Chapter 40: Multivariate autoregressive models

W. Penny and L. Harrison

April 28, 2006

Introduction

Functional neuroimaging has been used to corroborate functional specialisation as a principle of organization in the human brain. However, disparate regions of the brain do not operate in isolation and more recently neuroimaging has been used to characterise the network properties of the brain under specific cognitive states [Buchel et al. 1997a, Buchel and Friston 2000]. These studies address a complementary principle of organization, functional integration.

Functional MRI provides a unique opportunity to observe simultaneous recordings of activity throughout the brain evoked by cognitive and sensorimotor challenges. Each voxel within the brain is represented by a time series of neurophysiological activity that underlies the measured BOLD response. Given these multivariate, voxel-based time series, can we infer large-scale network behaviour among functionally specialised regions?

A number of methods have been proposed to answer this question including regression models [McIntosh et al. 1994, Friston et al. 1997, Friston et al. 1993, Friston et al. 1995], convolution models [Friston 2001, Friston and Buchel 2000] and state-space models [Buchel et al. 1998]. Regression techniques, underlying eg. the analysis of Psychophysiological Interactions (PPIs), are useful becasue they are easy to fit and can test for the modulatory interactions of interest [Friston et al. 1997]. However, this is at the expense of excluding temporal information, i.e. the history of an input or physiological variable. This is important as interactions within the brain, whether over short or long distances, take time and are not instantaneous. Structural Equation Modeling (SEM), as used by the neuroimaging community [McIntosh et al. 1994, Buchel et al. 1997b] has similar problems [1]. Convolution models, such as the Volterra approach, model temporal effects in terms of an idealized response characterized by the kernels of the model [Friston 2000]. A criticism of the Volterra approach is that it treats the system as a black box, meaning that it has no model of the internal mechanisms that may generate data. State-Space Models account for correlations within the data by invoking state variables, whose dynamics generate data. Recursive algorithms, such as the Kalman Filter, may be used to estimate these states through time, given the data [Buchel et al. 1998].

This chapter describes an approach based on Multivariate Autoregressive (MAR) models. These are linear multivariate time series models which have a

[1] There exist versions of SEM that do model dynamic information, see [Cudeck 2002] for details of Dynamic Factor Analysis.

long history of application in econometrics. The MAR model characterises inter-regional dependencies within data, specifically in terms of the historical influence one variable has on another. This is distinct from regression techniques that quantify instantaneous correlations. We use MAR models to make inferences about functional integration from fMRI data.

The chapter is divided into 3 sections. First we describe the theory of MAR models, parameter estimation, model order selection and statistical inference. We have used a Bayesian technique for model order selection and parameter estimation which is introduced in Chapter 24 and is described fully in [Penny and Roberts 2002]. Second, we model neurophysiological data taken from an fMRI experiment addressing attentional modulation of cortical connectivity during a visual motion task [Buchel et al. 1997b]. The modulatory effect of one region upon the responses to other regions is a second order interaction which is precluded in linear models. To circumvent this we have introduced bilinear terms [Friston et al. 1997]. Thirdly, we discuss the advantages and disadvantages of MAR models, its use in spectral estimation and possible future developments.

Theory

Multivariate Autoregressive Models

Given a univariate time series, its consecutive measurements contain information about the process that generated it. An attempt at describing this underlying order can be achieved by modelling the current value of the variable as a weighted linear sum of its previous values. This is an Autoregressive (AR) process and is a very simple, yet effective, approach to time series characterisation [Chatfield 1996]. The order of the model is the number of preceding observations used, and the weights characterise the time series.

Multivariate Autoregressive models extend this approach to multiple time series so that the vector of current values of all variables is modelled as a linear sum of previous activities. Consider d time series generated from d variables within a system such as a functional network in the brain and where m is the order of the model. A MAR(m) model predicts the next value in a d-dimensional time series, y_n as a linear combination of the m previous vector values

$$y_n = \sum_{i=1}^{m} y_{n-i} A(i) + e_n \tag{1}$$

where $y_n = [y_n(1), y_n(2), ..., y_n(d)]$ is the nth sample of a d-dimensional time series, each $A(i)$ is a d-by-d matrix of coefficients (weights) and $e_n = [e_n(1), e_n(2), ..., e_n(d)]$ is additive Gaussian noise with zero mean and covariance R. We have assumed that the data mean has been subtracted from the time series.

The model can be written in the standard form of a multivariate linear regression model as follows

$$y_n = x_n W + e_n \tag{2}$$

where $x_n = [y_{n-1}, y_{n-2}, ..., y_{n-m}]$ are the m previous multivariate time series samples and W is a $(m \times d)$-by-d matrix of MAR coefficients (weights). There are therefore a total of $k = m \times d \times d$ MAR coefficients.

If the nth rows of Y, X and E are y_n, x_n and e_n respectively and there are $n = 1..N$ samples then we can write

$$Y = XW + E \tag{3}$$

where Y is an $(N-m)$-by-d matrix, X is an $(N-m)$-by-$(m \times d)$ matrix and E is an $(N-m)$-by-d matrix. The number of rows $N-m$ (rather than N) arises as samples at time points before m do not have sufficient preceding samples to allow prediction.

MAR models are fully connected in that each region is, by default, assumed connected to all others. However, by fitting the model to data and testing to see which connections are signficantly non-zero, one can infer a sub-network that mediates the observed dynamics. This can be implemented using Bayesian inference as described below.

These sub-networks are related to the concept of 'Granger causality' [Granger 1969], which is defined operationally as follows. Activity in region X 'Granger causes' activity in region Y if any of the connections from X to Y, over all time lags, are non-zero. These causality relationships can be summarised by directed graphs as described in [Eichler 2005]. An example will be presented later on in the chapter.

Nonlinear autoregressive models

Given a network model of the brain we can think of two fundamentally different types of coupling; linear and nonlinear. The model discussed so far is linear. Linear systems are described by the principle of superposition, which is that inputs have additive effects on the response that are independent of each other. Nonlinear systems are characterised by inputs which interact to produce a response.

In [Buchel et al. 1997b], nonlinear interactions were modelled using 'bilinear terms'. This is the approach adopted in this chapter. Specifically, to model a hypothesized interaction between variables $y_n(j)$ and $y_n(k)$ one can form the new variable

$$I_n(j,k) = y_n(j)y_n(k) \tag{4}$$

This is a 'bilinear variable'. This is orthogonalised with respect to the original time series and placed in an augmented MAR model with connectivity matrices $\tilde{A}(i)$.

$$[y_n, I_n(j,k)] = \sum_{i=1}^{m} [y_{n-i}, I_{n-i}(j,k)]\tilde{A}(i) + e_n \tag{5}$$

The relevant entries in $\tilde{A}(i)$ then reflect modulatory influences eg. a change of the connection strength between $y(j)$ and other time series due to the influence of $y(k)$.

It should be noted that each bilinear variable introduces only one of many possible sources of nonlinear behaviour into the model. The example above specifically models nonlinear interactions between $y_n(j)$ and $y_n(k)$, however other bilinear terms could involve, for instance, the time series $y_n(j)$ and inputs $u(t)$. The inclusion of these terms are guided by the hypothesis of interest eg. does 'time' change the connectivity between earlier and later stages of processing in the dorsal visual pathway? Here $u(t)$ would model time.

Maximum Likelihood Estimation

Reformulating MAR models as standard multivariate linear regression models allows us to retain contact with the large body of statistical literature devoted to this subject; see eg. p. 423 of [Tiao 1992]. The Maximum Likelihood (ML) solution (see eg. [Weisberg 1980]) for the MAR coefficients is

$$\hat{W} = (X^T X)^{-1} X^T Y \tag{6}$$

The maximum likelihood noise covariance, S_{ML}, can be estimated as

$$S_{ML} = \frac{1}{N-k}(Y - X\hat{W})^T(Y - X\hat{W}) \tag{7}$$

where $k = m \times d \times d$. We define $\hat{w} = vec(\hat{W})$ where vec denotes the columns of $\hat{W}$ being stacked on top of each other (for more on the vec notation, see [Muirhead 1982]). To recover the matrix $\hat{W}$ we simply 'un-stack' the columns from the vector $\hat{w}$.

The ML parameter covariance matrix for $\hat{w}$ is given by [Magnus and Neudecker 1997] (page 321)

$$\hat{\Sigma} = S_{ML} \otimes (X^T X)^{-1} \tag{8}$$

where $\otimes$ denotes the Kronecker product (see, eg page 477 in [Tiao 1992]) The optimal value of m can be chosen using a model order selection criterion such as the Minimum Description Length (MDL). See eg. [Neumaier and Schneider 2000].

Bayesian Estimation

It is also possible to estimate the MAR parameters and select the optimal model within a Bayesian framework [Penny and Roberts 2002]. This has been shown to give better model order selection and is the approach used in this chapter. The maximum-likelihood solution is used to initialise the Bayesian scheme.

In what follows $\mathsf{N}(m, Q^{-1})$ is a multivariate Gaussian with mean m and precision (inverse covariance) Q. Also, $\mathsf{Ga}(b, c)$ is the Gamma distribution with parameters b and c defined in chapter 24. The gamma density has mean bc and variance b^2c. Finally, $\mathsf{Wi}(s, B)$ denotes a Wishart density [Tiao 1992]. The Bayesian model uses the following prior distributions

$$\begin{aligned} p(W|m) &= \mathsf{N}(0, \alpha^{-1} I) \\ p(\alpha|m) &= \mathsf{Ga}(b, c) \\ p(\Lambda|m) &= |\Lambda|^{-(d+1)/2} \end{aligned} \tag{9}$$

where m is the order of the model, α is the precision of the Gaussian prior distribution from which weights are drawn and Λ is the noise precision matrix (inverse of R). In [Penny and Roberts 2002] it is shown that the corresponding posterior distributions are given by

$$\begin{aligned} p(W|Y, m) &= \mathsf{N}(\hat{W}_B, \hat{\Sigma}_B) \\ p(\alpha|Y, m) &= \mathsf{Ga}(\hat{b}, \hat{c}) \\ p(\Lambda|Y, m) &= \mathsf{Wi}(s, B) \end{aligned} \tag{10}$$

The parameters of the posteriors are updated in an iterative optimisation scheme described in the Appendix. Iteration stops when the 'Bayesian evidence' for

model order m, $p(Y|m)$, is maximised. A formula for computing this is also provided in the appendix. Importantly, the evidence is also used as a model order selection criterion, that is, to select the optimal value of m. This is discussed at length in Chapters 24 and 35.

Bayesian Inference

The Bayesian estimation procedures outlined above results in a posterior distribution for the MAR coefficients $P(W|Y,m)$. Bayesian inference can then take place using confidence intervals based on this posterior. See, for example, page 84 of [Tiao 1992]. The posterior allows us to make inferences about the strength of a connection between two regions. Because this connectivity can be expressed over a number of time lags our inference is concerned with the vector of connection strengths, a, over all time lags. To make contact with classical (non-Bayesian) inference, we say that a connection is 'significantly non-zero' or simply 'significant' at level α if the zero vector lies outside the $1-\alpha$ confidence region for a. This is shown schematically in Figure 1.

Application

The responsiveness of dorsal visual brain regions in neuroimaging studies suggests attention is associated with changes in connectivity [Assad et al. 1995] [Craven and Savoy 1995]. In this chapter we use data from an fMRI study investigating attentional modulation of connectivity within the dorsal visual pathways [Buchel et al. 1997b]. This provides a testbed for assessing how MAR models estimate changes in connectivity.

In brief, the experiment was performed on a 2T MRI scanner. The visual stimulus involved random dots moving radially outwards at a fixed rate. Subjects were trained beforehand to detect changes in velocity of radial motion. Attentional set was manipulated by asking the subject to attend to changes in velocity or to just observe the motion. Both of these states were separated by periods of "fixation" where the screen was dark and only a fixation dot was visible. Each block ended with a "stationary" condition in which a static image of the previously moving dots was shown. Unknown to the subjects, the radial velocity remained constant throughout the experiment such that the only experimental manipulation was attentional set.

Categorical comparisons using General Linear Model (GLM) analyses (see eg. Chapter 8) were used to identify changes in brain activity dependent on attentional set. This revealed activations throughout right and left hemispheres in the primary visual cortex V1/2 complex, visual motion region V5 and regions involved in the attentional network including posterior parietal cortex (PPC) and in the right prefrontal cortex (PFC). Regions of interest (ROI) were defined with a diameter of 8mm centred around the most significant voxel and a representative time series was defined by the first eigenvariate of the region. For details of the experimental design and acquisition see [Buchel et al. 1997b]. We analyse data from three subjects. Time series from subject 1 are shown in Figure 2.

Inspecting the four time series reveals a number of characteristics. The time series from the V1/2 complex shows a dependence on the presentation of

the moving image with a small difference between attention and non-attention. However, in the higher brain areas of PPC and PFC attentional set is the dominant influence, with a marked increase in activity during periods of attention. The relative influence each region has on others is not obvious from visual inspection but, as we shall see, can be revealed from a MAR analysis.

Three models were tested using the regions and interaction terms shown below.

- Model 1: V1/V2, V5, PPC and PFC
- Model 2: V1/V2, V5 and $I_{v1,ppc}$
- Model 3: V5, PPC and $I_{v5,pfc}$

where $I_{v1,ppc}$ denotes an interaction between V1/V2 and PPC and $I_{v5,pfc}$ an interaction between V5 and PFC. These variables were created as described in the earlier section on nonlinear autoregressive models. The interaction terms can be thought as 'virtual' nodes in a network.

For each type of model, we computed the model evidence as a function of model order m. The results in Figure 3 show that the optimal order for all three models was $m = 4$ (subject 1). A model order of $m = 4$ was then used in the results reported below.

Model 1 was applied, to right hemisphere data only, to identify the functional network connecting key regions in the visual and attentional systems. Figure 4 shows connections in this model which, across all time lags, are significantly non-zero for subject 1. Over the three subjects, all V1/V2 to V5 connections ($\alpha < 0.0004$) and all PFC to PPC ($\alpha < 0.02$) connections were significant. We can therefore infer that activity in V1/V2 Granger causes activity in V5, and PFC Granger causes PPC. Also, the V5 to PPC connection was significant ($\alpha < 0.0009$) in two out of three subjects.

The second model was applied both to left and right hemisphere data. Figure 5 shows significantly non-zero connections for left hemisphere data from subject 1. This shows that activity in PPC changes the connectivity between V1/V2 and V5. The same was true for the other two subjects. For the right hemisphere data, however, only subject 1 showed an effect ($\alpha < 0.03$).

The third model was applied to right hemisphere data. Figure 6 shows significantly non-zero connections for data from subject 1. This shows that activity in PFC changes how PPC responds to V5. Subject 2 also showed this effect ($\alpha < 0.03$) but subject 3 did not.

Discussion

We have proposed the use of MAR models for making inferences about functional integration using fMRI time series. One motivation for this is that the previously dominant model used for making such inferences in the existing fMRI/PET literature, namely Structural Equation Modelling, as used in [McIntosh et al. 1994, Buchel et al. 1997b], is not a time series model. Indeed, inferences are based solely on the instantaneous correlations between regions - if the series were randomly permuted over time, SEM would give the same results. Thus SEM throws away temporal information.

Further, MAR models may contain loops and self-connections yet parameter estimation can proceed in a purely linear framework ie. there is an analytic solution that can be found via linear algebra. In contradistinction, SEM models with loops require non-linear optimisation. The reason for this is that MAR models do not contain instantaneous connections. The between-region connectivity arises from connections between regions at different time-lags. Due to temporal persistence in the activity of each region this captures much the same effect, but in a computationally simpler manner.

In this chapter we applied Bayesian MAR models to fMRI data. Bayesian inferences about connections were then made on the basis of the estimated posterior distribution. This allows for the identification of a sub-network of connections that mediate the observed dynamics. These connections describe causal relations, in the sense of Granger [Granger 1969].

This is in the spirit of how General Linear Models are used for characterising functional specialisation; all conceivable factors are placed in one large model and then different hypotheses are tested using t or F-contrasts [Frackowiak et al. 1997]. We note that this approach is fundamentally different to the philosophy underlying SEM. In SEM, only a few connections are modelled. These are chosen on the basis of prior anatomical or functional knowledge and are interpreted as embodying causal relations. Thus, with SEM, causality is ascribed a priori [Pearl 1998]. But with MAR, causality can be inferred from data.

MAR models can also be used for spectral estimation. In particular they enable parsimonious estimation of coherences (correlation at particular frequencies), partial coherences (the coherence between two time series after the effects of others have been taken into account), phase relationships [Marple 1987] [Cassidy and Brown 2000] and directed transfer functions [Kaminski et al. 1997]. MAR models have been used in this way to investigate functional integration from EEG and ECOG recordings [Bressler et al. 1999]. This provides a link with a recent analysis of fMRI data [Muller et al. 2001] which looks for sets of voxels that are highly coherent. MAR models provide a parametric way of estimating this coherence, although in this chapter we have reported the results in the time domain.

A further aspect of MAR models is that they capture only linear relations between regions. Following [Buchel et al. 1997b], we have extended their capabilities by introducing bilinear terms. This allows one to infer that activity in one region modulates connectivity between two other regions. Such inferences are beyond the current capabilities of Dynamic Causal Models for fMRI (see chapter 41).

It is also possible to include further higher order terms, for instance, second-order interactions across different lags. Frequency domain characterisation of the resulting models would then allow us to report bi-spectra [Priestley 1988]. These describe the correlations between different frequencies which may be important for the study of functional integration [Friston 2000] (see also Chapter 39).

A key aspect of our approach has been the use of a mature Bayesian estimation framework [Penny and Roberts 2002]. This has allowed us to select the optimal MAR model order. One promising direction for extending the model is to replace Gaussian priors with sparse priors. This would effectively remove most connections allowing the model to be applied to a very large number of regions. This approach has been investigated in a non-Bayesian

framework using penalised regression and pruning based on false discovery rates [Valdes-Sosa et al. 2005].

Appendix

Bayesian estimation

Following the algorithm developed in [Penny and Roberts 2002], the parameters of the posterior distributions are updated iteratively as follows.

$$\begin{aligned}
\Lambda_D &= \hat{\Lambda} \otimes (X^T X) \\
\hat{\Sigma}_B &= (\Lambda_D + \hat{\alpha} I)^{-1} \\
\hat{W}_B &= \hat{\Sigma}_B \Lambda_D \hat{W} \\
\frac{1}{\hat{b}} &= \frac{1}{2} \hat{W}_B^T \hat{W}_B + \frac{1}{2} \mathsf{Tr}(\hat{\Sigma}_B) + \frac{1}{b} \\
\hat{c} &= \frac{k}{2} + c \\
\hat{\alpha} &= \hat{b}\hat{c} \\
s &= N \\
B &= \frac{1}{2}(Y - X\hat{W}_B)^T (Y - X\hat{W}_B) \\
&+ \sum_n (I \otimes x_n) \hat{\Sigma}_B (I \otimes x_n)^T \\
\hat{\Lambda} &= sB^{-1}
\end{aligned} \tag{11}$$

The updates are initialised using the Maximum-Likelihood solution. Iteration terminates when the Bayesian log-evidence increase by less than 0.01%. The log-evidence is computed as follows

$$\begin{aligned}
\log p(Y|m) &= \frac{N}{2} \log |B| - KL_N(p(W|m), p(W|Y,m)) \\
&- KL_{Ga}(p(\alpha|m), p(\alpha|Y,m)) + \log \Gamma_d(N/2)
\end{aligned} \tag{12}$$

where KL_N and KL_{Ga} denote the Kullback-Liebler (KL) divergences for Normal and Gamma densities defined in chapter 24. Essentially, the first term in the above equation is an accuray term and the KL terms act as a penalty for model complexity. See chapters 24 and 35 for more on model comparison.

Testing the significance of connections

The connectivity between two regions can be expressed over a number of time lags. Therefore, to see if the connectivity is significantly non-zero we make an inference about the vector of coefficients a, where each element of that vector is the value of a MAR coefficient at a different time lag. First we specify $(k \times k)$ $(k = m \times d \times d)$ sparse matrix C such that

$$a = C^T w \tag{13}$$

returns the estimated weights for connections between the two regions of interest. For an MAR(m) model, this vector has m entries, one for each time-lag.

The probability distribution is given by $p(a) = \mathsf{N}(\mu, V)$ and is shown schematically in Figure 1. The mean and covariance are given by

$$\begin{aligned} \mu &= C^T \hat{w} \\ V &= C^T \hat{\Sigma}_B C \end{aligned} \tag{14}$$

where $\hat{w} = vec(\hat{W}_B)$ and $\hat{\Sigma}_B$ are the Bayesian estimates of the parameters of the posterior distribution of regression coefficients from the previous section. In fact, $p(a)$ is just that part of $p(w)$ that we are interested in.

The probability α that the zero vector lies on the $1-\alpha$ confidence region for this distribution is then computed as follows. We first note that this probability is the same as the probability that the vector m lies on the edge of the $1-\alpha$ region for the distribution $\mathsf{N}(0, V)$. This latter probabilitiy can be computed by forming the test statistic

$$d = \mu^T V^{-1} \mu \tag{15}$$

which will be the sum of $r = rank(V)$ independent, squared Gaussian variables. As such it has a χ^2 distribution

$$p(d) = \chi^2(r) \tag{16}$$

This results in the same test for multivariate effects in General Linear models described in Chapter 25. In the present context, if a are the autoregressive coefficients from region X to region Y, and the above test finds them to be significantly non-zero, then we can conclude that X Granger causes Y [Eichler 2005].

References

[Assad et al. 1995] J.A. Assad and R.M. Maunsell. Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature*, 373 (6514),518-521,1995.

[Tiao 1992] G.E.P Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis*. John Wiley, 1992.

[Bressler et al. 1999] S.L. Bressler, M. Ding, and W. Yang. Investigation of cooperative cortical dynamics by multivariate autoregressive modeling of event-related local field potentials. *Neurocomputing*, 26-27:625–631, 1999.

[Buchel et al. 1997a] C. Buchel and K.J. Friston. Characterizing functional integration. In R.S.J. Frackowiak, K.J Friston, C.D. Frith, R.J. Dolan, and J.C. Mazziotta, editors, *Human Brain Function*, pages 127–140. Academic Press USA, 1997.

[Buchel et al. 1997b] C. Buchel and K.J. Friston. Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex*, 7:768–778, 1997.

[Buchel et al. 1998] C. Buchel and K.J. Friston. Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. *Human Brain Mapping*, 6:403–408, 1998.

[Buchel and Friston 2000] C. Buchel and K.J. Friston. Assessing interactions among neuronal systems using functional neuroimaging. *Neural Networks*, 13:871–882, 2000.

[Cassidy and Brown 2000] M.J.Cassidy and P.Brown. Stationary and non-stationary autoregressive models for electrophysiological signal analysis and functional coupling studies. *IEEE Trans. Biomed. Eng.*, 49, 1142-1152, 2002.

[Chatfield 1996] C. Chatfield The Analysis of Time Series Chapman and Hall, 1996

[Craven and Savoy 1995] K.M. O'Craven and R.L. Savoy. Voluntary attention can modulate fmri activity in human MT/MST. *Investigational Ophthalmological Vision Science*, 1995.

[Cudeck 2002] R. Cudeck. *Structural Equation Modeling: Present and Future.* Scientific Software International, Lincolnwood, IL, USA, 2002.

[Eichler 2005] M. Eichler Evaluating effective connectivity. *Phil. Trans. R. Soc. B.* 360, 953-967, 2005.

[Frackowiak et al. 1997] R.S.J. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, and J.C. Mazziotta, editors. *Human Brain Function.* Academic Press USA, 1997.

[Friston 2000] K.J. Friston. The labile brain.I. neuronal transients and nonlinear coupling. *Phil Trans R Soc London B*, 355:215–236, 2000.

[Friston 2001] K.J. Friston. Brain function, nonlinear coupling, and neuronal transients. *The Neuroscientist*, 7, 406-418, 2001.

[Friston et al. 1997] K.J. Friston, C. Beuchel, G.R. Fink, J. Morris, E. Rolls, and R.J. Dolan. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6:218–229, 1997.

[Friston and Buchel 2000] K.J. Friston and C. Buchel. Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13):7591–7596, 2000.

[Friston et al. 1993] K.J. Friston, C.D. Frith, and R.S.J. Frackowiak. Time-dependent changes in effective connectivity measured with PET. *Human Brain Mapping*, 1:69–79, 1993.

[Friston et al. 1995] K.J. Friston, L.G. Ungerleider, P. Jezzard, and R. Turner. Charcterizing modulatory interactions between V1 and V2 in human cortex: A new treatment of functional MRI data. *Human Brain Mapping*, 2:211–224, 1995.

[Granger 1969] C. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrika* 37, 424-438.

[Juang 2001] J.N Juang. *Identification and control of mechanical systems.* Cambridge University Press, 2001.

[Kaminski et al. 1997] M. Kaminski, K. Blinowska, and W. Szelenberger. Topographic Analysis of coherence and propagation of EEG activity during sleep and wakefulness. *Electroencephalography and Clinical Neurophysiology*, 102:216–227, 1997.

[Magnus and Neudecker 1997] J.R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, 1997.

[Marple 1987] S.L. Marple. *Digital spectral analysis with applications*. Prentice-Hall, 1987.

[McIntosh et al. 1994] A.R McIntosh, C.L Grady, L.G Underleider, J.V Haxby, S.I Rapoprt, and B. Horwitz. Network analysis of cortical visual pathways mapped with pet. *Journal of Neuroscience*, 14, 655-666, 1994.

[Muirhead 1982] R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley, 1982.

[Muller et al. 2001] K. Muller, G. Lohmann, V. Blosch and D. Yves von Cramon. On multivariate spectral analysis of fMRI time series. *Neuroimage*, 14, 347-356, 2001.

[Neumaier and Schneider 2000] A. Neumaier and T. Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.*, 27, 27-57, 2000.

[Pearl 1998] J. Pearl. Graphs, Causality, and Structural Equation Models *Sociological Methods and Research*, 27, 226-284, 1998

[Penny and Roberts 2002] W.D. Penny and S.J. Roberts. Bayesian multivariate autoregressive models with structured priors. *IEE Proc.-Vis Image Signal Process*, 149(1):33–41, 2002.

[Priestley 1988] M.B Priestley. *Nonlinear and non-stationary time series analysis*. Harcourt Brace Jovanovich, 1988.

[Valdes-Sosa et al. 2005] P.A. Valdes-Sosa and J. Sanchez-Bornot and A. Lage-Castellanos and M. Vega-Hernandez and J. Bosch-Bayard and L. Melie-Garcia and E. Canales-Rodriguez. Estimating brain functional connectivity with sparse multivariate autoregression. *Phil. Trans. R. Soc. B.* 360, 969-981, 2005.

[Weisberg 1980] S. Weisberg. *Applied Linear Regression*. John Wiley, 1980.

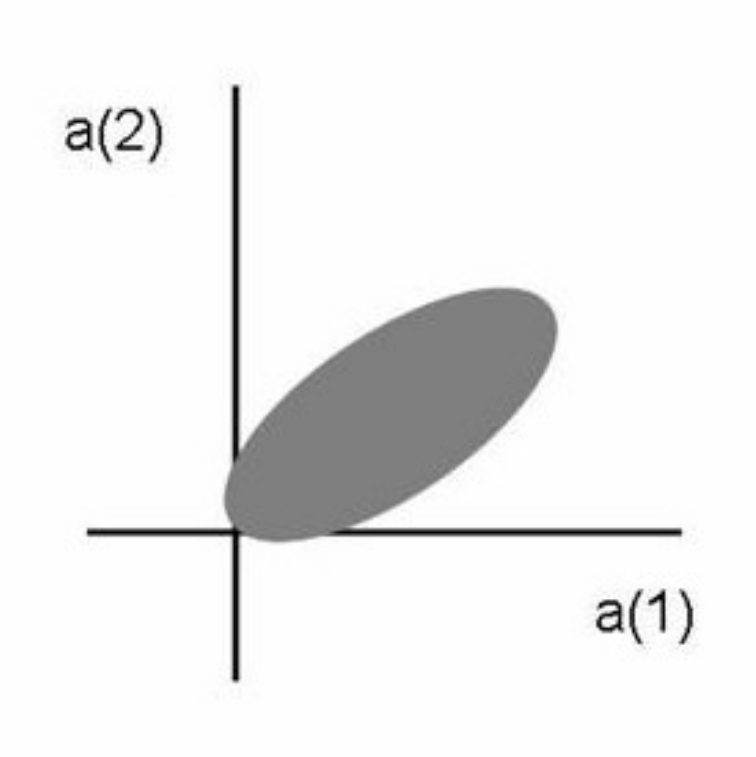

Figure 1: For a MAR(2) model the vector of connection strengths, a, between two regions consists of two values, $a(1)$ and $a(2)$. The probability distribution over a can be computed from the posterior distribution of MAR coefficients as shown in the appendix and is given by $p(a) = \mathsf{N}(\mu, V)$. Connectivity between two regions is then deemed significant at level α if the zero-vector lies on the $1 - \alpha$ confidence region. The figure shows an example $1 - \alpha$ confidence region for a MAR(2) model.

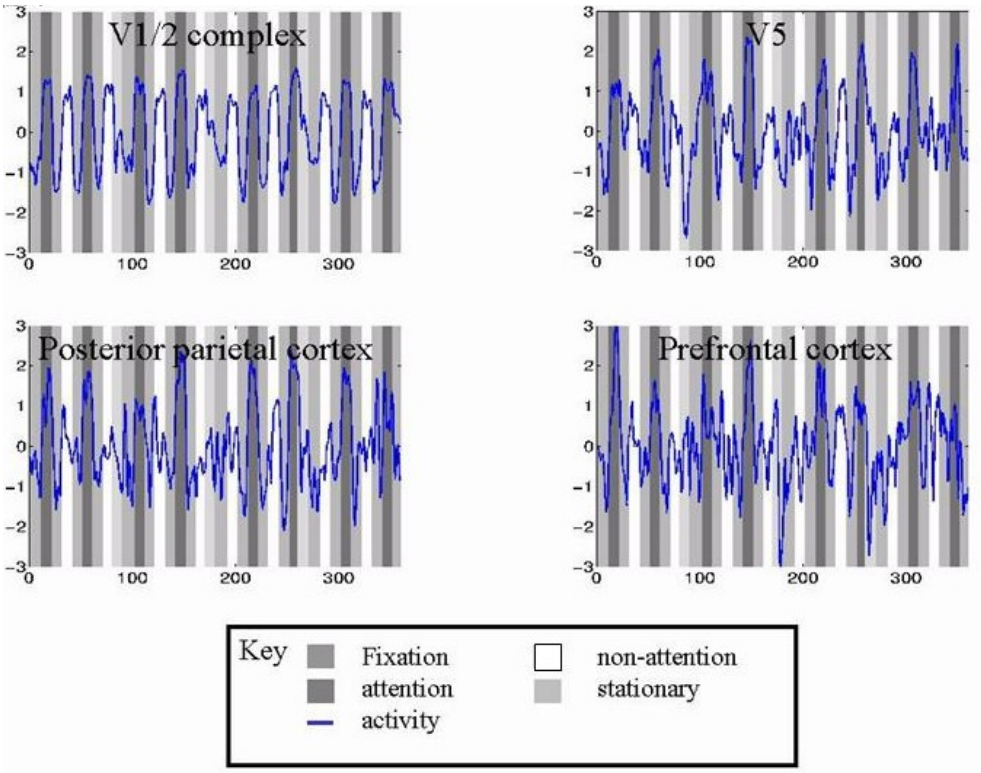

Figure 2: These are the time series of regions V1/2 complex, V5, PPC and PFC from subject 1, in the right hemisphere. All plots have the same axes of activity (adjusted to zero mean and unit variance) vs scan number (360 in total). The experiment consisted of four conditions in four blocks of 90 scans. Periods of "attention" and "non-attention" were separated by a "fixation" interval where the screen was dark and the subject fixated on a central cross. Each block ended with a "stationary" condition were the screen contained a freeze frame of the previously moving dots. Epochs of each task are indicated by the background grayscale (see key) of each series. Visually evoked activity is dominant in the lower regions of the V1/2 complex whereas attentional set becomes the prevalent influence in higher regions PPC and PFC.

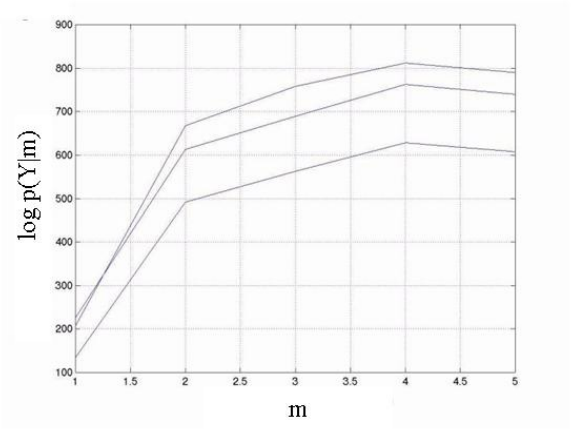

Figure 3: Plots of log-evidence, computed using the expression in the appendix, for each of the three MAR models for subject 1. The optimal order is $m = 4$ in each case.

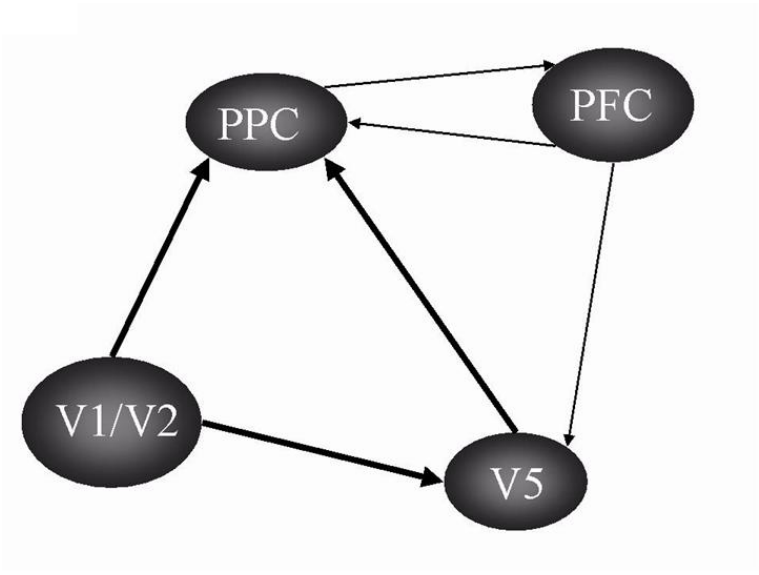

Figure 4: Inferred connectivity for model 1. Arrows indicate Granger causal relations. Thin arrows indicate $0.001 \leq \alpha \leq 0.05$ and thick $\alpha \leq 0.001$.

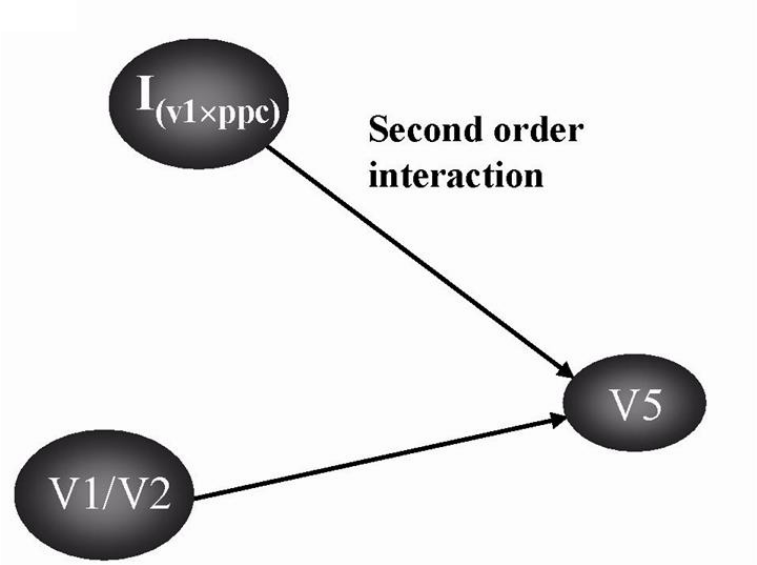

Figure 5: Inferred connectivity for model 2. Arrows show Granger causal relations ($0.001 \leq \alpha \leq 0.05$).The model supports the notion that PPC modulates the V1/V2 to V5 connection.

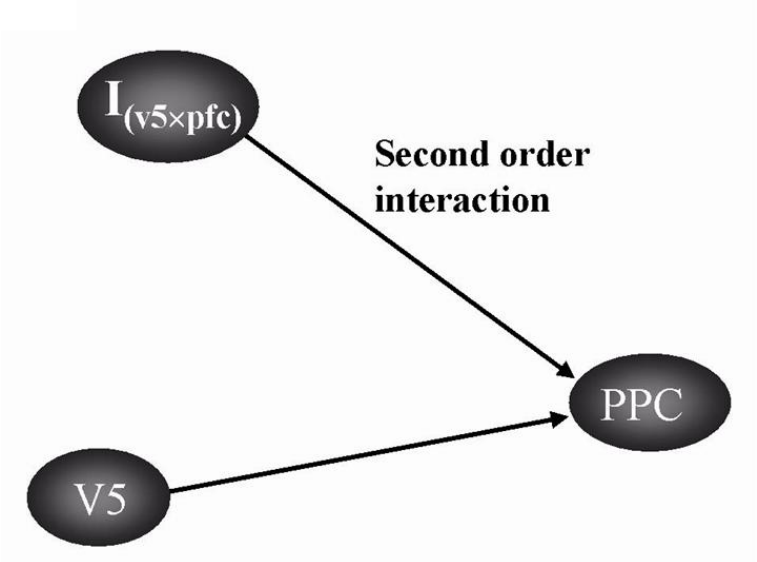

Figure 6: Inferred connectivity for model 3. Arrows show Granger causal relations ($0.001 \leq \alpha \leq 0.05$). The model supports the notion that PFC modulates the V5 to PPC connection.