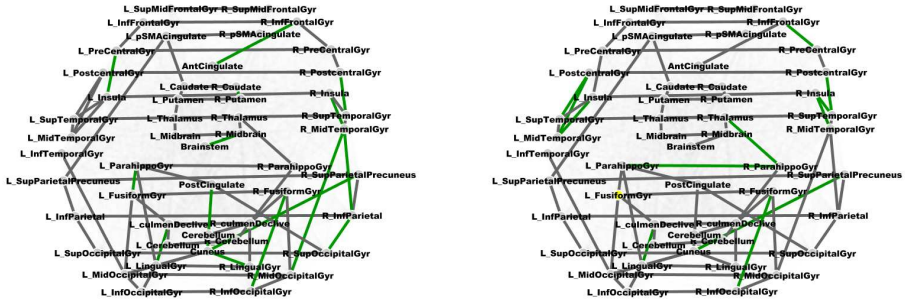

Figure 8: PD-On vs Control and PD-On vs PD-Off Encode Phase

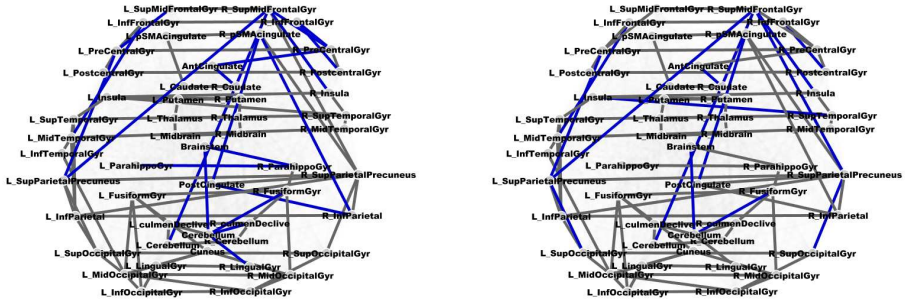

Figure 9: Control vs PD-Off and Control vs PD-On Encode Phase

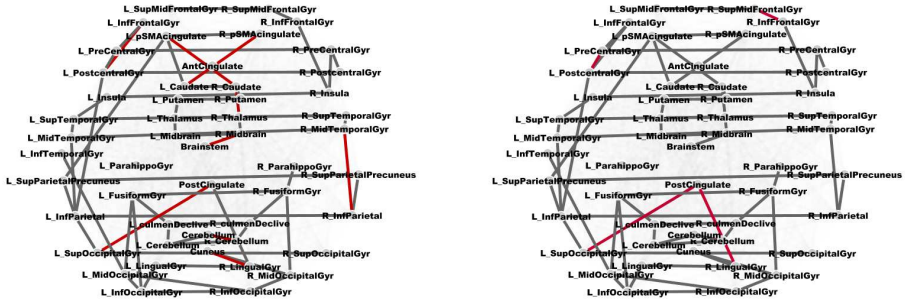

Figure 10: PD-Off vs Control PD-Off vs PD-On Encode Phase

7 Proposed Work

The main focus of this project is on a critical review and evaluation of these algorithms, and comparison to other methods for functional connectivity. Although comparisons exist in the literature, some of the algorithms and models considered (graphical Gaussian model, LINGAM) are novel to fMRI, and existing literature reviews often only compare 2-3 models, and contain misleading or incorrect characterizations of some techniques. Another area of exploration on this project is the ability of algorithms to scale to whole-brain connectivity.

7.1 Datasets

It is important that the data used correspond to well-known functional pathways. Three datasets will be used: data from Ishai et al. (2000) studying visual object recognition, and two unpublished datasets, the Parkinson's dataset previously described and another from Angela Yu. The latter two datasets allows a particularly powerful means to test these methods because subjects were run on and off medications affecting neurotransmitters. This allows a within-subjects comparison of the effects of medication: the drug effects are known, and should be detected by functional connectivity methods, while removing potential confounds due to differences in subjects or tasks.

7.1.1 Ishai

This dataset comes from a study of visual object representation(Ishai et al., 2000). As one of the first publicly-available datasets, this data has been reanalyzed in seval publications, and involves the highly studied visual object recognition system. Models applied to this dataset are easily reproducible, and we have strong prior expectations about what networks the method should recover.

7.1.2 Parkinson's Dataset

The dataset used comes from a study (in preparation) from Deborah Harrington's laboratory. This dataset is described in section. A feature of this dataset is that the Parkinson's group were run on and off dopamine-replacement medication. This dataset also includes Diffusion Tensor Imaging of each subject, which allows estimation of the number of anatomical connections between regions. Another possible validation of the model is to see if the methods find connections that respect known connectivity, or even to include this information in the search procedure as prior information.

7.1.3 Angela Yu Dataset

This data comes from an unpublished study by Angela Yu. The task involved decision making under uncertainty. Subjects were administered placebo, cloni-

dine (affecting norepinephrine) and scopolamine (affecting acetylcholine) during experimental sessions.

7.2 Experiment 1 : Simulation Study

The first experiment will involve an evaluation of structure-learning techniques using simulated data. An extensive comparison of methods appears in Tsamardinos et al. (2006). The proposed experiment will involve a smaller subset of algorithms, and will evaluate them with simulated fMRI time series. An important outcome will be investigating the effects of preprocessing, including removing autoregressive trends from the data. Algorithms will be evaluated based on their ability to identify the existence of connections, and for directed models, the direction of connections.

7.3 Experiment 2 : fMRI application

Methods which are show to perform well on simulated data will then be applied to the fMRI datasets. Algorithms will be evaluated for ability to discover pathways which correspond to known anatomical connections, and areas known to be involved in the task. The fMRI data will be reduced to a set of timecourses following the standard practices (i.e. using the mean of a functionally or anatomically defined region of interest), resulting in a few dozen timecourses or less. Another issue to investigate in this experiment is reproducibility of these results should also be evaluated, for example using split-half resampling on fMRI datasets to evaluate methods. A related issue is the impact of preprocessing steps, including detrending, highpass filtering, and the means of generating an average timecourse from a region of interest. Finally, the two datasets with drug and disease effects allow assessment of ability to uncover these effects within the same subject subject to manipulation of neurotransmitters. The methods in these cases should identify correctly the same medication effects across subjects.

7.4 Experiment 3

Another direction of this research is investigation of the ability of these methods to scale to a large number of variables. A promising set of methods to scale to full-brain conenctivity are l_1 regularized Gaussian graphical model techniques.

The idea of the l_1 or lasso technique is best explained in linear regression, where the objective is to find the β which minimizes the sum-squared error

$$||Y - X\beta||_2^2$$

The Lasso (Tibshirani, 1996) additionally penalizes the absolute value of the weights.

$$||Y - X\beta||_2^2 + \theta \sum_{j=1}^{p} |\beta_j|$$

This penalty on weights leads to a model which sets many of the coefficients to zero, leading to a more parsimonious model and helping to control overfitting when the number of regressors is large. Meinshausen and Buhlmann (2006) prove that under a few assumptions,the lasso solution (regressing each variable against all others) will find the correct pattern of conditional independence in the graphical Gaussian Model. Since then there have been several fast algorithms using convex optimization. D'Aspremont et al. (2008); Banerjee and El Ghaoui (2008); Banerjee and Natsoulis (2006); Duchi et al. (2008); Yuan and Lin (2007). Friedman et al. (2007b) impose a lasso penalty and use a coordinate wise descent procedure from Friedman et al. (2007a) to make it fast for large models.

As noted above, it may be possible to extend this insight to the LiNGAM algorith, which uses the inverse of the ICA mixing matrix as a step in the structure discovery algorithm.

Several other algorithms perform well in low-sample, high dimensional situations including the PC algorithm, and similar approximate constraint-based approached (Kalisch and Bühlmann, 2007; Castelo and Roverato, 2006; Opgen-Rhein and Strimmer, 2007).

8 Potential Outcomes and Contributions

The results of the first experiment and a thorough literature review will indicate how well structures may be recovered from fMRI data. An important result will be quantifying the ambiguity of these methods, which is an issue not discussed much to data in this literature.

Subsequent evaluation on fMRI imaging data will also allow comparisons of methods. The unique ability to compare methods within-subjects to discover neurotransmitter effects allows for a comparison with fewer confounding effects, and a measure of the ability of the methods to uncover effects which should lead to a change in the functional interaction between brain regions.

The third experiment applying structure learning at the voxel level should reveal similar results at a gross level to the resuls from using an average time course from an ROI. If the results differ substantially from the results obtained using an average timecours from regions of interest, this may indicate that this methodology is flawed, and may suggest improved means of summarizing a region of interest.

The most important contribution of this project will be in summarizing these methods in an accessible form, and explaining how they differ from correlation-based approaches widely used. This may lead to different interpretation of the results from functional connectivity studies to date.

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E., 2006. A Resilient, Low-Frequency, Small-World Human Brain Functional Network with

Highly Connected Association Cortical Hubs. Journal of Neuroscience 26 (1), 63.

Banerjee, O., El Ghaoui, L., 2008. Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. Journal of Machine Learning Research 9, 485–516.

Banerjee, O., Natsoulis, G., 2006. Convex optimization techniques for fitting sparse Gaussian graphical models. Proceedings of the 23rd international conference on Machine learning, 89–96.

Battle, A., Chechik, G., Koller, D., 2007. Temporal and Cross-Subject Probabilistic Models for fMRI Prediction Tasks. Advances in Neural Information Processing Systems: Proceedings of the 2006 Conference.

Bell, A., Sejnowski, T., 1995. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation 7 (6), 1129–1159.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 57 (1), 289–300.

Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., Sharma, T., 2000. How Good Is Good Enough in Path Analysis of fMRI Data? Neuroimage 11 (4), 289–301.

Castelo, R., Roverato, A., 2006. A Robust Procedure For Gaussian Graphical Model Search From Microarray Data With p Larger Than n. The Journal of Machine Learning Research 7, 2621–2650.

Comon, P., et al., 1994. Independent component analysis, a new concept. Signal Processing 36 (3), 287–314.

D'Aspremont, A., Banerjee, O., El Ghaoui, L., 2008. First-Order Methods for Sparse Covariance Selection. SIAM Journal on Matrix Analysis and Applications 30, 56.

Dempster, A., 1972. Covariance selection. Biometrics 28 (1), 157–175.

Duchi, J., Gould, S., Koller, D., 2008. Projected Subgradient Methods for Learning Sparse Gaussians. UAI.

Edwards, D., 2000. Introduction to Graphical Modelling. Springer.

Friedman, J., Hastie, T., Hoefling, H., Tibshirani, R., 2007a. Pathwise coordinate optimization. Annals of Applied Statistics, to appear.

Friedman, J., Hastie, T., Tibshirani, R., 2007b. Sparse inverse covariance estimation with the graphical lasso. Biostatistics.

Friston, K., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19 (4), 1273–1302.

Genovese, C., Lazar, N., Nichols, T., 2002. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. Neuroimage 15 (4), 870–878.

Hanson, S., Hanson, C., Halchenko, Y., Matsuka, T., Zaimi, A., 2007. Bottom-up and top-down brain functional connectivity underlying comprehension of everyday visual action. Brain Structure and Function 212 (3), 231–244.

Ishai, A., Ungerleider, L., Martin, A., Haxby, J., 2000. The Representation of Objects in the Human Occipital and Temporal Cortex. Journal of Cognitive Neuroscience 12 (90002), 35–51.

Just, M., Cherkassky, V., Keller, T., Minshew, N., 2004. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. Brain 127 (8), 1811.

Kalisch, M., Bühlmann, P., 2007. Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm. The Journal of Machine Learning Research 8, 613–636.

Kiiveri, H., Speed, T., 1982. Structural analysis of multivariate data: A review. Sociological Methodology 13, 209–289.

Langheim, F., Leuthold, A., Georgopoulos, A., 2006. Synchronous dynamic brain networks revealed by magnetoencephalography. Proc Natl Acad Sci US A 103 (2), 455–459.

Lauritzen, S., 1996. Graphical Models. Oxford University Press, USA.

Li, J., Wang, Z., McKeown, M., 2006. Dynamic Bayesian networks (DBNS) demonstrate impaired brain connectivity during performance of simultaneous movements in Parkinson's disease. Biomedical Imaging: Macro to Nano, 2006. 3rd IEEE International Symposium on, 964–967.

Li, J., Wang, Z., McKeown, M., 2007. A Multi-Subject, Dynamic Bayesian Networks (DBNS) Framework for Brain Effective Connectivity. Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on 1.

Li, J., Wang, Z., McKeown, M., 2008a. Controlling the false discovery rate in modeling brain functional connectivity. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, 2105–2108.

Li, J., Wang, Z., Palmer, S., McKeown, M., 2008b. Dynamic Bayesian network modeling of fMRI: A comparison of group-analysis methods. Neuroimage.

Marrelec, G., Doyon, J., Pelegrini-Issac, M., Benali, H., 2005. Heading for data-driven measures of effective connectivity in functional MRI. Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on 3.

Marrelec, G., Horwitz, B., Kim, J., Pélégrini-Issac, M., Benali, H., Doyon, J., 2007. Using partial correlation to enhance structural equation modeling of functional MRI data. Magnetic Resonance Imaging 25 (8), 1181–1189.

Marrelec, G., Krainik, A., Duffau, H., Pélégrini-Issac, M., Lehéricy, S., Doyon, J., Benali, H., 2006. Partial correlation for functional brain interactivity investigation in functional MRI. Neuroimage 32 (1), 228–237.

Meinshausen, N., Buhlmann, P., 2006. High dimensional graphs and variable selection with the lasso. Annals of Statistics 34 (3), 1436–1462.

Opgen-Rhein, R., Strimmer, K., 2007. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. BMC Systems Biology 1 (1), 37.

Pearl, J., 2000. Causality: Models, Reasoning, and Inference. Cambridge University Press.

Rajapakse, J., Zhou, J., 2007. Learning effective brain connectivity with dynamic Bayesian networks. Neuroimage 37 (3), 749–760.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. Neuroimage 25 (1), 230–242.

Roweis, S., Ghahramani, Z., 1999. A Unifying Review of Linear Gaussian Models. Neural Computation 11 (2), 305–345.

Salvador, R., Suckling, J., Schwarzbauer, C., Bullmore, E., 2005. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Philosophical transactions-Royal Society of London. Biological sciences 360 (1457), 937–946.

Shimizu, S., Hoyer, P., Hyvärinen, A., Kerminen, A., 2006. A Linear Non-Gaussian Acyclic Model for Causal Discovery. The Journal of Machine Learning Research 7, 2003–2030.

Speed, T., Kiiveri, H., 1986. Gaussian Markov Distributions over Finite Graphs. Ann. Statist 14 (1), 138–150.

Spirtes, P., Glymour, C., Scheines, R., 2000. Causation, Prediction, and Search. MIT Press.

Stein, J., Wiedholz, L., Bassett, D., Weinberger, D., Zink, C., Mattay, V., Meyer-Lindenberg, A., 2007. A validated network of effective amygdala connectivity. Neuroimage 36 (3), 736–745.

Storkey, A., Simonotto, E., Whalley, H., Lawrie, S., Murray, L., McGonigle, D., 2007. Learning Structural Equation Models for fMRI. ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 19, 1329.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58 (1), 267–288.

Tsamardinos, I., Brown, L., Aliferis, C., 2006. The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning 65 (1), 31–78.

Welchew, D., Ashwin, C., Berkouk, K., Salvador, R., Suckling, J., Baron-Cohen, S., Bullmore, E., 2005. Functional disconnectivity of the medial temporal lobe in Aspergers syndrome. Biological Psychiatry 57 (9), 991–998.

Welchew, D., Honey, G., Sharma, T., Robbins, T., Bullmore, E., 2002. Multidimensional Scaling of Integrated Neurocognitive Function and Schizophrenia as a Disconnexion Disorder. Neuroimage 17 (3), 1227–1239.

Whittaker, J., 1990. Graphical models in applied multivariate statistics. Wiley New York.

Yuan, M., Lin, Y., 2007. Model selection and estimation in the Gaussian graphical model. Biometrika 94 (1), 19–35.
URL http://biomet.oxfordjournals.org/cgi/content/abstract/94/1/19

Zhang, L., Samaras, D., Alia-Klein, N., Volkow, N., Goldstein, R., 2006. Modeling neuronal interactivity using dynamic bayesian networks. Advances in neural information processing systems 18, 1595–1602.

Zheng, X., Rajapakse, J., 2006. Learning functional structure from fMR images. Neuroimage 31 (4), 1601–1613.

Zhuang, J., LaConte, S., Peltier, S., Zhang, K., Hu, X., 2005. Connectivity exploration with structural equation modeling: an fMRI study of bimanual motor coordination. Neuroimage 25 (2), 462–470.