ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 0 (2003) 000–000

www.elsevier.com/locate/ynimg

Multivariate autoregressive modeling of fMRI time series

L. Harrison,* W.D. Penny, and K. Friston

Wellcome Department of Imaging Neuroscience, University College London, 12 Queen Square, London WC1N 3BG, UK

Received 9 July 2002; revised 7 January 2003; accepted 3 March 2003

Abstract

We propose the use of multivariate autoregressive (MAR) models of functional magnetic resonance imaging time series to make inferences about functional integration within the human brain. The method is demonstrated with synthetic and real data showing how such models are able to characterize interregional dependence. We extend linear MAR models to accommodate nonlinear interactions to model top-down modulatory processes with bilinear terms. MAR models are time series models and thereby model temporal order within measured brain activity. A further benefit of the MAR approach is that connectivity maps may contain loops, yet exact inference can proceed within a linear framework. Model order selection and parameter estimation are implemented by using Bayesian methods.

Introduction

Functional neuroimaging has been used to corroborate functional specialization 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 characterize the network properties of the brain under specific cognitive states (Buchel and Friston, 1997, 2000). These studies address a complementary principle of organization, functional integration.

Functional magnetic resonance imaging (fMRI) 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 specialized regions? To answer this question models are needed to describe the underlying connectivities implied by the data.

Effective connectivity is defined as the influence a neuron (or neuronal population) has on another. At the neuronal level this is equivalent to the effect presynaptic activity has on postsynaptic response, otherwise known as synaptic efficacy. Models of effective connectivity are designed to identify a suitable metric of influence among interconnected components (or regions of interest) in the brain. The notion of inferring influence from recorded data is, however, much more general. Consider the trajectory of an object as the result of external forces acting on it. These forces, which may be represented by equations of motion, determine the object's path. The equations of motion are an example of a model of the physical system. The observed data can be understood and analyzed by using this model. There are many different approaches to modeling a dynamic physical system; however, the motivation is the same for all: to identify operational principles responsible for generating the data.

There are two main approaches to modeling dynamic systems (e.g., physical bodies acted on by external forces or neuronal firing within a network), which can be used to understand spatial and/or temporal order within measured data, such as functional imaging data. These include equations of motion (as above); alternatively, we may model the systems behaviour by simply quantifying relationships within the measured data only. The first approach includes state-space models, e.g., used by the Kalman filter, while the second includes simple regression analysis and convolution models (such as the Volterra approach) to identify statistical dependencies, or patterns, within the data (Juang, 2001).

Both approaches have been used to measure effective

* Corresponding author. Fax: +02078131420.
E-mail address: lharris@fil.ion.ucl.ac.uk

1053-8119/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1053-8119(03)00160-5

connectivity among cortical regions from neuroimaging data starting with regression models (Friston et al., 1993, 1995, 1997; McIntosh et al., 1994), then input-output models (Friston, 2001; Friston and Buchel, 2000), and later state-space models (Buchel and Friston, 1998). Regression techniques, such as psychophysiological interactions, are advantageous as they are easy to solve, yet may be used to approximate nonlinear modulatory interactions (Friston et al., 1997). However, this is at the expense of ignoring temporal information, i.e., the history of an input (experimental task) or physiological variable (imaging data). This is important as interactions within the brain, whether over short or long distances, take time and are not instantaneous (which is implicit within regression models). Furthermore, the instantaneous state of any brain system that conforms to a dynamical system will depend on the history of its input. Structural equation modeling (SEM), as used by the neuroimaging community (Buchel and Friston, 1997; McIntosh et al., 1994), has similar problems.[1] Input-output 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 the data. State-space models account for correlations within the data by invoking state variables whose dynamics generate the data. Recursive algorithms, such as the Kalman filter, may be used to estimate these states through time given the data. This approach was used to estimate variable regression coefficients between V1 and V5 activity in the study by Buchel and Friston (1998).

The MAR model fits into this classification scheme as one that models temporal effects across different variables (e.g., regions of interest), without using state variables. They characterize interregional dependencies within the data, specifically in terms of the historical influence one variable has on another. This is distinct from regression techniques that quantify instantaneous correlations, yet is similar to the Volterra model in that the relative influences, over time, are estimated. These considerations have motivated the investigation of MAR models, which may, in some instances, be suitable for making inferences about functional integration in fMRI.

The study is divided into three 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 described fully by Penny and Roberts, (2002). Second, we test the method with synthetic data before modeling real neurophysiological data taken from an fMRI experiment addressing attentional modulation of cortical connectivity during a visual motion task (Buchel and Friston, 1997). The modulatory effect of one region upon the responses to other regions is a second-order interaction that is precluded in linear models. To circumvent this we have introduced bilinear terms (Friston et al., 1997). We assess the ability of bilinear MAR models to capture top-down modulatory effects of the prefrontal cortex (PFC) and posterior parietal cortex (PPC) on motion sensitive regions in the dorsal visual pathway during attention. In the final section we discuss the advantages of MAR models, its use in spectral estimation, and future developments of the Bayesian approach used to estimate MAR parameters.

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 modeling 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 characterization. The order of the model is the number of preceding observations used and the weights are the parameters of the model estimated from the data that uniquely characterize the time series.

Multivariate autoregressive models extend this approach to multiple time series so that the vector of current values of all variables is modeled as a linear sum of previous activities. Consider d time series generated from d variables (brain regions) within a system such as a functional network in the brain and where p is the order of the model. Here the scalar p denotes order; however, later we will use $p(\alpha)$ to mean the probability of α. A MAR (P) model predicts the next value in a d-dimensional time series, y_n, as a linear combination of the P previous vector values

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

where $y_n = [y_n(1), y_n(2),\ldots, y_n(d)]$ is the nth sample of a d-dimensional time series, each $A(i)$ is a d-by-d matrix of coefficients (weights) and $e_n = [e_n(1), e_n(2),\ldots, e_n(d)]$ 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 $x_n = [y_{n-1}, y_{n-2},\ldots, y_{n-p}]$ are the p previous multivariate time series samples and W is a $(p \times d)$-by-d matrix of MAR coefficients (weights). There are therefore a total of $k = p \times d \times d$ MAR coefficients.

[1] There exist versions of SEM that do model dynamic information; see Cudeck (2002) for details of dynamic factor analysis.

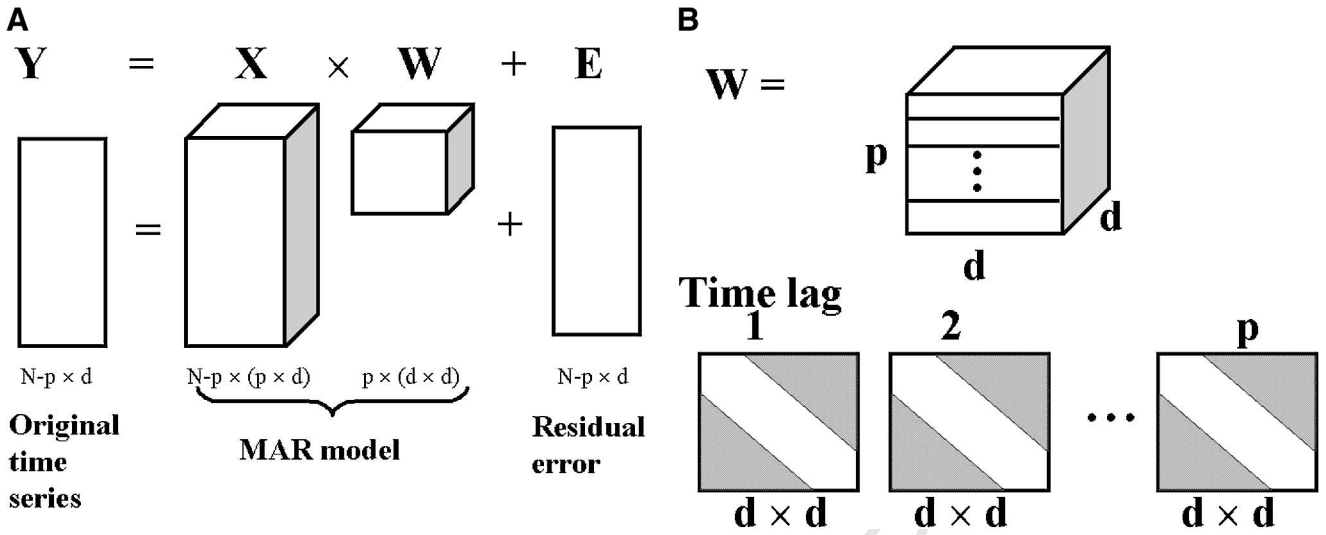


Fig. 1. A schematic of Eq. (3) (main text). (A) The original d-dimensional time series Y is modeled as a MAR process (XW) plus residual error (E). (B) W is a matrix containing all the weights that characterize the interactions among the elements of d. It consists of p layers (one for each time lag used in the model), each layer containing a $d \times d$ matrix of weights. The p layers of W have been placed in sequential order at the bottom of the figure. The diagonal elements are "self-connections" while the off-diagonals reflect the dependence between different variables in the original d-dimensional series. W can be used to compare network properties associated with different cognitive tasks.

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

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

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

MAR models quantify the linear dependence of one region upon all others in the network. The weights in W can be interpreted as characterizing the influence each region has upon it. Independence between a pair of regions results in a weight of zero while dependence is reflected in a nonzero magnitude.

A schematic representation of Eq. (3) is shown in Fig. 1. Fig. 1A shows the original d-dimensional time series (Y) modeled as a MAR process (XW) plus residual error (E). W characterizes the d-dimensional series as a network of connection strengths between all possible pairs of elements in the original series. Fig. 1B is a schematic of W, which consists of P layers, one for each time lag used in the model. Each layer is a $d \times d$ matrix of weights (shown as the squares along the bottom of the figure, again, one for each time lag). The diagonal entries in these are "self-connections" and the off-diagonals are connections between regions. Any dependence among elements in the d-dimensional time series (brain regions) is reflected in nonzero off-diagonal coefficients.

The use of MAR models for characterizing networks of cortical activity associated with cognitive tasks allows one to quantify the dependence among all possible combinations of pairs of regions in the model. This way, connectivity architectures can be compared across different cognitive tasks. For example, the instruction to attend or not to a stimulus will induce plastic changes in connectivity, which may be quantified and compared by using MAR models.

Nonlinear autoregressive models

Given a network coupling 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 upon the response that are independent of each other. However, if the inputs interact to produce a response, the response can no longer be described by a linear combination of the inputs. This is an example of a nonlinear interaction.

In the study by Buchel and Friston (1997), such nonlinear interactions have been modeled by making use of bilinear terms and is the approach adopted here. 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 the bilinear variable. This is then orthogonalized with respect to the original time series. We then form an augmented MAR model with an extra time series and augmented connectivity matrices $\tilde{A}(i)$.

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

The relevant entries in $\tilde{A}(i)$ then reflect modulatory influ-

ences, e.g., 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 the inputs $[u(t)]$. The inclusion of these terms are guided by the hypothesis of interest, e.g., 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, e.g., Box and Tiao (1992) (p. 423).

The maximum likelihood (ML) solution [see, e.g., Weisberg (1980)] for the MAR coefficients is

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

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

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

where $k = p \times d \times d$. 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 "unstack" the columns from the vector $\hat{w}$.

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

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

where $\otimes$ denotes the Kronecker product [see, e.g., p. 477 in Box and Tiao (1992)]

The optimal value of P can be chosen by using a model order selection criterion such as the minimum description length (MDL). See, e.g., 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 study. The maximum-likelihood solution is used to initialise the Bayesian scheme.

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

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

where p 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 the study by Penny and Roberts (2002), it is shown that the corresponding posterior distributions are given by

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

The parameters of the posteriors are updated in an iterative optimization scheme described in the Appendix. Iteration stops when the "Bayesian evidence" for model order p, $p(Y|p)$, is maximized. 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 p.

Bayesian Inference

The Bayesian estimation procedures outlined above result in a posterior distribution for the MAR coefficients $P(W|Y,p)$. Bayesian inference can then take place using confidence intervals based on this posterior. See, for example, page 84 of Box and 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 nonzero" or simply "significant" at level α if the zero vector lies outside the $1 - \alpha$ confidence region for a. This is shown schematically in Fig. 2. We also refer to α as the "P value" (see Appendix B). F2

Application

Synthetic data

To test the face validity of the method two sets of synthetic data were generated, which are shown in Fig. 3. All time series were generated from known MAR(2) models. The known values were compared with estimates of F3

tapraid3/6k-nimage/6k-nimage/6k0603/6k1502-03a | knepper1 | S=5 | 5/20/03 | 9:31 | Art: 1502

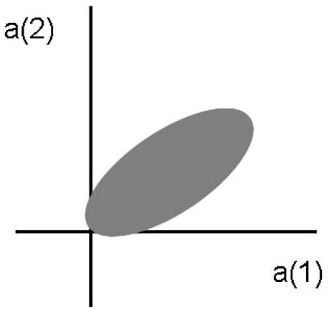


Fig. 2. For a MAR(2) model the vector of connection strengths, *a,* between two regions consists of two values, $a(1)$ and $a(2)$. The probability distribution over a can be computed from the posterior distribution of MAR coefficients as shown in Appendix B and is given by $p(a) = N(m,V)$. Connectivity between two regions is then deemed significant at level alpha 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.

model order and weights. Each data set contained six time series, the first being independent (Fig. 3A) with all off-diagonal MAR terms equal to zero while the second (Fig. 3B) included two sets of three time series that were dependent within sets but independent between sets.

The upper right figure in Fig. 3A shows the true MAR(2) structure from which time series were generated (lower left). The model has only nonzero diagonal terms and no covariance structure and thereby generates independent time series. Each series is essentially an AR(2) process. The simulated time series were modeled as a MAR(p) process, using Bayesian evidence to select the optimal order. This is shown in Fig. 4A, demonstrating, as anticipated, an optimal model order of 2. Parameter estimates are shown for comparison on the right of Fig. 3A. The matrix W is represented in two ways; the first (upper right) shows only the conditional means of the parameter estimates in the same format as the known MAR(2) model. The general character of the known structure has been captured with dominant diagonal terms; however, off-diagonal terms have many nonzero values. A more complete representation of the posterior distribution $[p(W|Y,p)]$ is shown below, which depicts the variance about the estimated means. Each plot within the matrix of graphs contains weight estimates at all time lags in the model for one connection. Zero is indicated and posterior distributions shown in relation it. Those parameters that straddle zero, despite having a nonzero mean, are not considered significant. All connections that are significantly nonzero are circled. The overall structure of the true parameters is reflected in the estimates with the exception of one connection (circled off diagonal coefficient).

The second set of synthetic data contained a mixture of dependence and independence and is shown in Fig. 3B. The format is the same as Fig. 3A. The known MAR(2) model has two subgroups characterized by dependence within each subgroup (to the degree of the coefficients magnitude) and independence between subgroups (reflected in the zero coefficients). Modeling the time series (lower left) as a MAR(p) and using Bayesian evidence (Fig. 4A), the correct model order ($p = 2$) was identified. Parameter estimates (right) again reflect the known MAR(2) structure.

The accuracy of model order selection using Bayesian evidence was generally stable; however, occasionally an incorrect order was calculated by using smaller data sets (e.g., < 250). Increasing the number of data points to 500 produced robust and correct estimates in all cases. The stability of these results is worth noting as estimates of the zero coefficents produced the most "false positives" (usually only 1 of a possible 30 in the first data set and 18 in the second) while "false negatives" occurred less frequently. Given these occasional discrepancies, the Bayesian framework for estimating model parameters was able to differentiate and quantify interdependence within a MAR process.

fMRI data

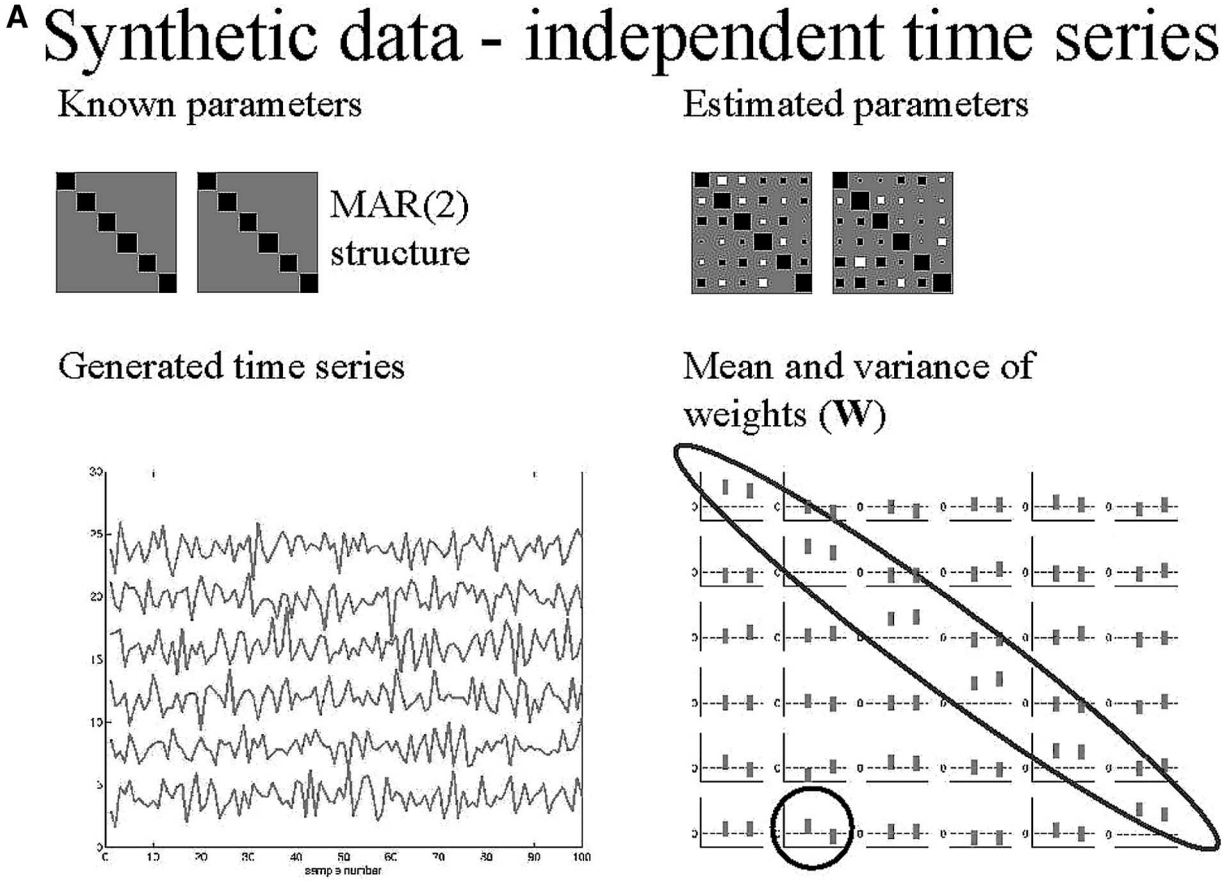
Attentional effects on the responsiveness of motion sensitive area V5 and PPC measured in electrophysiological and neuroimaging studies suggest attention is associated with changes in connectivity (Assad and Mausell, 1995; O'Chaven and Savoy, 1995). In this study we use data from an fMRI study investigating attentional modulation of connectivity within the dorsal visual pathways (Buchel and Friston, 1997). This provides a testbed for assessing how MAR models estimate changes in connectivity.

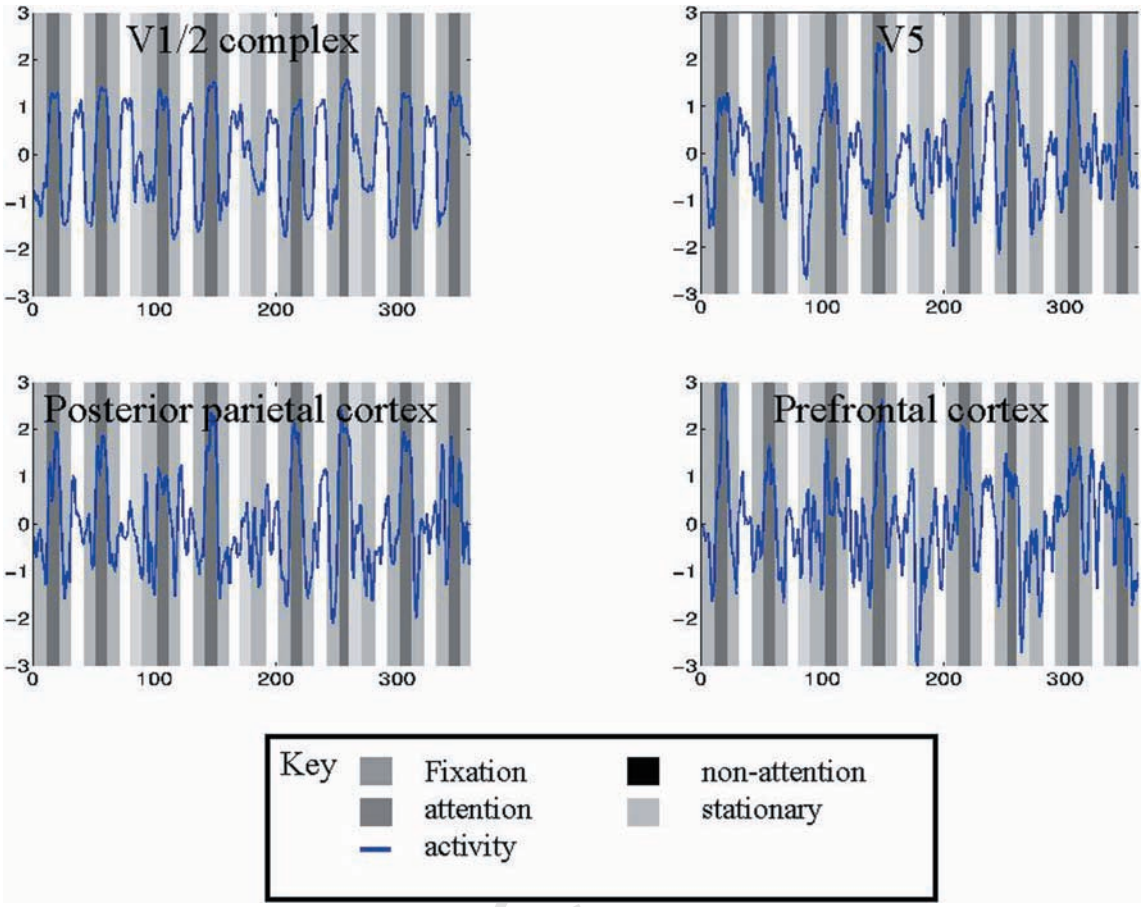
In brief, the experiment was performed on a 2-T MRI scanner on several subjects. The visual stimulus involved random dots moving radially outward at a fixed rate. Subjects were trained beforehand to detect changes in velocity of radial motion. Attentional set was manipulated by asking the subject to 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 being attentional set.

Categorical comparisons using SPM(t) 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 8 mm centered around the most significant (<0.05, corrected) 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 and Friston (1997). The time series of the right hemisphere regions, in one subject, are shown in Fig. 4.

Inspecting the four time series reveals a number of characteristics worth noting. The series from the V1/2 complex show a dependence on the presentation of the moving image

tapraid3/6k-nimage/6k-nimage/6k0603/6k1502-03a | knepper1 | S=5 | 5/20/03 | 9:31 | Art: 1502





COLOR

AQ: 29 Fig. 4. These are the representative time series of regions V1/2 complex, V5, and PPC, and PFC from one subject, in the right hemisphere. All plots have the same axes of activity (adjusted to zero mean and unit variance) versus 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, and each block ended with a "stationary" condition where 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.

with a small difference between attention and nonattention. 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 upon others, or indeed any nonadditive interaction, is not obvious from visual inspection alone. Modeling the series as a MAR process provides a quantitative approach to these putative effects.

Three models were tested by using the regions and bilinear terms shown below. Bilinear terms for interactions between V1/2 complex and PPC are written as $I_{v1,ppc}$, and regions V5 and PFC as $I_{v5,pfc}$. These time series were entered into bilinear MAR models. The interaction terms can be thought as "virtual" nodes in a network. Models 1 and 3 involved only right hemisphere PFC as no significant attention related activation in the left PFC was found.

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

The motivation for the first model was to ask a very general question: given time series of brain activity over the

Fig. 3. Two synthetic data sets are generated from known MAR(2) models (A and B). Both A and B have the same layout with the known MAR(2) model (upper left) from which time series are generated (sample in lower left). See Fig. 4A, for a plot of the Bayesian evidence verses model order identifying a model order of $p = 2$. Parameter estimates are shown on the right in two formats. The first (upper right) is a plot of estimated means of weighting parameters for comparison with the known MAR model. Below this is a more comprehensive representation of the posterior distribution [$p(W|Y,p)$] of the estimates. There are 6×6 plots with mean and variance (two standard deviations) shown at all time lags (2) in relation to zero. Distributions that do not straddle zero are significantly nonzero. A and B differ in that the known MAR(2) models used in A produce largely independent series whereas B generates mixed dependences between series. Estimates that are significantly different (<0.05) from zero across both time lags are circled and reveal the same MAR(2) underlying process that generated the series originally.

Table 1
P values for testing individual connection strengths across all time lags for Model 1 for all three subjects, right hemisphere only[a]

Direction	Regions	Subject 1	Subject 2	Subject 3
Bottom-up	v1-v5	(0.0000[c])	(0.0004[c])	(0.0000[c])
	v1-ppc	(0.0000[c])	[d]	[d]
	v1-pfc	[d]	(0.007[b])	[d]
	v5-ppc	(0.0002[c])	(0.0009[c])	[d]
	ppc-pfc	(0.0182[b])	[d]	[d]
Top-down	pfc-v1	[d]	[d]	(0.0045[b])
	pfc-v5	(0.0005[c])	[d]	[d]
	pfc-ppc	(0.0197[b])	(0.0017[b])	(0.0001[c])

[a] The table is divided in two with the upper part showing values for ascending connections (bottom-up) and the lower part descending connections (top-down) within the visual and attentional systems.

[b,c,d]The most significant coefficients are shown schematically for Subject 1 in Fig. 6. [b] *P* values between 0.05 and 0.001; [c] *P* values < 0.001; [d] for estimates not significantly different from zero ($P > 0.05$).

entire experiment, during all four states, is there, on average, a functional network connecting key regions in the visual and attentional systems? The second and third models were motivated by SEM and Volterra analyses of the same data reported by Buchel and Friston (1997) and Friston and Buchel (2000), respectively. These different methodologies address the same issues of modulatory, or bilinear interactions, at different levels of the visual and attentional pathways and provide a convenient benchmark with which to validate the current approach. Models 2 and 3 therefore specifically address whether or not attentional influence from top-down regions could be mediated by second-order interactions. If there is no bilinear interaction among the regions then this should be reflected as zero-valued weights within a bilinear MAR model.

All pairwise connection strengths (MAR coefficents across all time lags) were tested separately across all time lags for significant differences from zero. This raises an important issue of how apriori knowledge of connectivity should determine the analytic strategy. By testing all connections we are essentially approaching the model with no prior knowledge. Alternatively, a model led approach is directed by prior knowledge of connectivity, thereby only testing connections established by, for example, anatomical studies. Given the motivation of the first model, to test for average connectivities over all tasks, we used the former strategy.

AQ: 3 The results of all connections that attained signifcance (*P* values <0.05) are shown for three subjects in Tables 1, 2, and 3. Connectivity maps have been used to illustrate the dependencies in Figs. 6, 7, and 8 for Models 1, 2, and 3, respectively. The width of the arrows are scaled to the *P* values (thin for *P* values between 0.05 and 0.001 and thick for values <0.001). The optimal model order was selected by using Bayesian evidence. Plots of which are shown for one subject and all three models in Fig. 5B. The optimal model order was $P = 4$ for all models.

The first model characterizes the network of connectivities on average over all attentional states. Table 1 reveals clear feedforward and backward connections within the network. All V1 to V5 connections reached very high significance ($P < 0.0004$) and for V5 to PPC connections this was true for two subjects ($P < 0.0009$). Top-down connections between PFC and PPC were demonstrated in all subjects ($P < 0.02$).

Table 2
P values for connection strengths across all time lags for Model 2 for all three subjects[a]

Direction	Regions	Subject 1	Subject 2	Subject 3
Top-down	$I_{v1,ppc}$-v5 right	(0.0205[b])	[c]	[c]
	$I_{v1,ppc}$-v5 left	(0.0384[b])	(0.0568)	(0.0474[b])

[a] Left and right hemispheres were modeled separately. The table includes top-down connections in both hemispheres. The bilinear term $I_{v1,ppc}$ introduces second-order interactions among PPC, V1, and V5. There is a consistent dependence between $I_{v1,ppc}$ and V5 in the left hemisphere for all three subjects. Diagrams of the weight matrix *W* and connectivity maps are shown in Fig. 7.

A schematic of these results is shown in Fig. 6 for one subject. Fig. 6A is the weight matrix *W* shown as a 4×4 matrix with each element containing a representation of the posterior distribution $[p(W|Y,p)]$ of weights for one connection at all time lags in the model, the same as in Fig. 3. Given the hierarchical sequence of the regions the upper off-diagonal terms correspond to dependence ascending through the cortical hierarchy. For example the plots in position (1,2) are the coefficients characterizing the influence V1 has upon V5 at all time lags. In short, upper diagonal coefficients quantify the influence of forward connections. The lower off-diagonal terms complement these and characterize backward projections. The forward driving influence of V1 on V5, V5 on PPC, and PPC on PFC is evident. Back projecting influences of PFC upon V5 and PPC are also shown. These are important observations; however, they are limited in that they represent a linear characterization that precludes attentional modulation. The second and third models were designed to test whether attentional areas modulate coupling within the visual system.

Bilinear terms $I_{v1,ppc}$ and $I_{v5,pfc}$ were included in Models 2 and 3, respectively. The results of Model 2 are displayed in a similar fashion to Model 1 in Table 2 and Fig. 7. Both left and right hemispheres were modeled separately and show consistent results in the left for the connection be-

Table 3
Similar layout for Model 3, right hemisphere only[a]

Direction	Regions	Subject 1	Subject 2	Subject 3
Top-down	$I_{v5\text{-}pfc}$-ppc	(0.056)	(0.03[b])	[c]

[a] Second-order interactions are modeled between $I_{v5,pfc}$ and PPC and show significant coefficients in two of the three subjects. Weight matrix and connectivity maps are shown in Fig. 8.

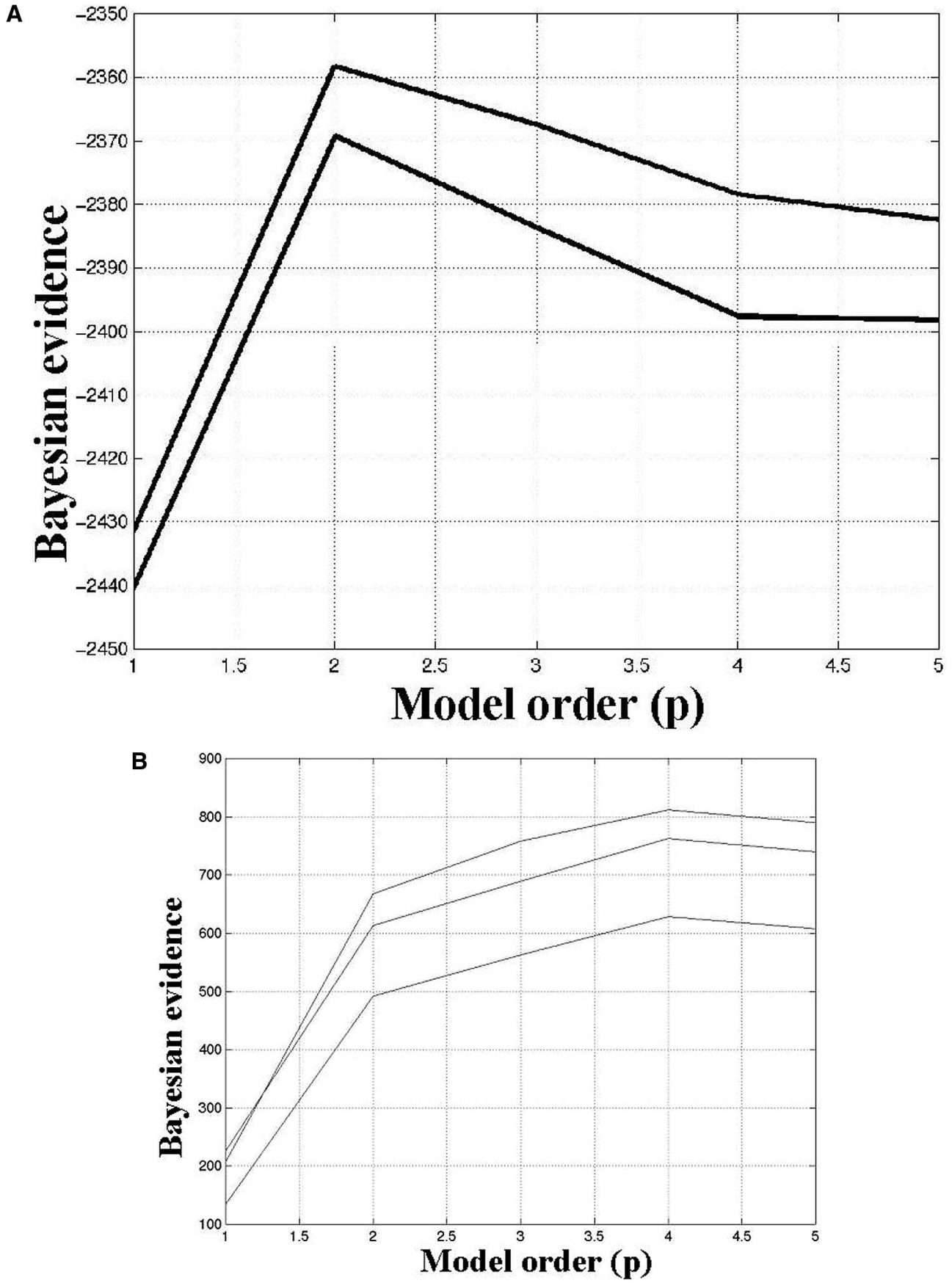
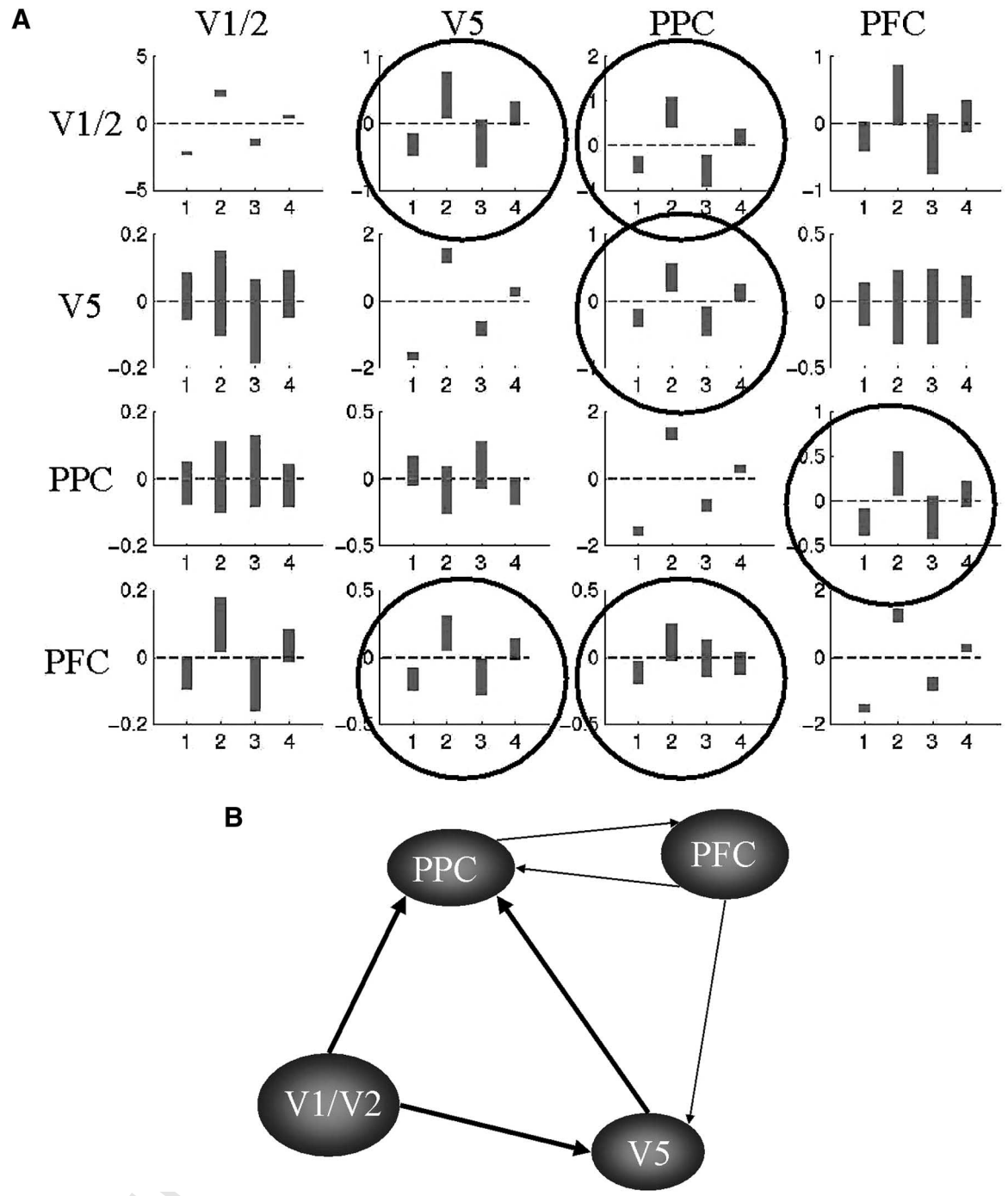


Fig. 5. Model order selection using the Bayesian evidence versus model order p. A contains two plots calculated for the synthetic data described in Fig. 3. In both data sets (independent and mixed dependence) the optimal order was identified as $p = 2$. B shows three plots generated from one subject for all three models of the real data described in Fig. 5. The optimal order was $p = 4$ for each subject.

tapraid3/6k-nimage/6k-nimage/6k0603/6k1502-03a | knepper1 | S=5 | 5/20/03 | 9:31 | Art: 1502



AQ: 31 Fig. 6. Model 1 including regions V1/2 complex, V5, PPC, and PFC. (A) *W* from a MAR(4) model, with each region indicated on the rows and columns of the diagram. Each element within the matrix of plots contains means and variances of parameter estimates at all time lags for that connection (same as Fig. 3). Diagonals are "self-connections" and, given the order of variables, the upper off-diagonal coefficients represent ascending, feed forward influence, and can be interpreted as characterizing driving connections through the cortical hierarchies. The lower off-diagonals complement these and characterize the influence higher regions have upon lower, estimating back-projecting activity. All significant connections are circled. (B) Connectivity map of all connection strengths that were significantly different from zero across all time lags for model 1. Arrow width is scaled to the *P* value of the connection strength estimated in *W* (thin for *P* values between 0.05 and 0.001 and thick for <0.001).

tween $I_{v1,ppc}$ and V5 ($P < 0.05$). This indicates that PPC changes the connection between V1 and V5. Only Subject 1 reached significance on the right for this interaction. The bilinear effect is depicted in Fig. 7B and C, first (left figure) showing the bilinear variable as a "virtual" node and second (right figure) its implicit physiological interpretation. Given the posterior density of the connection strength, bilinear terms are seen to account for a significant component of the activity observed within these regions.

The results of Model 3 are shown in Table 3 and Fig. 8.

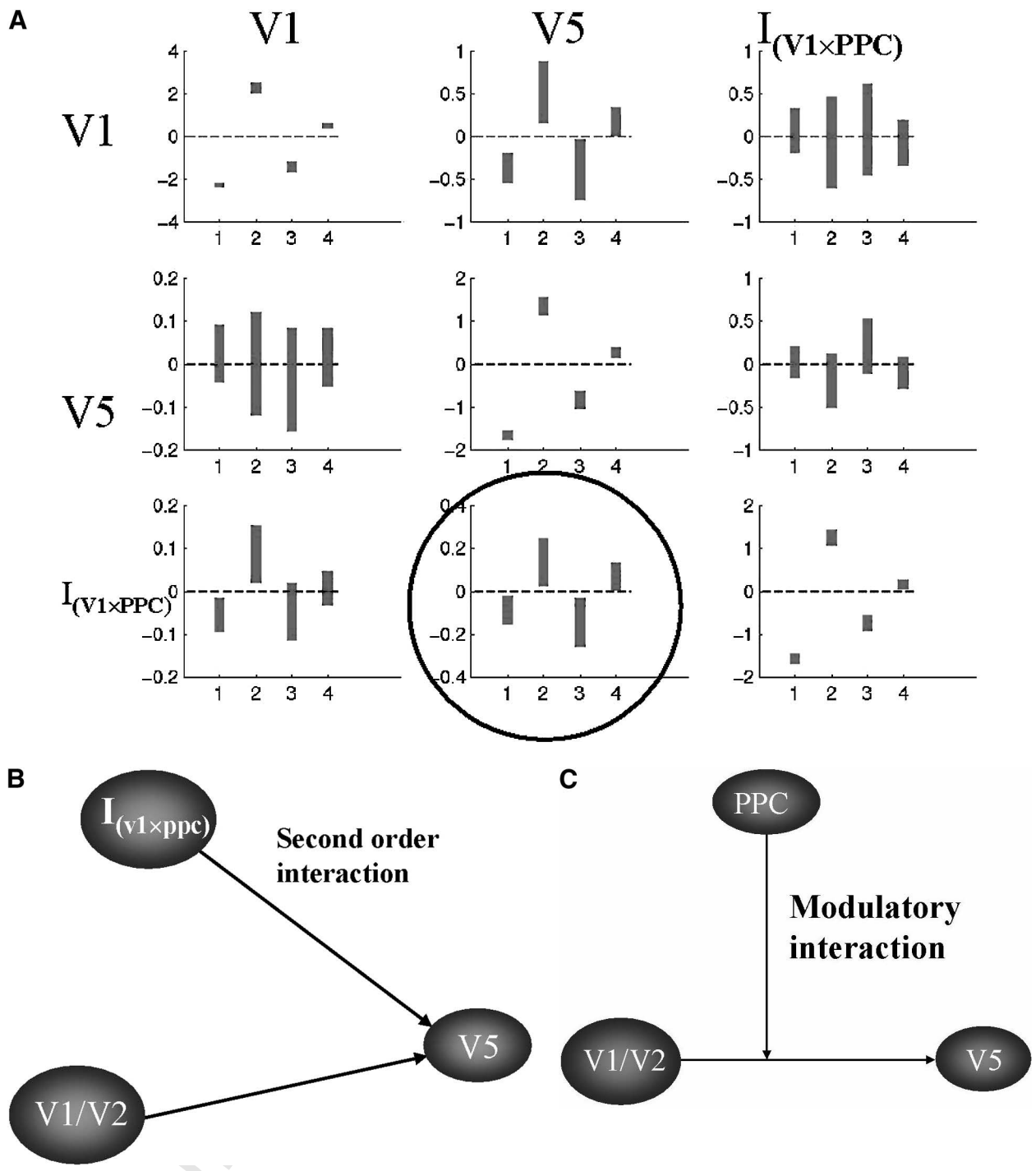
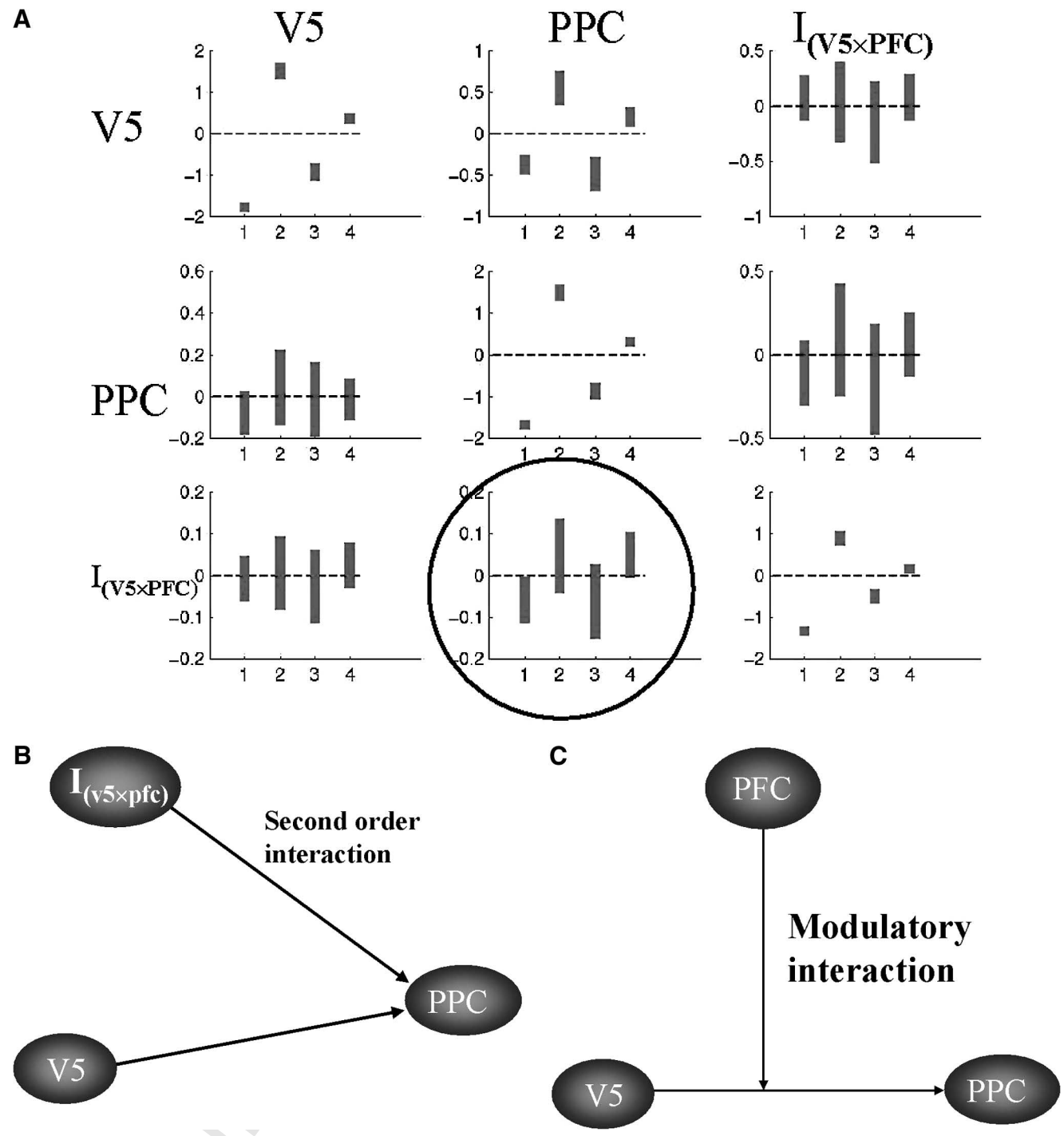


Fig. 7. Model 2 including regions V1/2 complex, V5, and bilinear term $I_{v1,ppc}$ for one subject (right hemisphere). (A) W is shown in the same layout as Fig. 6. The weight of interest is between $I_{v1,ppc}$ and V5 as it describes a second-order interaction characterizing the modulatory role upon downstream processing (upon the connection strength between V1 and V5). These coefficients were significantly nonzero (circled). A diagram of the model including the bilinear term as a "virtual" node is shown in B and the physiological interpretation of this is shown in C, where $I_{v1,ppc}$ is not shown but is represented implicitly (the thin arrow represents a P value of between 0.05 and 0.001).

Second-order connections between $I_{v5,pfc}$ and PPC reach significance with P values between 0.03 and 0.056 in two subjects. Fig. 8B demonstrates a similar interaction as in Fig. 7B, suggesting that attention may be mediated by PFC and modulation of the connection strength between V5 and PPC. That is, PFC changes how PPC responds to V5, the responsiveness being greater during attention than nonattention. This is a bilinear effect, similar to that found in Model 2.



AQ: 32 AQ: 33 Fig. 8. Model 3 including regions, V5, PPC, and bilinear term $I_{v5,pfc}$ for one subject (right hemisphere). The significance of the connection between $I_{v5,pfc}$-PPC is tested as in previous models and connectivity maps are shown as in Fig. 6. The model supports the notion that PFC plays a modulatory role during attention upon the connectivity between V5 and PPC.

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 dominant model, used for making such inferences in the existing fMRI/PET literature, namely structural equation modeling, as used by Buchel and Friston (1997) and McIntosh et al. (1994), is not a time series model. Indeed, inferences are based solely on the instantaneous correlations between regions; i.e., if the time series were randomly permuted SEM would give the same results. Thus SEM throws away temporal information.

This deficiency is not shared by MAR models that are proper time series models.

Further, MAR models may contain loops and self-connections yet parameter estimation can proceed in a purely linear framework; i.e., there is an analytic solution that can be found via linear algebra. In contradistinction, SEM models with loops require nonlinear optimization. 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 (i.e., the activities are similar from one sample to the next) this captures much the same effect, but in a computationally simpler manner.

Given that MAR models extract temporal information, how can this be interpreted? MAR models derive temporal information from the auto and crosscovariance function, which is used to estimate coefficients at different lags. The temporal profile of the coefficient estimates characterizes the temporal aspects of the dependencies. For example, the coefficients can alternate from positive to negative and decay with increasing lags, due to oscillatory interactions. As BOLD is a measurement of the hemodynamic response to neuronal processes, temporal information is smoothed, rendering the coefficients a summary of neuronal activity observed during the hemodynamic response. This may confound the interpretation of the exact timing of an interaction; however, general observations regarding the brains response are possible. In Fig. 6A the first and second coefficient of connectivity between PPC and PFC, both forward and backward, are similar. This suggests that on average, throughout the entire experiment, that there is an equivalent level of connectivity between the two regions, over a period of two time lags (approximately 6 s). The models in Figs. 7 and 8 support the hypothesis of a modulatory interaction, modeled using a bilinear term, from PPC and PFC, whose effect becomes less prominent with time, i.e., up to four time lags (approximately 12 s).

In this study we have used "off-the-shelf" MAR models in which every region is connected to every other region. Bayesian inferences about connections are then made on the basis of the estimated posterior distribution. This is in the spirit of how general linear models are used for characterizing functional specialization; 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 modeled and these are chosen on the basis of prior anatomical or functional knowledge. In cases where this knowledge is available this may be the preferred approach and, in the future, we envisage the use of MAR models that are not fully connected (see Bayesian estimation below).

MAR models can 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, 2002), 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) that looks for sets of voxels that are highly coherent. MAR models provide a parametric way of estimating this coherence, although in this study we have reported the results in the time domain.

A further aspect of our off-the-shelf MAR models is that they capture only linear relationships between regions. Following Buchel and Friston (1997), we have extended their capabilities by introducing bilinear terms. It is also possible to include further higher order terms, for instance, second-order interactions across different lags. Frequency domain characterization of the resulting models would then allow us to report bispectra [28]. These describe the correlations between different frequencies that may be important for the study of functional integration (Friston, 2000).

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. In the future, the Bayesian approach could be greatly extended. In the study by Penny and Roberts (2002), we show how MAR coefficients can be placed into groups. For example, all of those connecting the same two regions could be placed in the same group. Different groups could then be associated with different prior precisions. Groups with prior means of zero and infinite prior precision in effect then specify the absence of a connection. In this way, we could design MAR models with sparse connectivities. More generally, prior means and precisions could be estimated from data using a Bayesian framework that could be specified so as to include time series from multiple subjects. These further extensions will be the subject of subsequent studies.

Acknowledgments

This work was funded by the Wellcome Trust.

Appendix A

Bayesian estimation

Following the algorithm developed in the study by Penny and Roberts (2002), the parameters of the posterior distributions are updated iteratively as follows:

$$\Lambda_D = \hat{\Lambda} \otimes (X^T X) \tag{11}$$

$$\hat{\sum}_B = (\Lambda_D + \hat{\alpha} I)^{-1}$$

$$\hat{W}_B = \hat{\sum}_B \Lambda_D \hat{W}$$

$$\frac{1}{\hat{b}} = \frac{1}{2}\hat{W}_B^T \hat{W}_B + \frac{1}{2}\mathrm{Tr}(\hat{\sum}_B) + \frac{1}{b}$$

$$\hat{c} = \frac{k}{2} + c$$

$$\hat{\alpha} = \hat{b}\hat{c}$$

$$s = N$$

$$B = \frac{1}{2}(Y - X\hat{W}_B)^T(Y - X\hat{W}_B) + \sum_n (I \otimes x_n)\hat{\sum}_B (I \otimes x_n)^T$$

$$\hat{\Lambda} = sB^{-1}$$

The updates are initialized using the maximum-likelihood solution. Iteration terminates when the Bayesian evidence increase by less than 0.01%. The Bayesian evidence is computed as follows

$$p(Y|p) = \frac{N}{2}\log|B| - KL(p(W|p),p(W|Y,p)) - KL(p(\alpha|p),p(\alpha|Y,p)) + \log\Gamma_d(N/2) \tag{12}$$

where $KL(p1,p2)$ denotes the Kullback-Liebler (KL) divergence between densities $p1$ and $p2$. Expressions for these are given in in the study by Penny and Roberts (2002). Essentially, the first term in the above equation is an accuray term and the KL terms act as a penalty for model complexity.

Appendix B

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 nonzero 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 $(k \times k)$ $(k = p \times d \times d)$ sparse matrix C such that

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

returns the estimated weights for connections between the two regions of interest. For a MAR(p) model, this vector has p entries, one for each time lag. The probability distribution is given by $p(a) = N(m, V)$ and is shown schematically in Fig. 2. The mean and covariance are given by

$$m = C^T \hat{w} \tag{14}$$

$$V = C^T \hat{\sum}_B C$$

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 $N(0, V)$. This latter probabilitiy can be computed by forming the test statistic

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

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

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

References

Assad, J.A., Maunsell, R.M., 1995. Neuronal correlates of inferred motion in primate posterior parietal cortex. Nature XXX, XXX–XXX.

Box, G.E.P., Tiao, G.C., 1992. Bayesian Inference in Statistical Analysis. John Wiley, New York.

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

Buchel, C., Friston, K.J., 1997. Characterizing functional integration, in: Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Mazziotta, J.C. (Eds.), Human Brain Function. Academic Press, New York, pp. 127–140.

Buchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.

Buchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb. Cortex 7, 768–778.

Buchel, C., Friston, K.J., 1998. Dynamic changes in effective connectivity characterized by variable parameter regression and kalman filtering. Hum. Brain Mapp. 6, 403–408.

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

Cudeck, R., 2002. Structural Equation Modeling: Present and Future.

Muller, K., et al., 2001. On multivariate spectral analysis of fmri time series. Neuroimage XX, XXX–XXX.

Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Mazziotta, J.C. (Eds.), 1997. Human Brain Function. Academic Press, New York.

Friston, K.J., 2000. The labile brain. I. Neuronal transients and nonlinear coupling. Phil. Trans. R. Soc. Lond. B 355, 215–236.

Friston, K.J., 2001. Brain function, nonlinear coupling, and neuronal transients. The Neuroscientist, XX, XXX–XXX.

Friston, K.J., Beuchel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6, 218–229.

Friston, K.J., Buchel, C., 2000. Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc. Natl. Acad. Sci. USA 97, 7591–7596.

Friston, K.J., Frith, C.D., Frackowiak, R.S.J., 1993. Time-dependent changes in effective connectivity measured with PET. Hum. Brain Mapp. 1, 69–79.

Friston, K.J., Ungerleider, L.G., Jezzard, P., Turner, R., 1995. Characterizing modulatory interactions between v1 and v2 in human cortex: a new treatment of functional MRI data. Hum. Brain Mapp. 2, 211–224.

Juang, J.N., 2001. Identification and Control of Mechanical Systems. AQ: 15 Cambridge University Press.

Kaminski, M., Blinowska, K., Szelenberger, W., 1997. Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness. Electroencephalogr. Clin. Neurophysiol. 102, 216–227.

Magnus, J.R., Neudecker, H., 1997. Matrix Differential Calculus with AQ: 16 Applications in Statistics and Econometrics. John Wiley, New York.

AQ: 17 Marple, S.L., 1987. Digital Spectral Analysis with Applications. Prentice-Hall.

McIntosh, A.R., Grady, C.L., Underleider, L.G., Haxby, J.V., Rapoprt, S.I., Horwitz, B., 1994. Network analysis of cortical visual pathways AQ: 18 mapped with pet. J. Neurosci. XXX, XXX–XXX.

Cassidy, M.J., Brown, P., 2002. Stationary and non-stationary autoregressive models for electrophysiological signal analysis and functional coupling studies. Preprint. AQ: 19

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. John Wiley, New York. AQ: 20

Neumaier, A., Schneider, T., 2000. Estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Trans. Math. Softw. XXX, XXX–XXX. AQ: 21

O'Craven, K.M., Savoy, R.L., 1995. Voluntary attention can modulate fmri activity in human mt/mst. Invest. Ophthalmol. Vis. Sci. XXX, XXX–XXX. AQ: 22

Penny, W.D., Roberts, S.J., 2002. Bayesian multivariate autoregressive models with structured priors. IEE Proc. Vis. Image Signal Proc. 149, 33–41.

Priestley, M.B., 1988. Nonlinear and Non-Stationary Time Series Analysis. Harcourt Brace Jovanovich. AQ: 23

Weisberg, S., 1980. Applied Linear Regression. John Wiley, New York. AQ: 24