Published in final edited form as:

Neuroimage. 2011 January 15; 54(2): 807–823. doi:10.1016/j.neuroimage.2010.09.052.

Multivariate dynamical systems models for estimating causal interactions in fMRI

Srikanth Ryali¹, Kaustubh Supekar^{2,3}, Tianwen Chen¹, and Vinod Menon^{1,4,5}

¹ Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94305

² Graduate Program in Biomedical Informatics, Stanford University School of Medicine, Stanford, CA 94305

³ Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, CA 94305

⁴ Program in Neuroscience, Stanford University School of Medicine, Stanford, CA 94305

⁵ Department of Neurology & Neurological Sciences, Stanford University School of Medicine, Stanford, CA 94305

Abstract

Analysis of dynamical interactions between distributed brain areas is of fundamental importance for understanding cognitive information processing. However, estimating dynamic causal interactions between brain regions using functional magnetic resonance imaging (fMRI) poses several unique challenges. For one, fMRI measures Blood Oxygenation Level Dependent (BOLD) signals, rather than the underlying latent neuronal activity. Second, regional variations in the hemodynamic response function (HRF) can significantly influence estimation of causal interactions between them. Third, causal interactions between brain regions can change with experimental context over time. To overcome these problems, we developed a novel state-space Multivariate Dynamical Systems (MDS) model to estimate intrinsic and experimentally-induced modulatory causal interactions between multiple brain regions. A probabilistic graphical framework is then used to estimate the parameters of MDS as applied to fMRI data. We show that MDS accurately takes into account regional variations in the HRF and estimates dynamic causal interactions at the level of latent signals. We develop and compare two estimation procedures using maximum likelihood estimation (MLE) and variational Bayesian (VB) approaches for inferring model parameters. Using extensive computer simulations, we demonstrate that, compared to Granger causal analysis (GCA), MDS exhibits superior performance for a wide range of signal to noise ratios (SNRs), sample length and network size. Our simulations also suggest that GCA fails to uncover causal interactions when there is a conflict between the direction of intrinsic and modulatory influences. Furthermore, we show that MDS estimation using VB methods is more robust and performs significantly better at low SNRs and shorter time series than MDS with MLE. Our study suggests that VB estimation of MDS provides a robust method for estimating and interpreting causal network interactions in fMRI data.

Address for Correspondence: Srikanth Ryali, Ph.D. and Vinod Menon, Ph.D., Department of Psychiatry & Behavioral Sciences, 780 Welch Rd, Room 201, Stanford University School of Medicine, Stanford, CA 94305-5778, sryali@stanford.edu and menon@stanford.edu, Tel no: 1-650-736-0128, Fax no: 1-650-736-7200.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Causality; Dynamical Systems; Variational Bayes; Bilinear; Expectation Maximization; Kalman Smoother; Deconvolution

Introduction

Functional magnetic resonance imaging (fMRI) has emerged as a powerful tool for investigating human brain function and dysfunction. fMRI studies of brain function have primarily focused on identifying brain regions that are activated during performance of perceptual or cognitive tasks. There is growing consensus, however, that localization of activations provides a limited view of how the brain processes information and that it is important to understand functional interactions between brain regions that form part of a neurocognitive network involved in information processing (Bressler and Menon 2010; Friston, 2009c; Fuster, 2006). Furthermore, evidence is now accumulating that the key to understanding the functions of any specific brain region lies in disentangling how its connectivity differs from the pattern of connections of other functionally related brain areas (Passingham et al., 2002). A critical aspect of this effort is to better understand how causal interactions between specific brain areas and networks change dynamically with cognitive demands (Abler et al., 2006; Deshpande et al., 2008; Friston, 2009b; Goebel et al., 2003; Mechelli et al., 2003; Roebroeck et al., 2005; Sridharan et al., 2008). These and other related studies in the literature highlight the importance of dynamic causal interactions for understanding brain function at the systems level.

In recent years, several methods have been developed to estimate causal interactions in fMRI data (Deshpande et al., 2008; Friston et al., 2003; Goebel et al., 2003; Guo et al., 2008; Rajapakse and Zhou, 2007; Ramsey et al., 2009; Roebroeck et al., 2005; Seth, 2005; Smith et al., 2009; Valdes-Sosa et al., 2005). Of these, Granger causal analysis (GCA) (Roebroeck et al., 2005; Seth, 2005) and dynamic causal modeling (DCM) (Friston et al., 2003) are among the more commonly used approaches thus far. There is a growing debate about the relative merits and demerits of these approaches for estimating causal interactions using fMRI data (Friston, 2009a, b; Roebroeck et al., 2009). The main limitations of GCA highlighted by this debate are that: (1) GCA estimates causal interactions in the observed Blood-Oxygenation-Level-Dependent (BOLD) signals, rather than in the underlying neuronal responses; (2) GCA may not be able to accurately recover causal interactions because of regional variations in hemodynamic response; and (3) GCA does not take into account the experimentally induced modulatory effects while estimating causal interactions (Friston, 2009a, b). The main limitations of DCM highlighted in this debate are that: (1) DCM is a confirmatory method wherein several causal models are tested and the model with the highest evidence is chosen. This is problematic if the number of regions under investigation is large since a large number of models need to be tested increases exponentially with the increase in the number of regions; (2) conventional DCM uses a deterministic model to describe the dynamics of the latent neuronal signals which may not be adequate to capture the dynamics of the underlying neuronal processes (a stochastic version was recently proposed (Daunizeau et al., 2009)); and (3) the assumptions used by DCM for deconvolution of hemodynamic response have not yet been adequately verified (Roebroeck et al., 2009). Here, we develop a new method that incorporates the relative merits of both GCA and DCM while attempting to overcome their limitations.

We propose a novel multivariate dynamical systems (MDS) approach (Bishop, 2006) for modeling causal interactions in fMRI data. MDS is based on a state-space approach which can be used to overcome many of the aforementioned problems associated with estimating

causal interactions in fMRI data. State-space models have been successfully used in engineering applications of control systems and machine learning (Bishop, 2006) but their use in neuroscience has been limited. Notable examples of state-space models include Hidden Markov models (HMM) which are widely used in speech recognition applications (Rabiner, 1989) and Kalman filters for object tracking (Koller and Friedman, 2009). Critically, state-space models can be represented as probabilistic graphical models (Koller and Friedman, 2009), which as we show below (Figure 1), greatly facilitate representation and inference for causal modeling of fMRI data.

Critically, MDS estimates causal interactions in the underlying latent signals, rather than the observed BOLD-fMRI signals. In order to estimate causal interactions from the observed fMRI data, it is important to take into account variations in hemodynamic response function (HRF) across different brain regions (David et al., 2008). MDS is a state-space model in which a “state equation” is used to model the unobserved states of the system and an “observation equation” is used to model the observed data as a function of latent state signals (Figure 1). The state equation is a vector autoregressive model incorporating both intrinsic and modulatory causal interactions. Intrinsic interactions reflect causal influences independent of external stimuli and task conditions, while modulatory interactions reflect context dependent influences. The observation models produce BOLD-fMRI signals as a linear convolution of latent signals and basis functions spanning the space of variations in HRF.

The latent signals and unknown parameters that characterize causal interactions between brain regions are estimated using two different approaches. In the first approach, we use expectation maximization (EM) to obtain maximum likelihood estimates (MLE) of the parameters and test the statistical significance of the estimated causal relationships between brain regions using a nonparametric approach. We refer to this approach as MDS-MLE. In the second approach, we use a Variational Bayes (VB) approach to compute the posterior distribution of latent variables and parameters which cannot be computed analytically using a fully Bayesian approach. We refer to this approach as MDS-VB. By representing MDS as probabilistic graphical network (Figure 1), we show that MDS-VB provides an elegant analytical solution for computing the posterior distributions and deriving causal connectivity estimates which are sparse and more readily interpretable.

We first describe our MDS model and discuss MLE and VB approaches for estimating intrinsic and modulatory causal interactions between multiple brain regions. We test the performance of MDS using computer-simulated datasets as a function of network size, fMRI time points and signal to noise ratio (SNR). We evaluate performance of our MDS models with extensive computer simulations and examine several metrics, including sensitivity, false positive rate and accuracy, in terms of correctly identifying both intrinsic and modulatory causal interactions. Finally, we contrast our results with those obtained with GCA.

Methods

Notation

In the following sections, we represent matrices by upper case letters and scalars and vectors by lower-case letters. Random matrices are represented by bold face letters whereas random vectors and scalars are represented by bold face lower-case letters.

MDS Model—Consider the following state-space model to represent the multivariate fMRI time series

$$s(t) = As(t-1) + \sum_{j=1}^J v_j(t) C_j s(t-1) + Du(t) + w(t) \quad (1)$$

$$x_m(t) = [s_m(t) s_m(t-1) \dots s_m(t-L+1)]' \quad (2)$$

$$y_m(t) = b_m \Phi x_m(t) + e_m(t) \quad (3)$$

In Equation (1), $s(t)$ is a $M \times 1$ vector of latent signals at time t of M regions, A is an $M \times M$ connection matrix wherein $A(m, n)$ denotes the strength of intrinsic causal connection (which is independent of external stimuli or task condition) from n -th region to m -th region. C_j is an $M \times M$ connection matrix ensued by modulatory input $v_j(t)$, J is the number of modulatory inputs. The non-diagonal elements of C_j represent the coupling of brain regions in the presence of modulatory input $v_j(t)$. Therefore, latent signals $s(t)$ in M regions at time t is a bilinear function of modulatory inputs $v_j(t)$ and its previous state $s(t-1)$. D is an $M \times M$ diagonal matrix wherein $D(i, i)$ denotes external stimuli strength to i -th region. $u(t)$ is an $M \times 1$ binary vector whose elements represent the external stimuli to m -th region under investigation. $w(t)$ is an $M \times 1$ state noise vector whose distribution is assumed to be Gaussian distributed with covariance matrix $Q(w(t) \sim N(0, Q))$. Additionally, state noise vector at time instances $1, 2, \dots, T$ ($w(1), w(2) \dots w(T)$) are assumed to be identical and independently distributed (iid). Equation (1) represents the time evolution of latent signals in M brain regions. More specifically, the latent signals at time t , $s(t)$ is expressed as a linear combination of latent signals at time $t-1$, external stimulus at time t ($u(t)$), bilinear combination of modulatory inputs $v_j(t)$, $j = 1, 2, \dots, J$ and its previous state, and state noise $w(t)$. The latent dynamics modeled in Equation (1) gives rise to observed fMRI time series represented by Equations (2) and (3).

We model the fMRI time series in region “ m ” as a linear convolution of HRF and latent signal $s_m(t)$ in that region. To represent this linear convolution model as an inner product of two vectors, the past L values of $s_m(t)$ are stored as a vector. $x_m(t)$ in equation (2) represents an $L \times 1$ vector with L past values of latent signal at m -th region.

In Equation (3), $y_m(t)$ is the observed BOLD signal at t of m -th region. Φ is a $p \times L$ matrix whose rows contain bases for HRF. b_m is a $1 \times p$ coefficient vector representing the weights for each basis function in explaining the observed BOLD signal $y_m(t)$. Therefore, the HRF in m -th region is represented by the product $b_m \Phi$. The BOLD response in this region is obtained by convolving HRF ($b_m \Phi$) with the L past values of the region's latent signal ($x_m(t)$) and is represented mathematically by the vector inner product $b_m \Phi x_m(t)$.

Uncorrelated observation noise $e_m(t)$ with zero mean and variance σ_m^2 is then added to generate the observed signal $y_m(t)$. $e_m(t)$ is also assumed to be uncorrelated with $w(\tau)$ at all t and τ . Equation (3) represents the linear convolution between the embedded latent signal $x_m(t)$ and the basis vectors for HRF. Here, we use the canonical HRF and its time derivative as bases, as is common in most fMRI studies (Penny et al., 2005; Smith et al., 2009).

Equations (1–3) together represent a state-space model for estimating the causal interactions in latent signals based on observed multivariate fMRI time series. This model can be seen both as a multivariate extension of univariate time series models (Makni et al., 2008; Penny et al., 2005), and also as an extension of GCA wherein vector autoregressive model for latent, rather than BOLD-fMRI, signals are used to model the causal interactions among

brain regions. Furthermore, our MDS model also takes into account variations in HRF as well as the influences of modulatory and external stimuli in estimating causal interactions between the brain regions.

Estimating causal interactions between M regions specified in the model is equivalent to estimating the unknown parameters A and $C_j, j = 1, 2, \dots, J$. In order to estimate A and C_j 's, the other unknown parameters $D, Q, \{b_m\}_{m=1}^M$ and $\{\sigma_m^2\}_{m=1}^M$ and the latent signal $\{s(t)\}_{t=1}^T$ based on the observations $\{y_m(t)\}_{m=1}^M, t = 1, 2, \dots, T$, where T is the total number of time samples, needs to be estimated. We use the following MLE and VB methods for estimating the parameters of the MDS model.

Maximum Likelihood Estimation (MLE)

Estimation: Maximum likelihood estimates of MDS model parameters and latent signals are obtained by maximizing the log-likelihood of the observed fMRI data. We use EM algorithm to estimate the unknown parameters and latent variables of the model. EM algorithm is an iterative method consisting of two steps viz., E-step and M-step. In the E-step, the posterior distribution of latent variables is computed given the current estimates of parameters. In the M-step, given the current posterior distribution of latent variables, the parameters of the model are estimated by maximizing the conditional expectation of log of complete likelihood given the data. The E and M steps are repeated until convergence. The log-likelihood of the data is guaranteed to increase or remain the same for every iteration of E and M steps. Also, the application of EM algorithm asymptotically gives maximum likelihood estimates of the parameters. In the E step, the posterior distributions are obtained by using Kalman filtering and smoothing algorithms (Bishop, 2006). The detailed equations for E and M steps are given in Appendix A. We refer this solution to MDS using MLE as MDS-MLE.

Inference: The statistical significance of intrinsic ($A(m, n)$) and modulatory ($C_j(m, n), j = 1, 2, \dots, J$) causal connections estimated using the EM approach was tested using a Bootstrap method. In this approach, the distribution of connection strengths, under the null hypothesis that there are no connections between the regions, was generated by estimating A and C from 100 surrogate datasets constructed from the observed data. A surrogate dataset was obtained by applying a Fourier transform to observed signal at the m -th region and then randomizing its phase response by adding a random phase shift at every frequency. The phase shifts were obtained by randomly sampling in the interval $[0, 2\pi]$. Inverse Fourier transform was then applied to generate one instance of surrogate data (Prichard and Theiler, 1994). Randomization of the phase response destroys the causal interactions between the brain regions while preserving their power spectra. EM algorithm was then run on this surrogate data to obtain A and C under the null hypothesis. This procedure was repeated on 100 surrogate datasets and the empirical distributions under null hypothesis were obtained for elements of A and C . The statistical significance of each connection was then estimated using these distributions at p value of 0.01. with Bonferroni correction to account for multiple comparisons.

Variational Bayes (VB)

Estimation of Posterior Distributions: In this approach, we use a VB framework to obtain the posterior distributions of the unknown parameters and latent variables. Let $\Theta = \{A, C_1, \dots, C_J, D, Q, R, B\}$ represent the unknown parameters and $S = \{s(t), t = 1, 2, \dots, T\}$ be the latent variables of the model. Given the observations $Y = \{y(t), t = 1, 2, \dots, T\}$ and the probabilistic model, the Bayesian approach aims to find the joint posterior $p(S, \Theta|Y)$. However, obtaining this posterior distribution using a fully Bayesian approach is

analytically not possible for most models including MDS. In the VB approach, we make analytical approximation to $p(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$. Let $q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$ be any arbitrary probability distribution then the log of the marginal distribution of observations can be written as (Bishop, 2006)

$$\log P(\mathbf{Y}) = L(q) + KL(q \| p) \quad (4)$$

Where

$$L(q) = \int d\mathbf{S} d\boldsymbol{\Theta} q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y}) \log \frac{p(\mathbf{Y}, \mathbf{S}, \boldsymbol{\Theta})}{q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})} \quad (5)$$

$$KL(q \| p) = - \int d\mathbf{S} d\boldsymbol{\Theta} q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y}) \log \frac{p(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})}{q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})} \quad (6)$$

$KL(q \| p)$ is the Kullback-Leibler divergence between $q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$ and $p(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$. $KL(q \| p) \geq 0$, with equality, if and only if, $q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y}) = p(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$. Therefore, $L(q)$ serves as a lower bound on the log of the evidence ($\log P(\mathbf{Y})$). The maximum of this lower bound occurs when KL divergence is zero for which the optimal choice of $q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$ is $p(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$. Since $p(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$ is not tractable, certain assumptions on the form of $q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$ are made and then the optimal distribution is found by maximizing the lower bound $L(q)$. In this work, we assume that the posterior distribution $q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y})$ factorizes over $\mathbf{S}$ and $\boldsymbol{\Theta}$, i.e.,

$$q(\mathbf{S}, \boldsymbol{\Theta}|\mathbf{Y}) = q_{\mathbf{S}}(\mathbf{S}|\mathbf{Y}) q_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\mathbf{Y}) \quad (7)$$

We note that no further assumptions are made on the functional form of these distributions $q_{\mathbf{S}}(\mathbf{S}|\mathbf{Y})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\mathbf{Y})$. These quantities are obtained by taking functional derivatives of $L(q)$ with respect to $q_{\mathbf{S}}(\mathbf{S}|\mathbf{Y})$ and $q_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\mathbf{Y})$. It can be shown that

$$\log q_{\mathbf{S}}(\mathbf{S}|\mathbf{Y}) \propto E_{\boldsymbol{\Theta}}(\log p(\mathbf{Y}, \mathbf{S}, \boldsymbol{\Theta})) \quad (8)$$

$$\log q_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\mathbf{Y}) \propto E_{\mathbf{S}}(\log p(\mathbf{Y}, \mathbf{S}, \boldsymbol{\Theta})) \quad (9)$$

Equations (8) and (9) are respectively VB-E and VB-M steps. Expectations are computed with respect to $q_{\boldsymbol{\Theta}}(\boldsymbol{\Theta}|\mathbf{Y})$ in equation (8) and with respect to $q_{\mathbf{S}}(\mathbf{S}|\mathbf{Y})$ in Equation (9). In the VB-E step, the distribution of latent signal $s(t)$, for each t , is updated given the current distribution of the parameters $\boldsymbol{\Theta}$. For reasons described below, $s(t)$ has a Gaussian distribution and in this step updating the distribution amounts to updating the mean and variance of the Gaussian distribution. Therefore, in the VB-E step, estimating means of $s(t)$ at every t is equivalent to estimating the latent signals. In the VB-M step, the distributions for model parameters are updated given the update distributions for latent signal $s(t)$. These VB-E and VB-M steps are repeated until convergence. Note that we do not make any assumptions about the factorization of $\boldsymbol{\Theta}$ and $\mathbf{S}$. Any further conditional independencies in these sets are derived from the probabilistic graphical model of MDS shown in Figure (1). The details of the derivation of the posterior probabilities using the graphical model are

given in Appendix-B. Figure 2 shows a flow chart of various steps involved in both MDS-VB and MDS-MLE methods.

Choice of Priors and Inference: The Bayesian approach allows the specification of both informative and non-informative priors on the model parameters. The specification of these priors helps in regularizing the solution and avoids over fitting in the case where number of parameters to be estimated is large compared to the number of observations. This is generally the case where the number of brain regions to be modeled is large. Since we do not have a priori information on these parameters, we specify non-informative conjugate priors on these parameters. Here, we briefly explain the notion of non-informative conjugate priors (Bishop, 2006). Let z be a Gaussian random variable with mean μ and variance σ^2 . If $\sigma^2 \rightarrow \infty$ or $1/\sigma^2 \rightarrow 0$ (precision), the distribution becomes flat and the random variable z can take any value between $-\infty$ and ∞ with equal probability. Here, we refer to such distributions as non-informative. Let x and y be two random variables with probabilities $p(x)$ and $p(y)$ respectively. $p(y)$ is said to be a conjugate prior for $p(x)$ if the functional form of the posterior $p(x/y)$ is same as that of $p(y)$. Specifying conjugate priors leads to elegant analytical solutions, and it also allows us to specify priors in such a way that we get sparse and interpretable solutions. For example, one can specify the Gaussian priors, on the elements of connection matrices A and C . If the prior on each element of A (or C_j) is

$$A_{i,j} \sim N\left(0, \frac{1}{\lambda_{i,j}}\right) \quad (10)$$

where, $\lambda_{i,j}$ is the prior precision for $A_{i,j}$. Such a specification of priors helps in automatic relevance determination (ARD) of the connections $A_{i,j}$ between the regions (Tipping, 2001). During the learning process of A and λ 's, a significant proportion of $\lambda_{i,j}$'s go towards infinity and the corresponding connections $\lambda_{i,j}$'s have posterior distributions whose mean values shrink towards its prior mean which is zero. The elements of the matrix A which do not have significant values become very small and only the elements which are significant survive. Therefore, adopting this procedure helps in automatically identifying the relevant entries of the matrix A and hence the name "Automatic relevance determination". This is very important because unlike the MLE approach, inference on connection weights (A and C_j 's) is now straightforward. The details of prior specification for various parameters are given in Appendix-B. We test the significance of parameters by thresholding the corresponding posterior probabilities at a p-value of 0.01 with Bonferroni correction to account for multiple comparisons.

Simulated Datasets

(A) Datasets with modulatory effects and external stimuli: We assess the performance of MDS using a number of computer-simulated datasets generated at various SNRs (10, 5 and 0 dB), for different number of brain regions or nodes ($M = 2, 3$ and 5) and for different number of time samples ($T = 200, 300$ and 500).

Figure 3 shows the intrinsic and modulatory connectivity of three networks with 2, 3 and 5 nodes. For example, in the two nodes network (Figure 3A), node 1 receives an external input and there is an intrinsic causal connection from node 1 to node 2 with a weight of -0.3 ($A(2,1) = -0.3$). A modulatory input induces a connection from node 1 to node 2 with weight of 0.5 ($C(2,1) = 0.5$) and whose sign is opposite to that of intrinsic connection. Similarly in the five-node structure (Figure 3C), node 1 receives the external input and has causal influences on nodes 2, 3 and 4 (matrix elements A). Nodes 4 and 5 have bidirectional influences. Modulatory inputs induce causal influences from node 1 to 2 and from node 3 to

2 (matrix C). Note that all three networks have intrinsic and modulatory connections from node 1 to 2 with weights -0.3 and 0.5 respectively. We simulated networks with these weights to explicitly test whether MDS could recover these interactions which could be missed by GCA because of opposing signs of the weights of intrinsic and modulatory connections. We set the autocorrelations of the time series (diagonal elements of matrix A) to 0.7 (Ge et al., 2009; Roebroeck et al., 2005). Our simulations also modeled variations in HRFs across regions. Figure 4 shows the simulated HRFs at each of the network nodes in the structure. The HRFs were constructed in such a way that the direction of hemodynamic delays is confounded with the direction of latent signal delay, making the task of recovering network parameters more challenging. For example in the 5-node network, node 1 drives nodes 2, 3 and 4 at the level of latent signals. But the HRF at node 1 peaks later than that in the nodes 2, 3 and 4. These HRFs were simulated using a linear combination of canonical HRF and its temporal derivative. Analogous to most fMRI data acquisition paradigms, we assume that the sampling interval, also referred to as repetition time (TR), is 2 seconds. This corresponds to an embedding dimension of $L = 16$, which is also the length of the HRF. Note that the duration of the canonical HRF is approximately about 32 seconds.

Figure 5 shows the experimental and modulatory inputs applied to the nodes shown in Figure 3. The external input was simulated to reflect an event-related fMRI design with stimuli occurring at random instances with the constraint that the time difference between two consecutive events is at least 2 TR (Figure 5A). This input can also be a slow event or block design. In the MDS framework, there is no restriction on the nature of the experiment design. The modulatory input is assumed to be a boxcar function (Figure 5B). The modulatory inputs indicate the time periods wherein the network configuration could change because of context specific influences such as changes in attention, alertness and explicit experimental manipulations.

The simulated datasets were generated using the model described in the equations (1)–(3). The latent noise covariance was fixed at $Q = 0.1I_M$, where I_M is identity matrix of size M . The observed noise variance at m -th region for a given SNR was computed as

$$\sigma_m^2 = \text{Var}(y_m) 10^{-0.1SNR} \quad (11)$$

We assume that the canonical HRF and its temporal derivative span the space of HRFs. Therefore, they constitute the rows of Θ which would be a 2×16 matrix. The coefficients of the matrices A and C for each network structure are shown in Figure 3.

We generated 25 datasets for each SNR , network structure and time samples. The performance of the method was assessed using the performance metrics described in the next section.

(B) Datasets without modulatory effects and external stimuli: To examine the performance of the methods when there are no modulatory or external stimulus effects, we simulated 25 datasets for a 3-node network shown in Figure 3(B) at 5 dB SNR. We set the weights to the same values as in the previous case except that there are no causal interactions from modulatory inputs. The weights corresponding to external stimuli were all set to zero. The diagonal elements (autocorrelations) in matrix A were set to 0.8 , 0.7 and 0.6 respectively. These datasets were created to provide a more appropriate, albeit less general, comparison of MDS with GCA.

(C) Effects of fMRI down-sampling: Interactions between brain regions occur at finer time scales compared to the sampling intervals of fMRI signals. fMRI signals are typically sampled at TR = 1 or 2 seconds while neuronal processes occur at millisecond resolution. To investigate the effects of fMRI down sampling on MDS performance, we adopted the approach described by Deshpande and colleagues (Deshpande et al., 2009). We generated datasets with 1 millisecond sampling interval at 0 dB network shown in Figure 3A. We obtained neuronal signals in node 1 and node 2 with a delay of d_n milliseconds between them. In this case, node 1 drives node 2 under both intrinsic and modulatory conditions with weights shown in Figure 3A. The autocorrelations in node 1 and 2 were set to 0.8 and 0.7 respectively. We then convolved neuronal signal at node 1 with a canonical HRF generated again at 1 KHz sampling rate and then re-sampled to sampling interval of TR = 2 seconds to obtain fMRI signal. In node 2, we convolved the “neuronal” signal with the HRF which was delayed by d_h milliseconds with respect to the HRF in node 1, and again re-sampled to TR = 2 seconds to obtain fMRI signal. We obtained simulated datasets at various neuronal delays $d_n = \{0, 200, 400, 600, 800, 1000\}$ and HRF delays $d_h = \{0, 500, 2500\}$ milliseconds (Deshpande et al., 2009). We also examined two cases for HRF delays: (1) HRF delay is in the same direction of neuronal delay and (2) HRF delay is in the opposite direction of neuronal delay. The second case represents the scenario where HRF confounds the causal interactions at neuronal level. We generated 25 simulated datasets for each combination of d_n and d_h . Supplementary Table S1 summarizes the characteristics of each dataset used for evaluating the performance of MDS.

Performance Metrics—The performance of MDS in discovering the intrinsic and modulatory causal interactions in simulated datasets was assessed using various performance metrics such as sensitivity, false positive rate and accuracy in correctly identifying causal intrinsic and modulatory interactions, where:

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (12)$$

$$\text{false positive rate} = \frac{FP}{TN + FP} \quad (13)$$

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

where, TP is the number of true positives, TN is the number of true negatives, FN is the number of false negatives and FP is the number of false positives. These performance metrics are computed for each of the 25 datasets and then are averaged to obtain the overall performance.

Results

Applying MDS – An example

We first illustrate the performance of MDS-MLE and MDS-VB by computing the estimated intrinsic and modulatory connections and the deconvolved (or estimated) latent signals of a five node network simulated at 5 dB noise shown in Figure 3C. The MDS approach, using either MLE or VB, simultaneously estimates both latent signals and unknown parameters in the model using E and M steps respectively. The left and right panels in Figure 6

respectively show the actual and estimated latent signals and the actual and estimated BOLD signal at the five nodes in the network using MDS-MLE and MDS-VB. The estimated BOLD signal $\widehat{y}_m$ at m -th node using these methods was computed as follows

$$\widehat{y}_m = \widehat{b}_m' \Phi \widehat{x}_m(t) \quad (15)$$

Where, $\widehat{b}_m'$ are the estimated coefficients (using MLE or VB) corresponding to the basis functions spanning the subspace of HRFs and $\widehat{x}_m'$ (using MLE or VB) is the estimated latent signal at the m -th node. As shown in this figure, both MDS-MLE and MDS-VB were able to recover the latent and BOLD signals at this SNR. Table-1 shows the mean square error (MSE) between the estimated and latent signals and estimated and actual BOLD-fMRI responses in each node using these two methods. The MSE in estimating these signals is very low by both methods. Figures 7A and B respectively show the estimated intrinsic and modulatory causal interactions by MDS-MLE and MDS-VB in the simulated five node network. MDS-VB correctly identified both intrinsic (solid lines) and modulatory connections (dotted lines) in this network as shown in Figure 7B. MDS-MLE also correctly recovered both intrinsic and modulatory networks but it introduced an additional false modulatory connection from node 3 to node 1 as shown in Figure 7A.

We next compare the performance of MDS with that of GCA using the same simulated data. This analysis was performed using the multivariate GCA toolbox developed by Seth (Seth, 2010). We applied GCA on the same dataset to verify whether it can recover the causal connections (either intrinsic or modulatory). As shown in Figure 7C, GCA likely missed both the intrinsic and modulatory interactions from node 1 to 2 but it was able to recover modulatory interactions from node 3 to 4 in addition to other connections. However, unlike MDS, GCA cannot distinguish between intrinsic and modulatory interactions. GCA missed the connection from node 1 to 2 because these nodes have both intrinsic and modulatory interactions with opposing actions. Since GCA does not model these interactions separately, the net connection strength between these nodes is not significant. On the other hand, MDS models these interactions explicitly and therefore is able to recover both types of connections. This example demonstrates that GCA cannot recover all the connections under these conditions while both MDS methods could recover all the connections and at the same time differentiate between the different types of interactions.

Performance of MDS on simulated data with modulatory effects and external stimuli

We evaluated the performance of MDS-MLE and MDS-VB on simulated datasets by computing sensitivity, false positive rate and accuracy in finding intrinsic and modulatory causal interactions as a function of SNR, network size and the number of time samples. Figures 8, 9 and 10 respectively show the performance of MDS-MLE and MDS-VB for time samples $T = 500, 300, 200$. For each T and network size, the performance of MDS-MLE and MDS-VB is evaluated for $SNR = 0, 5$ and 10 dB. In each of these figures panels A, B and C show the performance of MDS-MLE and MDS-VB with respect to sensitivity, false positive rate and accuracy in identifying the intrinsic and modulatory interactions for 2, 3 and 5 node networks respectively. The performance of these methods improved with the increase in SNR and time samples (T).

Between the two MDS methods, MDS-VB showed superior performance compared to MDS-MLE across all the SNRs, time samples and network size. MDS-VB showed significantly greater performance at low SNR and for shorter time series. For example, for $SNR = 0$ dB, $T = 200$ time points and number of nodes = 5 (Figure 10C), the sensitivity of

MDS-VB in recovering intrinsic and modulatory interactions was about 0.75 and 0.6 respectively while MDS-MLE has sensitivity of only about 0.3 and 0.5 respectively. The accuracy of MDS-VB is also high (>0.8) under all conditions (panel (C) in Figures 8, 9 and 10) because the sensitivities are high and false positive rates of this method is very low (panels A and B in Figures 8, 9 and 10). More generally, in cases with high noise, lower sample length, and larger network size MDS-VB consistently outperforms MDS-MLE.

Comparison of MDS and GCA on simulated data with modulatory effects and external stimuli

Finally, we compared the performance of GCA with MDS methods on the same datasets and performance metrics. Figures 8, 9 and 10 respectively show the comparative performance of GCA with MDS-VB and MDS-MLE for $T = 500, 300$ and 200 at various SNRs and network sizes. The results suggest that performance of GCA is poor compared to both MDS methods with respect to sensitivity and accuracy in identifying causal interactions between brain nodes. Since GCA cannot distinguish between intrinsic and modulatory interactions, we computed the performance metrics by considering both the connection types. The performance of GCA declined with the decrease in SNR for networks of size 3 and 5 (Figures 8–10). The performance of GCA is worse even for a 2-node network as shown in Figures 8A, 9A and 10A. The sensitivity of GCA for this network is less than 10% because intrinsic and modulatory interactions have weights with opposite signs. Since GCA does not model these interactions explicitly, it does not detect the interactions between the two nodes. On the other hand, both MDS methods showed better sensitivity and accuracy in identifying both types of interactions at all SNRs for this network.

Comparison of MDS and GCA on simulated data in the absence of modulatory effects and external stimuli

Table 2 shows the relative performance of MDS-VB, MDS-MLE and GCA on 25 datasets simulated for a 3-node network at 5 dB SNR without any modulatory effects and external stimuli. The performance of GCA improved and recovered the causal network with sensitivity of 0.9, FPR of 0 and accuracy of 0.98. In this case, the performance of GCA is comparable to MDS, suggesting that in the absence of modulatory effects and external stimuli GCA can perform as well as MDS even in the presence of HRF variations.

Effects of fMRI down-sampling on MDS performance

We examined the performance of MDS-VB on simulated data in which latent signals were generated at various delays using a sampling interval of 1 millisecond and convolved with various delays in the HRF. Causal interactions were then estimated based on the observed time series obtained at sampling interval of 2 seconds. We examined the performance of MDS-VB under four different cases:

(a) No latent signal delay but HRF is delayed by 500 and 2500 milliseconds—In this case, there are no causal interactions between the two nodes with respect to the latent signals but the observed fMRI time series are delayed with respect to each other because of delays in their respective HRFs. In this case, MDS-VB was accurate in that it did not recover any causal interactions (both intrinsic and modulatory) despite variations in HRF.

(b) No HRF delays—In this case, HRFs in both nodes are identical. As shown in Figure S1(A), the sensitivity of MDS-VB in recovering both intrinsic and modulatory interactions is above 0.9 (left panel) with FPRs below 0.1 (middle panel) and therefore with accuracies of above 0.9 (right panel) at all the latent signal delays. The performance of MDS-VB improved with the increase in latent signal delays.

(c) Latent and HRF delays both in the same direction—In this scenario, HRF delays do not confound the causal interactions between the nodes at the latent signal level. For HRF delay of 500 and 2500 milliseconds, the performance metrics shown in Figures S1 (B) and S1(C) suggest that MDS-VB is able to recover both intrinsic and modulatory causal interactions reliably. For the HRF delay of 2500 milliseconds, there is a small drop in sensitivity for latent signal delays of 800 and 1000 milliseconds.

(d) HRF delays oppose the delays in latent signals—This is the most difficult situation for any method because HRF delays confound the causal interactions at latent signal level. Figure S2 (B) shows the performance of MDS-VB when HRF in node 2 peaks 500 milliseconds before the HRF in node 1 while node 1 drives node 2 at latent signal level. The performance of MDS-VB improved with the increase in latent signal delay. Figure S2(C) shows the performance of MDS-VB for HRF delay of 2500 milliseconds. Although the sensitivities for latent signal delays from 200 to 600 milliseconds are higher (left panel) but are accompanied by greater false positives (middle panel) and therefore have poor accuracies (right panel). The performance of MDS-VB improved for latent signal delays of 800 and 1000 milliseconds in recovering causal interactions.

Discussion

We have developed a novel dynamical systems method to model intrinsic and modulatory interactions in fMRI data. MDS uses a vector autoregressive state-space model incorporating both intrinsic and modulatory causal interactions. Intrinsic interactions reflect causal influences independent of external stimuli and task conditions, while modulatory interactions reflect context dependent influences. Our proposed MDS method overcomes key limitations of commonly used methods for estimating the causal relations using fMRI data.

Critically, causal interactions in MDS are modeled at the level of latent signals, rather than at the level of the observed BOLD-fMRI signals. Our simulations clearly demonstrate that this has the added effect of eliminating the confounding effects of regional variability in HRF. The parameters and latent variables of the state-space model were estimated using two different methods. In the MDS-MLE method, the statistical significance of the parameters of the state equation, which represent the causal interactions between multiple brain nodes, was tested using a Bootstrap method. In the MDS-VB method, we used non-informative priors to facilitate automatic relevance detection. We first discuss findings from our simulations, and show that MDS-VB provides the robust and accurate solutions even at low SNRs and smaller number of observed samples (time points). We then contrast the performance of MDS with the widely used GCA method. In this context, we highlight instances where GCA works reasonably well and where it fails. Finally, we discuss several important conceptual issues concerning the investigation of dynamic causal interactions in fMRI, contrasting MDS with other recently developed methods.

Performance of MDS on simulated datasets – contrasting MLE and VB approaches

In the following sections, we evaluate and discuss the performance of MDS under various scenarios. Importantly, we demonstrate, for the first time, that VB approaches provide better estimates of model parameters than MLE based approaches. We investigated the performance of MDS-MLE and MDS-VB on simulated datasets generated at SNRs of 0, 5 and 10 dB for network structures of sizes 2, 3 and 5 and time samples of 200, 300 and 500. We simulated regional HRF variations in such a way that the directions of hemodynamic response delays were in opposite direction to the delays in the latent signals (Figure 4). HRF delays could therefore influence the estimation of causal interactions when applied directly

on the observed BOLD-fMRI signals. This makes the problem of estimating causal interactions particularly challenging and provides novel insights into strengths and weaknesses of multiple approaches used here.

The performance of MDS was found to be robust when tested under various simulated conditions. Specifically, MDS was able to reliably recover both intrinsic and modulatory causal interactions from the simulated datasets and its performance was found to be superior to the conventional approaches such as GCA. Among MDS methods, the performance of MDS-VB was found to be superior to MDS-MLE with respect to performance metrics such as sensitivity, false positive rate and accuracy in identifying intrinsic and modulatory causal interactions (Figures 7–10). MDS-VB showed significantly improved performance over MDS-MLE, especially at adverse conditions such as low SNRs, large network size and for less number of observed samples (Figure 10C).

The superior performance of MDS-VB can be attributed to the regularization imposed by priors in this method. Our priors not only regularized the solution but also helped in achieving sparse solutions. By using sparsity promoting priors, the weights corresponding to insignificant links are driven towards zero and therefore enable automatic relevance detection (Tipping, 2001). This approach is not only useful for regularizing solutions when the number of unknown parameters is high, but also for providing sparse and interpretable solutions. This feature of VB can be especially important in analyzing networks with large number of nodes, an aspect often overlooked in most analyses of causality in complex networks.

Another advantage of Bayesian analysis lies in computing the statistical significance of the network connections estimated by MDS methods. In the MLE approach, we need to resort to a Bootstrap approach, which can be computationally expensive. MDS-VB, on the other hand, provides posterior probabilities of each model parameter, as opposed to point estimates in MLE, which can be used to compute their statistical significance. From a computational perspective, MDS-VB is several orders of magnitude faster than MLE-MLE because it does not require nonparametric tests for statistical significance testing. Taken together, these findings suggest that MDS-VB is a superior and more powerful method than MDS-MLE.

Comparison with GCA

We demonstrated the importance of modeling the influence of both external and modulatory stimuli for estimating the causal networks by applying GCA on a five-node network. On this dataset, GCA failed to detect both the modulatory and intrinsic connections between nodes 1 and 2 (Figure 7C). As mentioned earlier, GCA missed this connection because the network has both intrinsic and modulatory connections between these nodes but with weights of opposite signs. Therefore, in GCA the net strength of this connection is very small and it did not survive a conservative test of statistical significance. Our MDS methods, on the other hand, uncovered both these connections. This phenomenon is most obvious in our simulations of the 2-node network wherein GCA could not find causal interactions between the nodes (Figures 8A, 9A and 10A). In this network also both intrinsic and modulatory connections have weights with opposite signs. These results demonstrate the importance of explicitly modeling the influence of external and modulatory stimuli. Overall, the performance of GCA, when applied on the datasets generated at various SNR, networks and for different number of observations, was found to be inferior when compared to both MDS methods. This was true with respect to both sensitivity and accuracy in identifying causal interactions between multiple brain nodes (Figures 8, 9 and 10). When compared to MDS, the performance of GCA drops significantly at lower SNRs. These results suggest that MDS is more robust against observation noise than GCA. Therefore, our simulations suggest that

MDS-VB outperforms GCA for networks consisting of less than 6 nodes. More extensive simulations however are needed to compare the performance of MDS with GCA, for larger networks.

Conventional GCA methods do not take into account dynamic changes in modulatory inputs and their effect on context dependent causal interactions. In order to compare GCA more directly with MDS, we examined causal interactions in the absence of modulatory influences. As expected, in this case, the performance of GCA was comparable to that of MDS. Together, these findings suggest that GCA can accurately recover causal interactions in the absence of modulatory effects. Although newer dynamic GCA methods have been proposed, they appear to be designed more for improving estimations of causal interactions rather than examining context dependent dynamic changes in causal interactions (Havlicek et al., 2010; Hemmelmann et al., 2009; Hesse et al., 2003; Sato et al., 2006). Further simulation studies are needed to assess how well dynamic GCA can estimate context specific modulatory effects.

We next contrast our findings using GCA and MDS in the context of the equivalence of ARIMA and structural time series models. GCA is based on autoregressive integrated moving average (ARIMA) models proposed by Box and Jenkins whereas MDS is a structural time series model (Box. et al., 1994). In the econometrics literature, it is well known that the linear structural time series models have equivalent ARIMA model representations (Box. et al., 1994). This equivalence has been under-appreciated in the neuroimaging literature, as demonstrated by the recent discussion regarding the relative merits of GCA and DCM (Friston, 2009a, b; Roebroeck et al., 2009). Our detailed simulations suggest that, under certain conditions, GCA is able to recover much of the causal network structure in spite of the presence of HRF delay confounds. This is most clearly illustrated using the simulations shown in Figure 7C, where we found that GCA could recover the network structure except for the intrinsic/modulatory connection from node 1 to 2. Our simulations also suggest that GCA may not be able to uncover causal connections when there is a conflict between intrinsic and modulatory connections (Figure 7C, 8, 9 and 10) but for other cases it is able to recover the underlying networks. In estimating the causal interactions, the estimated model order for GCA using the Akaike information criterion (AIC) was more than 3. Note that in our simulations, the causal interactions at the latent signal level were generated using VAR with model order 1 (Equation 1). It is plausible that in GCA, the higher model order is used to compensate for variations in HRF delay and experimental effects such as context specific modulatory connections (Deshpande et al., 2009). Our simulations suggest that this is indeed that case and that optimal model order selection in GCA results in improved estimation of causal interactions between nodes. Nevertheless, structural time series based models like MDS and DCM can provide better interpretation of network structure since they can distinguish between intrinsic and context specific modulatory causal interactions in latent signals.

Effects of down-sampling on MDS performance

In most fMRI studies, data are typically acquired at sampling rates of about 2 seconds (or TR = 2seconds). However, dynamical interactions between brain regions occur at faster time scales of 10–100 milliseconds. To examine the effects of downsampling fMRI data on the performance of MDS, we first simulated interactions between nodes at a sampling rates of 1KHz and then re-sampled the time series to 0.5 Hz after convolving with region-specific HRFs. MDS-VB was then applied on these datasets to estimate causal interactions between nodes. We also examined the influence of HRF delays on the estimation of causal interactions under four scenarios (Figures S1 and S2), similar to the strategy used by Deshpande and colleagues (Deshpande et al., 2009) to study the effect of HRF variability on GCA. In the first scenario, there were no causal interactions between nodes but HRFs were

delayed between the nodes. In this case, MDS-VB performed accurately and did not infer any false causal interactions. This shows that MDS-VB can model and remove the effects HRF variation while estimating causal interactions at latent signal level. In the second scenario, we introduced causal interactions between nodes, but without HRF variations. MDS reliably estimated causal interactions for various delays in latent signals (Figure S1A). In the third scenario, we introduced causal interactions between nodes and also varied HRFs such that delays in latent signals and HRFs were in the same direction. In this case, MDS was able to recover both intrinsic and modulatory causal interactions accurately (Figure S1B, S1C). In the fourth scenario, when delays in latent signal and HRF opposed each other performance dropped significantly, just as with GCA (Deshpande et al., 2009). Further research is needed to examine whether causal interactions under this scenario are inherently unresolvable by MDS and other techniques such as DCM.

Comparison of MDS with other approaches

As noted above, like GCA, MDS can be used to estimate causal interactions between a large numbers of brain nodes. Unlike GCA, however, causal interactions are estimated on the underlying latent signals while simultaneously accounting for regional variations in the HRF. Furthermore, unlike GCA, MDS can differentiate between intrinsic and stimulus-induced modulatory interactions. Like DCM, MDS takes into account regional variations in HRF while estimating the causal interactions between brain regions. And like DCM, MDS also explicitly models external and modulatory inputs, allowing us to simultaneously estimate intrinsic and modulatory causal interactions between brain regions. Unlike DCM, however, MDS does not require the investigator to test multiple models and choose one with the highest model evidence. This overcomes an important limitation of DCM - as the number of brain regions of interest increases, an exponentially large number of models needs to be examined; as a result, the computational burden in evaluating these models and identifying the appropriate model can become prohibitively high. MDS overcomes such problems and as our study illustrates, MDS incorporates the relative merits of both GCA and DCM while attempting to overcome their limitations.

Both DCM and MDS are state-space models but DCM uses a deterministic state model, (although a stochastic version has been recently developed (Daunizeau et al., 2009)) whereas MDS employs a stochastic model. Modeling latent interactions as a stochastic process is important for taking into account intrinsic variations in latent signals that are not induced by experimental stimuli. Another important difference is that MDS uses empirical basis functions to model variations in HRF whereas DCM uses a biophysical Balloon model (Friston et al., 2003). Since the Balloon model is a nonlinear model, several approximations are required to solve it. In contrast, empirical HRF basis functions allow MDS to use a linear dynamical systems framework. The relative accuracy of these approaches is currently not known.

One important advantage of MDS is that it does not assume that the fMRI time series is stationary unlike methods based on vector autoregressive modeling such as GCA. This is important because the dynamics of the latent signals can be altered significantly by experimental stimuli, leading to highly non-stationary signals. In GCA, the time series is tested for stationarity by either examining the autocorrelation of the time series or by investigating the presence of unit roots. If the time series is found to be nonstationary then one commonly used approach to remove non-stationarity is to replace the original time series with a difference of the current and previous time points (Seth, 2010). A problem with the use of such a manipulation is it acts as a high-pass filter that can significantly distort the estimated causal interactions (Bressler and Seth, 2010).

Two methods based on dynamical systems based approach for modeling fMRI data have been proposed recently (Ge et al., 2009; Smith et al., 2009). Smith and colleagues used a switching linear dynamical systems model wherein modulatory inputs were treated as random variables (Smith et al., 2009). In contrast, MDS models them as deterministic quantities which are known for a given fMRI experiment. Modeling modulatory inputs as unknown random variables is useful for fMRI experiments in which the occurrence of modulatory inputs is unknown. However, for most fMRI studies modulatory inputs are known and modeling them as unknown quantities unnecessarily increases the number of parameters to be estimated. Also, the switching dynamical systems model makes additional assumptions in computing the probability distributions of the state variables (Murphy, 1998). Further, Smith and colleagues used an MLE approach to estimate latent signals and model parameters. As we show in this study, compared to MLE, a VB approach yields more robust, computationally efficient and accurate model estimation even when the SNR and the number of time points are low, as is generally the case with fMRI data. Another difference is that Smith and colleagues combine intrinsic and modulatory matrices i.e., for *every j-th* modulatory input, the connection matrix ($A_j = A + C_j$) is estimated from the data. In MDS, we estimate intrinsic and modulatory matrices C_j separately which explicitly dissociates intrinsic and modulatory effects on causal interactions between brain regions. Another difference lies in testing the statistical significance of the estimated causal connections. In MLE- MDS, we use a non-parametric approach and in MDS-VB, posterior probabilities of the model parameters are used for testing the significance of the causal interactions. Finally, it should be noted that the performance of this method under varying SNRs and sample sizes is not known since no simulations were performed.

Ge and colleagues (2009) used a different state-space approach to estimate causal interactions in the presence of external stimuli. They used vector autoregressive modeling for the state equation to model causal interactions among brain regions, whereas the observation model was nonlinear. They used an extended Kalman filtering approach to estimate the state variables and model parameters. This method was applied on local field potential data, so its usefulness for fMRI data is unclear. However, there are several differences between MDS and this approach. MDS has been developed explicitly for fMRI data to account for HRF variations in brain regions while simultaneously estimating causal interactions. In the work of Ge and colleagues, both state variables and unknown model parameters were treated as state variables and extended Kalman filtering is used to obtain maximum likelihood estimates of these variables (Ge et al., 2009). In MDS, we have taken a different approach - state variables are separated from model parameters. This allowed us to use sparsification promoting priors in the MDS-VB approach. Our results on simulated data suggest that MDS-VB outperforms MDS-MLE especially at low SNRs and smaller number of temporal observations. Finally, Ge and colleagues used Kalman filtering to estimate state variables while in MDS we used Kalman smoothing for estimating the latent signals (Ge et al., 2009). In Kalman smoothing, both past and future data is used to estimate latent signals whereas filtering uses only past values to estimate the current values. In general, smoothing provides better estimates of latent signal than the filtering approach (Bishop, 2006). Finally, although Ge and colleagues validated their approach on two three-node toy models, the performance of this method is not known under varying conditions such as different SNRs, network sizes and number of data samples.

Conclusions

The Bayesian multivariate dynamical system framework we have developed here provides a robust method for estimating and interpreting causal network interactions in simulated BOLD-fMRI data. Extensive computer simulations demonstrate that this MDS method is more accurate and robust than GCA and among the MDS methods developed here MDS-VB

exhibits superior performance over MDS-MLE. Critically, MDS estimates both intrinsic and experimentally induced modulatory interactions in the latent signals, rather than the observed BOLD-fMRI signals. Unlike DCM, our proposed MDS framework does not require testing multiple models and may therefore be more useful for analyzing networks with a large number of nodes and connections. One limitation of this work is that our simulations were based on datasets created using the same model as the one used to estimate causal interactions. In this vein, preliminary analysis using simulations with delayed latent signals at millisecond temporal resolution suggests that MDS can accurately recover intrinsic and modulatory causal interactions in the presence of confounding delays in HRF. Future studies will examine the performance of MDS using more realistic simulations in which causal influences are generated independently of any one particular model, as well as the application of MDS real experimental fMRI data (Menon et al. In Preparation).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grants from the National Institutes of Health (HD047520, HD059205, HD057610) and the National Science Foundation (BCS-0449927).

References

- Abler B, Roebroeck A, Goebel R, Hose A, Schonfeldt-Lecuona C, Hole G, Walter H. Investigating directed influences between activated brain areas in a motor-response task using fMRI. *Magn Reson Imaging*. 2006; 24:181–185. [PubMed: 16455407]
- Bishop, C. *Pattern Recognition and Machine Learning*. Springer; 2006.
- Box, GEP.; Jenkins, GM.; Reinsel, GC. *Time Series Analysis Forecasting and Control*. Pearson Education; 1994.
- Bressler SL, Menon V. Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn Sci*. 2010
- Bressler SL, Seth AK. Wiener-Granger Causality: A well established methodology. *Neuroimage*. 2010
- Cassidy MJ, Penny WD. Bayesian nonstationary autoregressive models for biomedical signal analysis. *IEEE Trans Biomed Eng*. 2002; 49:1142–1152. [PubMed: 12374338]
- Daunizeau J, Friston KJ, Kiebel SJ. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D*. 2009; 238:2089–2118. [PubMed: 19862351]
- Deshpande G, Hu X, Stilla R, Sathian K. Effective connectivity during haptic perception: a study using Granger causality analysis of functional magnetic resonance imaging data. *Neuroimage*. 2008; 40:1807–1814. [PubMed: 18329290]
- Deshpande G, Sathian K, Hu X. Effect of hemodynamic variability on Granger causality analysis of fMRI. *Neuroimage*. 2009
- Friston K. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLoS Biol*. 2009a; 7:e33. [PubMed: 19226186]
- Friston K. Dynamic causal modeling and Granger causality Comments on: The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *Neuroimage*. 2009b
- Friston KJ. Modalities, modes, and models in functional neuroimaging. *Science*. 2009c; 326:399–403. [PubMed: 19833961]
- Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage*. 2003; 19:1273–1302. [PubMed: 12948688]
- Fuster JM. The cognit: a network model of cortical representation. *Int J Psychophysiol*. 2006; 60:125–132. [PubMed: 16626831]

- Ge T, Kendrick KM, Feng J. A novel extended Granger Causal Model approach demonstrates brain hemispheric differences during face recognition learning. *PLoS Comput Biol*. 2009; 5:e1000570. [PubMed: 19936225]
- Glover GH. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*. 1999; 9:416–429. [PubMed: 10191170]
- Goebel R, Roebroeck A, Kim DS, Formisano E. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn Reson Imaging*. 2003; 21:1251–1261. [PubMed: 14725933]
- Guo S, Wu J, Ding M, Feng J. Uncovering interactions in the frequency domain. *PLoS Comput Biol*. 2008; 4:e1000087. [PubMed: 18516243]
- Havlicek M, Jan J, Brazdil M, Calhoun VD. Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *Neuroimage*.
- Hemmelmann D, Ungureanu M, Hesse W, Wustenberg T, Reichenbach JR, Witte OW, Witte H, Leistritz L. Modelling and analysis of time-variant directed interrelations between brain regions based on BOLD-signals. *Neuroimage*. 2009
- Hesse W, Moller E, Arnold M, Schack B. The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies. *J Neurosci Methods*. 2003; 124:27–44. [PubMed: 12648763]
- Koller, D.; Friedman, N. Probabilistic graphical models Principles and Techniques. The MIT Press; 2009.
- Makni S, Beckmann C, Smith S, Woolrich M. Bayesian deconvolution of [corrected] fMRI data using bilinear dynamical systems. *Neuroimage*. 2008; 42:1381–1396. [PubMed: 18603003]
- Mechelli A, Price CJ, Noppeney U, Friston KJ. A dynamic causal modeling study on category effects: bottom-up or top-down mediation? *J Cogn Neurosci*. 2003; 15:925–934. [PubMed: 14628754]
- Murphy, KP. Technical report. DEC/Compaq Cambridge Research Labs; 1998. Switching Kalman Filters.
- Passingham RE, Stephan KE, Kotter R. The anatomical basis of functional localization in the cortex. *Nat Rev Neurosci*. 2002; 3:606–616. [PubMed: 12154362]
- Penny W, Ghahramani Z, Friston K. Bilinear dynamical systems. *Philos Trans R Soc Lond B Biol Sci*. 2005; 360:983–993. [PubMed: 16087442]
- Prichard D, Theiler J. Generating surrogate data for time series with several simultaneously measured variables. *Phys Rev Lett*. 1994; 73:951–954. [PubMed: 10057582]
- Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE*. 1989; 77:257–285.
- Rajapakse JC, Zhou J. Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage*. 2007; 37:749–760. [PubMed: 17644415]
- Ramsey JD, Hanson SJ, Hanson C, Halchenko YO, Poldrack RA, Glymour C. Six problems for causal inference from fMRI. *Neuroimage*. 2009; 49:1545–1558. [PubMed: 19747552]
- Roebroeck A, Formisano E, Goebel R. Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage*. 2005; 25:230–242. [PubMed: 15734358]
- Roebroeck A, Formisano E, Goebel R. The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *Neuroimage*. 2009
- Sato JR, Junior EA, Takahashi DY, de Maria Felix M, Brammer MJ, Morettin PA. A method to produce evolving functional connectivity maps during the course of an fMRI experiment using wavelet-based time-varying Granger causality. *Neuroimage*. 2006; 31:187–196. [PubMed: 16434214]
- Seth AK. Causal connectivity of evolved neural networks during behavior. *Network*. 2005; 16:35–54. [PubMed: 16350433]
- Seth AK. A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Methods*. 2010; 186:262–273. [PubMed: 19961876]
- Smith JF, Pillai A, Chen K, Horwitz B. Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. *Neuroimage*. 2009

- Sridharan D, Levitin DJ, Menon V. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A*. 2008; 105:12569–12574. [PubMed: 18723676]
- Tipping M. Sparse Bayesian learning and relevant vector machine. *Journal of Machine Learning Research*. 2001; 1:211–244.
- Valdes-Sosa PA, Sanchez-Bornot JM, Lage-Castellanos A, Vega-Hernandez M, Bosch-Bayard J, Melie-Garcia L, Canales-Rodriguez E. Estimating brain functional connectivity with sparse multivariate autoregression. *Philos Trans R Soc Lond B Biol Sci*. 2005; 360:969–981. [PubMed: 16087441]

Appendix A

In this appendix, we provide detailed equations for estimating the model parameters and latent states of MDS using an expectation maximization algorithm.

Solving MDS Using Maximum Likelihood Estimation

The state space and observation equations (1)–(3) can be expressed in the standard state form so that Kalman filtering and smoothing recursive equations can be used to estimate the probability distribution of latent signals which constitutes the E-step in our EM algorithm (Penny et al., 2005).

Let

$$\mathbf{x}(t) = [\mathbf{s}'(t) \mathbf{s}'(t-1) \dots \mathbf{s}'(t-L+1)]' \quad (\text{A.1})$$

The equations (1)–(3) can be written in terms of $\mathbf{x}(t)$ as

$$\mathbf{x}(t) = \tilde{G}(t)\mathbf{x}(t-1) + \tilde{D}u(t) + \tilde{w}(t) \quad (\text{A.2})$$

$$y(t) = B\tilde{\Phi}\mathbf{x}(t) + e(t) \quad (\text{A.3})$$

where,

$$\begin{aligned} G(t) &= A + \sum_{j=1}^J v_j(t) C_j, & \tilde{G}(t) &= \tilde{A} + \sum_{j=1}^J v_j(t) \tilde{C}_j \\ \tilde{A} &= \begin{pmatrix} A & 0_{M \times (L-1)} \\ \psi & 0_M \end{pmatrix}, & \tilde{C}_j &= \begin{pmatrix} C_j & 0_{M \times (L-1)} \\ \psi & 0_M \end{pmatrix} \end{aligned} \quad (\text{A.4})$$

ψ is the $M(L-2) \times ML$ delay matrix that fills the lower rows of $\tilde{A}$.

Similarly,

$$\tilde{D} = \begin{pmatrix} D & 0_{M \times (L-1)} \\ 0_{M(L-2) \times ML} & 0_M \end{pmatrix} \quad (\text{A.5})$$

$$U(t) = \begin{bmatrix} u'(t) & 0_{1,M(L-1)} \end{bmatrix}' \quad (\text{A.6})$$

$$\tilde{w}(t) = \begin{bmatrix} w'(t) & 0_{1,M(L-1)} \end{bmatrix}' \quad (\text{A.7})$$

$$\tilde{w}(t) \sim N(0, \tilde{Q}) \quad (\text{A.8})$$

$$\tilde{Q} = \begin{pmatrix} Q & 0_{M \times (L-1)} \\ 0_{M(L-2) \times ML} & 0_M \end{pmatrix} \quad (\text{A.9})$$

$$B = \begin{bmatrix} b_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_M \end{bmatrix} \quad (\text{A.10})$$

$$\tilde{\Phi} = \begin{bmatrix} \cdots \\ \vdots & \ddots & \vdots \\ \cdots \end{bmatrix} \quad (\text{A.11})$$

$$e(t) = [e_1(t) \ e_2(t) \ \cdots \ e_m(t)]' \quad (\text{A.12})$$

$$e(t) \sim N(0, R) \quad (\text{A.13})$$

$$R = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2) \quad (\text{A.14})$$

Let $x(0)$ be the initial state and uncorrelated with state noise $\tilde{W}(t)$ and observation noise $e(t)$ and is normally distributed with mean μ_o and covariance Σ_o . The equations (9) and (10) are now in standard linear state-space system. Therefore, Kalman filtering and smoothing recursions can be used to carry out the E-step.

E-step

In E-step, the probability of latent signal $x(t)$, $t = 1, 2, \dots, T$ given the observed data $y(t)$, $t = 1, 2, \dots, T$ is computed. Since the state, observation noise and initial state $x(0)$ are assumed to be Gaussian, the latent signals $x(t)$, $t = 1, 2, \dots, T$ are also normally distributed whose means and covariances can be estimated by forward (filtering) and backward (smoothing) recursion steps in Kalman estimation framework. In the filtering step, mean and covariance of $x(t)$ are computed given the observations $y(\tau)$, $\tau = 1, 2, \dots, t$. In the smoothing step, the above means

and covariances are updated such that they represent mean and covariance at each time t given all the data $y(t)$, $t = 1, 2, \dots, T$.

Kalman Filtering

In the filtering step, the goal is to compute the following posterior distribution of $x(t)$ given the observations $y(\tau)$, $\tau = 1, 2, \dots, t$ and the parameters of the model

$$p(x(t)|y(1), y(2) \dots y(t)) = N(x_t^f, \sum_t^f) \quad (\text{A.15})$$

The mean and covariance of this distribution can be computed using the following forward recursive steps.

$$x_t^{f-1} = \tilde{G}x_{t-1}^{f-1} + \tilde{D}U(t) \quad (\text{A.16})$$

$$\sum_t^{f-1} = \tilde{G}(t) \sum_{t-1}^{f-1} \tilde{G}' + \tilde{Q} \quad (\text{A.17})$$

$$K(t) = \sum_t^{f-1} E' (E \sum_t^{f-1} E' + R)^{-1} \quad (\text{A.18})$$

where, $E = B\tilde{\Phi}$

$$x_t^f = x_t^{f-1} + K(t)(y(t) - E x_t^{f-1}) \quad (\text{A.19})$$

$$\sum_t^f = \sum_t^{f-1} - K(t)E \sum_t^{f-1} \quad (\text{A.20})$$

The above recursion is initialized using $x_1^0 = \mu_0$ and $\sum_1^0 = \sum_\sigma$.

Kalman Smoothing

In the smoothing step, the goal is to compute the posterior distribution of $x(t)$ given all the observations $y(\tau)$, $\tau = 1, 2, \dots, T$ and the parameters

$$p(x(t)|y(1), y(2) \dots y(T)) = N(x_t^T, \sum_t^T) \quad (\text{A.21})$$

The mean x_t^T and covariance $\sum_t^T$ at each time t can be estimated using the following backward recursions.

$$x_t^T = x_t^T + J_t(x_{t+1}^T - \tilde{G}(t)x_t^T) \quad (\text{A.22})$$

$$\sum_t^T = \sum_t^t + J_t(\sum_t^{T+1} - \sum_t^{t-1})J_t' \quad (\text{A.23})$$

where, J_t is defined as

$$J_t = \sum_t^t \tilde{G}'(t) (\sum_t^{t-1})^{-1} \quad (\text{A.24})$$

The above backward recursions are initialized noting the fact that for $t=T$, x_t^T and $\sum_t^T$ can be obtained from equations (21) and (22) respectively.

M-step

In the M-step, the goal is to find the unknown parameters $\theta = \{A, C_1, \dots, C_J, D, Q, B, R\}$ given the data and the current posterior distributions of $x(t)$, $t = 1, 2, \dots, T$. The parameters can be estimated by maximizing the expected complete log-likelihood of the data and therefore the resulting estimates are called as maximum likelihood estimates.

The complete log-likelihood of the data is given by

$$L = \log p(x(1), x(2) \dots x(T), y(1), y(2) \dots y(T) | \Theta) = \log p(x(1) | \Theta) + \sum_{t=2}^T \log p(x(t) | x(t-1), \Theta) + \sum_{t=1}^T \log p(y(t) | x(t), \Theta) \quad (\text{A.25})$$

Estimation of A, C_j's and D

The complete log-likelihood that depends on the parameters A, C and D is given by

$$L(A, C_1, \dots, C_J, D) \propto -0.5 \sum_{t=2}^T (x_s(t) - (A + \sum_{j=1}^J v_j(t)C_j)F x(t-1) - du(t))' Q^{-1} (x_s(t) - (A + \sum_{j=1}^J v_j(t)C_j)F x(t-1) - du(t)) \quad (\text{A.26})$$

Where, $F = [I_M \ 0_{M \times (L-1)}]$, $d = \text{diag}(D)$ and $x_s(t) = x(1:M, t) = s(t)$.

Taking expectations of equation (A.26) with respect to $p(x(t) | y(1), y(2), \dots, y(T))$ and then differentiating the resulting equation with A, C and d results in the following coupled linear equations:

$$[A \ C_1 \dots C_J \ d] \begin{pmatrix} \sum_{t=2}^T F(t)P(t-1)F(t)' & \sum_{t=2}^T F(t)x_{t-1}^T u(t) \\ \sum_{t=2}^T x_{t-1}^T u(t)F(t)' & \sum_{t=2}^T u(t)^2 \end{pmatrix} = \begin{bmatrix} \sum_{t=2}^T P_s(t, t-1)F(t)' & \sum_{t=2}^T m_s(t)u(t) \end{bmatrix} \quad (\text{A.27})$$

Where,

$$P(t) = \sum_t^T (x_t^T)(x_t^T)' \quad (\text{A.28})$$

$$m_s(t) = x_t^T(1:M) \quad (\text{A.29})$$

$$F(t) = [I_M v_1(t) I_M \dots v_J(t) I_M]' F \quad (\text{A.30})$$

($m_s(t)$ is first M elements of x_t^T),

$$P(t, t-1) = J_{t-1} \sum_t^T (x_t^T)(x_t^T)' \quad (\text{A.31})$$

$P_s(t, t-1)$ is the first $M \times M$ sub matrix of $P(t, t-1)$.

Estimation of Q

Taking expectations of Equation (A.26) with respect to $p(x(t)|y(1), y(2), \dots, y(T))$ and then differentiating the resulting equation with Q , the estimate of Q is given by

$$\begin{aligned} \hat{Q} = & \frac{1}{T-1} \sum_{t=2}^T (P_s(t) \\ & - P_s(t, \\ & t-1)F'G(t)' - m_s(t)u(t)'d \\ & - G(t)FP_s(t, \\ & t-1) \\ & + G(t)FP(t \\ & - 1)F'G(t)' \\ & + G(t)F x_{t-1}^T u^{(t)}d \\ & - du(t)m_s'(t) \\ & + du(t)(x_{t-1}^T)'F'G(t)' \\ & + d'u(t)^2d) \end{aligned} \quad (\text{A.32})$$

Where,

$P_s(t)$ is the first $M \times M$ sub matrix of $P(t)$. Note that the estimated values of A ($\hat{A}$), C ($\hat{C}$) and d ($\hat{d}$) obtained by solving equation (34) are used in equation (35) in place of A , C and d .

Estimation of B

Each row vector b_m , $m = 1, 2, \dots, M$ can be estimated (we assume the noise covariance matrix R to be diagonal) by maximizing the conditional expectation of the complete log-likelihood given in equation (A.25). By taking the derivative of the conditional expectation and equating it to zero, the estimate of b_m is given by

$$\widehat{b}_m' = (\Phi \sum_{t=1}^T P_m(t) \Phi')^{-1} \Phi \sum_{t=1}^T y_m(t) x_t^T(m) \quad (\text{A.33})$$

Where, $P_m(t) = E(s_m(t)s_m'(t)|y(1), y(2) \dots y(T))$ that can be easily found from $P(t)$. $y_m(t)$ and $x_t^T(m)$ are m -th elements of the vectors $y(t)$ and x_t^T respectively.

Estimation of R

The diagonal observation covariance matrix R can be estimated by maximizing the conditional expectation of the complete log-likelihood given in equation (A.25). The estimation of diagonal components of R are given by

$$\widehat{R}(m, m) = \frac{1}{T} \sum_{t=1}^T (y_m^2(t) - 2\widehat{b}_m' \Phi y_m(t) x_t^T(m) + \text{trace}(\widehat{b}_m' \Phi P_r(t) \Phi \widehat{b}_m)), m=1, 2, \dots M \quad (\text{A.34})$$

Estimation of μ_o and Σ_o

The maximum likelihood estimates of initial state mean μ_o and covariance Σ_o are given by

$$\widehat{\mu}_o = x_1^T \quad (\text{A.35})$$

$$\widehat{\Sigma}_o = \sum_1^T \quad (\text{A.36})$$

The above E and M steps are repeated until the change in log-likelihood of the data between two iterations is below a specified threshold.

Appendix – B

Solving MDS using VB framework

In VB, the goal is to find the posterior distributions of latent variables $q_S(S|Y)$ and parameters $q_\theta(\theta|Y)$ by maximizing the lower bound on the log evidence $L(q)$ given in equation (5).

VB-E-step

In this step, the posterior distributions of latent variables $q_S(S|Y)$ are estimated given the current posterior probability of the model parameters $q_\theta(\theta|Y)$. As in MLE approach, we compute the posteriors of embedded latent signals $x(t)$ from which the posterior of $s(t)$ can be obtained. The distribution over these latent variables is obtained using a sequential algorithm similar to Kalman smoothing which we used in the E-step of MLE approach. In the VB version of Kalman smoothing, the point estimates of the parameters are replaced by their expectations of the type $E(ZWZ')$ where Z is some parameter of the model and W a matrix. Although, these expectations are straightforward to compute, but are computationally expensive for higher order models. We, therefore, use the approximation $E(AWA') = E(A)WE(A')$, which gives qualitatively similar results and is computationally

efficient. This approach was also taken in (Cassidy and Penny, 2002). As a result, the VB-E step is same as the E-step in MLE approach. Therefore, mean and covariance of $\mathbf{x}(t)$ are

given by $\mathbf{x}_t^T$ and $\sum_t^T$.

VB-M step

In this step, the posterior distributions of model parameters $q_{\theta}(\theta|Y)$ are estimated given the current posterior probability of the latent variables $q_s(S|Y)$. Using the probabilistic graphical model in Figure 1, one can show that the joint posterior distribution of parameters $q_{\theta}(\theta|Y)$ further factorizes as

$$q_{\theta}(\theta|Y) = q(\mathbf{A}, \mathbf{C}_1, \dots, \mathbf{C}_J, \mathbf{D}, \mathbf{Q}) q(\mathbf{B}, \mathbf{R}) \quad (\text{B.1})$$

In this work, we also assume that the state and observation noise covariance matrices ($\mathbf{Q}$ & $\mathbf{R}$) to be diagonal. Therefore, the distribution of the elements in the rows of $\mathbf{A}$, $\mathbf{C}_1, \dots, \mathbf{C}_J$, $\mathbf{D}$ & $\mathbf{B}$ can be inferred separately. Consider the state equation for the m-th node

$$\mathbf{s}_m(t) = (\mathbf{a}_m + \sum_{j=1}^J v_j(t) \mathbf{c}_{j,m}) \mathbf{s}_m(t-1) + \mathbf{d}_m u(t) + \mathbf{w}_m(t), \quad \mathbf{w}_m(t) \sim N(0, \beta_m) \quad (\text{B.2})$$

where, $\mathbf{a}_m$ and $\mathbf{c}_{j,m}$ are m-th rows of $\mathbf{A}$ and $\mathbf{C}_j$ respectively and $\beta_m = \frac{1}{Q_{(m,m)}}$. In terms of embedded signal $\mathbf{x}(t)$, the above equation can be written as:

$$\mathbf{s}_m(t) = \theta'_m [F(t) \mathbf{x}(t); u(t)] + \mathbf{w}_m(t) \quad (\text{B.3})$$

Where, $\theta'_m = [\mathbf{a}_m, \mathbf{c}_{1,m}, \dots, \mathbf{c}_{J,m}, \mathbf{d}_m]$ $F(t) = [I_M v_1(t) I_M \dots v_J(t) I_M]^T F$

We assume the following Gaussian-Gamma conjugate priors for θ_m and β_m

$$p(\theta_m, \beta_m | \alpha) \sim N(0, (\beta_m \mathbf{A}_{\alpha})^{-1}) Ga(a_{\alpha}, b_{\alpha}) \quad (\text{B.4})$$

Where, $\mathbf{a}_{gr} = [\alpha_1, \alpha_2, \dots, \alpha_{2M+1}]$ are the hyperpriors on each element of θ_m and $\mathbf{A}_{\alpha} = \text{diag}(\alpha)$

Let the prior on α be

$$p(\alpha) = \prod_{i=1}^{2M+1} Ga(c_{\alpha}, d_{\alpha}) \quad (\text{B.5})$$

Therefore, by applying equation (9), the joint posterior for θ_m & β_m is given by

$$q(\theta_m, \beta_m | Y) = N(\bar{\theta}_m, \bar{\beta}_m^{-1} \sum_m) Ga(a_{m,N}, b_{m,N}) \quad (\text{B.6})$$

Where,

$$\sum_m^{-1} = \begin{pmatrix} \sum_{t=2}^T F(t)P(t-1)F(t)' & \sum_{t=2}^T F(t)x_{t-1}^T u(t) \\ \sum_{t=2}^T x_{t-1}^T u(t)F(t)' & \sum_{t=2}^T u(t)^2 \end{pmatrix} \quad (\text{B.7})$$

$$\bar{\theta}_m = \sum_m \left(\frac{\sum_{t=2}^T F(t)E(s_m(t)x(t-1))}{\sum_{t=2}^T P_s(t, t-1)F(t)'} \right) \quad (\text{B.8})$$

$$a_{m,N} = a_o + \frac{T+2M}{2} \quad (\text{B.9})$$

$$b_{m,N} = b_o + 0.5 \left(\sum_{t=2}^T E(s_m^2(t)) - \bar{\theta}_m' \sum_m^{-1} \bar{\theta}_m \right) \quad (\text{B.10})$$

The posterior for hyper parameters α is given by

$$q(\alpha|Y) = \prod_{i=1}^{2M+1} Ga(c_N, d_{Ni}) \quad (\text{B.11})$$

Where,

$$c_N = c_o + \frac{1}{2} \quad (\text{B.12})$$

$$d_{Ni} = d_o + \frac{1}{2} \left(\bar{\theta}_m^2(i) \frac{a_{m,N}}{b_{m,N}} + \sum_m (i, i) \right) \quad (\text{B.13})$$

The posteriors for θ_m , β_m and α are estimated for each $m = 1, 2, \dots, M$ from which the posteriors for A , C_1 , $\dots$, C_J , D & Q are computed.

Similarly, the posterior distribution for the model parameters in the output equation is computed. Since R is assumed to be diagonal, the observation equation in m -th node is given by

$$y_m(t) = b_m \Phi x_m(t) + e_m(t), \quad e_m(t) \sim N(0, \lambda_m) \quad (\text{B.14})$$

Where, $\lambda_m = \frac{1}{R_{(m,m)}}$. Again assuming Gaussian-Gamma conjugate priors for b_m and λ_m

$$p(\mathbf{b}_m, \lambda_m | \alpha) \sim N(0, (\lambda_m \mathbf{A}_\alpha)^{-1}) Ga(a_o, b_o) \quad (\text{B.15})$$

Where, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_P]$ are the hyper priors on each element of and $\mathbf{b}_m$ and $\mathbf{A}_\alpha = \text{diag}(\alpha)$

Let the prior on α be

$$p(\alpha) = \prod_{i=1}^P Ga(a_o, b_o) \quad (\text{B.16})$$

By applying equation (9), the joint posterior for $\mathbf{b}_m$ and λ_m is given by

$$q(\mathbf{b}_m, \lambda_m | Y) = N(\bar{\mathbf{b}}_m, \lambda_m^{-1} V_m) Ga(a_{m,N}, b_{m,N}) \quad (\text{B.17})$$

$$V_m^{-1} = \Phi \sum_{t=1}^T P_m(t) \Phi' + \bar{\mathbf{A}}_\alpha \quad (\text{B.18})$$

$$\bar{\mathbf{b}}_m = V_m \Phi \sum_{t=1}^T y_m(t) x_t^T(m) \quad (\text{B.19})$$

$$a_{m,N} = a_o + \frac{T+P-1}{2} \quad (\text{B.20})$$

$$b_{m,N} = b_o + 0.5 \left(\sum_{t=2}^T E(y_m^2(t)) - \bar{\mathbf{b}}_m' V_m^{-1} \bar{\mathbf{b}}_m \right) \quad (\text{B.21})$$

$$c_N = c_o + \frac{1}{2} \quad (\text{B.22})$$

$$d_{Ni} = d_o + \frac{1}{2} \left(\frac{1}{\bar{b}_m^2(i)} \frac{a_{m,N}}{b_{m,N}} + V_m(i, i) \right) \quad (\text{B.23})$$

$$\bar{\mathbf{A}}_\alpha(i, i) = \frac{c_N}{d_{Ni}} \quad (\text{B.24})$$

The posteriors for $\mathbf{b}_m$, λ_m and α are estimated for each $m = 1, 2, \dots, M$. In this work, we set the hyperparameters a_o , b_o , c_o , and d_o to 10^{-3} .

The VB-E and VB-M steps are repeated until convergence.

Appendix C

Initialization

The above iterative procedure (E and M step) needs to be initialized. In this work, we estimate the initial values of latent signals($\hat{s}(t)$) by estimating them at each node using the Weiner deconvolution method of (Glover, 1999) wherein the canonical HRF is used for this deconvolution step. We then estimate the initial values of A , C , d and Q by solving Equation (1) by least squares assuming that the $\hat{s}(t)$'s are true values. Similarly the parameters B and R are estimated from Equation 3 by least squares approach. The EM algorithm is then run using these initial values till the required convergence is obtained.

Figure 1.

Probabilistic graphical model for multivariate dynamical system (MDS). All conditional interdependencies in MDS can be inferred from this model. The state variables $s(t)$ are modeled as a linear dynamical system. The non-diagonal elements of matrices A and C represent the intrinsic and modulatory connection strengths respectively. The diagonal elements of D represent the weight of external stimulus at i -th node. $Q(m, m)$ is the state noise variance at m -th node. Each element of A , C and D has precision of α . Each element of α follows Gamma distribution with parameters c_0 and d_0 . The prior for $1/Q(m, m)$ follows Gamma distribution with parameters a_0 and b_0 . $y(t)$ represents the observed BOLD signal, the elements of B represent weights corresponding to the basis functions for HRFs and $R(m,$

m) is the observation noise variance at m -th node. Each element of B has precision of α . Each element of α follows the Gamma distribution with parameters c_0 and d_0 . The prior for $1/R(m, m)$ follows the Gamma distribution with parameters a_0 and b_0 . The random variables are indicated as open circles and deterministic quantities as rectangles.

Figure 2.

Flow chart showing major steps in implementation of MDS. Weiner deconvolution is used to get an initial estimate of latent signals and a least square estimation procedure is used to find an initial estimation of model parameters. The estimates of latent signals and model parameters are refined in E and M steps respectively. These steps are repeated until convergence. The significance of model parameters is then assessed in the inference step.

Figure 3. Simulated models with intrinsic and experimentally induced modulatory connections for (A) 2-node, (B) 3-node and (C) 5-node networks. Intrinsic connections are shown in solid lines and modulatory connections are shown in broken lines and highlighted with the connecting back circles. $A(i, j)$ and $C(i, j)$ are the weights of intrinsic and modulatory connections, respectively, between nodes i and j . $D(i, i)$ is the strength of external stimulus to the i -th node.

Figure 4. Variable regional hemodynamic response function used in the simulations for each node in (A) 2, (B) 3 and (C) 5 node networks shown in Figure 3.

Figure 5. Onset and duration of event related experimental stimuli (A) and modulatory inputs (B) used in the simulations.

Figure 6.

Left panel: Actual and estimated latent signals at each of the nodes of the 5-node network shown in Figure 3C. **Right panel:** Estimated and actual BOLD-fMRI signals at each node. MDS using both MLE and VB approaches, accurately recovered latent signals and predicted the fMRI signals based on the estimated model parameters and latent signals.

Figure 7.

(A) Intrinsic and modulatory connections estimated by MDS using Maximum likelihood estimates (MDS-MLE) and (B) estimates Variational Bayes estimates (MDS-VB). (C) Causal interactions estimated by Granger Causal Analysis (GCA). MDS-VB correctly identified both intrinsic and modulatory connections. MDS-MLE correctly estimated all the intrinsic and modulatory connections in the five node network but also introduced a false modulatory connection from node 3 to node 1. GCA missed both intrinsic and modulatory connections from node 1 to 2 for reasons described in the text.

Figure 8.

(A) Sensitivity, false positive rate (FPR) and Accuracy in identifying causal intrinsic and modulatory connections in the 2-node network (shown in Figure 3A) at SNRs of 0, 5 and 10 dB using MDS-MLE, MDS-VB and GCA. (B) Similar results for 3-node (shown in Figure 3B) and (C) 5-node (shown in Figure 3C) networks. The performance of MDS is superior to GCA for all 3 networks and SNRs. Among MDS methods, the performance of MDS-VB is superior to MDS-MLE. Sample size $T = 500$ time points.

Figure 9.

(A) Sensitivity, FPR and Accuracy in identifying causal intrinsic and modulatory connections in the 2-node network (shown in Figure 3A) at SNRs of 0, 5 and 10 dB using MDS-MLE, MDS-VB and GCA. (B) Similar results for 3-node (shown in Figure 3B) and (C) 5-node (shown in Figure 3C) networks. The performance of MDS is superior to GCA for all 3 networks and SNRs. Among MDS methods, the performance of MDS-VB is superior to MDS-MLE. In contrast to Figure 8, the sample size (T) here is **300** time points.

Figure 10.

(A) Sensitivity, FPR and Accuracy in identifying causal intrinsic and modulatory connections in the 2-node network (shown in Figure 3A) at SNRs of 0, 5 and 10 dB using MDS-MLE, MDS-VB and GCA. (B) Similar results for 3-node (shown in Figure 3B) and (C) 5-node (shown in Figure 3C) networks. The performance of MDS is superior to GCA for all 3 networks and SNRs. Among MDS methods, the performance of MDS-VB is superior to MDS-MLE. In contrast to Figure 8, the sample size (T) here is **200** time points.

Table 1

Mean square error (MSE) between original and estimated latent signals and BOLD-fMRI signals using MDS-MLE and MDS-VB at each node of the 5-node simulated network model (Figure 3C)

Nodes	Neuronal Signals		BOLD Signals	
	MLE	VB	MLE	VB
1	0.024	0.023	0.027	0.027
2	0.024	0.024	0.015	0.014
3	0.019	0.019	0.025	0.024
4	0.017	0.017	0.021	0.02
5	0.018	0.017	0.02	0.02

Table 2

Performance of MDS and GCA in the absence of modulatory effects and external stimuli

Method	Sensitivity	FPR	Accuracy
MDS-VB	0.98	0.02	0.98
MDS-MLE	0.92	0.03	0.96
GCA	0.9	0	0.98