



Identifying effective connectivity parameters in simulated fMRI: a direct comparison of switching linear dynamic system, stochastic dynamic causal, and multivariate autoregressive models

Jason F. Smith^{1*}, Kewei Chen^{2,3,4}, Ajay S. Pillai⁵ and Barry Horwitz¹

¹ Brain Imaging and Modeling Section, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD, USA

² Positron Emission Tomography Center and Banner Alzheimer's Disease Institute, Banner Good Samaritan Medical Center, Phoenix, AZ, USA

³ Department of Mathematics and Statistics, Arizona State University, Tempe, AZ, USA

⁴ Arizona Alzheimer's Disease Consortium, Phoenix, AZ, USA

⁵ Human Motor Control Section, National Institute on Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA

Edited by:

Pedro A. Valdes-Sosa, Cuban Neuroscience Center, USA

Reviewed by:

Anthony S. Dick, Florida International University, USA
Felix Carbonell, Biospective Inc., Canada

*Correspondence:

Jason F. Smith, Brain Imaging and Modeling Section, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, 10 Center Drive, Room 5d39, Bethesda, MD 20894, USA.
e-mail: smithjas@nidcd.nih.gov

The number and variety of connectivity estimation methods is likely to continue to grow over the coming decade. Comparisons between methods are necessary to prune this growth to only the most accurate and robust methods. However, the nature of connectivity is elusive with different methods potentially attempting to identify different aspects of connectivity. Commonalities of connectivity definitions across methods upon which base direct comparisons can be difficult to derive. Here, we explicitly define "effective connectivity" using a common set of observation and state equations that are appropriate for three connectivity methods: dynamic causal modeling (DCM), multivariate autoregressive modeling (MAR), and switching linear dynamic systems for fMRI (sLDSf). In addition while deriving this set, we show how many other popular functional and effective connectivity methods are actually simplifications of these equations. We discuss implications of these connections for the practice of using one method to simulate data for another method. After mathematically connecting the three effective connectivity methods, simulated fMRI data with varying numbers of regions and task conditions is generated from the common equation. This simulated data explicitly contains the type of the connectivity that the three models were intended to identify. Each method is applied to the simulated data sets and the accuracy of parameter identification is analyzed. All methods perform above chance levels at identifying correct connectivity parameters. The sLDSf method was superior in parameter estimation accuracy to both DCM and MAR for all types of comparisons.

Keywords: effective connectivity, fMRI BOLD, modeling and simulation, parameter estimation

INTRODUCTION

There is a growing consensus that cognition can be fruitfully analyzed by studying the dynamic integration of information computed by multiple, spatially discrete, functionally specialized brain regions (McIntosh, 1999, 2001). In studies of cognition utilizing functional magnetic resonance imaging (fMRI), as well as other neuro-imaging methods, application of functional and effective connectivity measures has become dramatically more common. Connectivity methods are used to identify approximations to the integrative dynamics of multiple brain regions at a coarse-grained level of analysis. For fMRI, the spatial scale is measured in millimeters and the temporal scale measured in seconds. Though, these scales are far from that of neurons, dendrites, and spikes, theory, and evidence strongly suggests that the dynamic patterns of interactions at any level of analysis can be analyzed roughly independently from the details of the smaller units that interact to form the patterns themselves (Kelso, 1995; Haken, 1996). Thus, adequate understanding of the dynamic interactions

between brain regions measured at the scale of fMRI may not require complex units such as multi-compartment spiking neurons as a computational substrate. The interregional interactions at this coarser level are potentially interesting in their own right and represent a rich dataset to mine for unique biomarkers of disease and differential abilities (Büchel et al., 1999; Mechelli et al., 2002; Bokde et al., 2009; Whalley et al., 2009; Rowe, 2010; Horwitz and Rowe, 2011).

Increased interest in connectivity analysis of neuro-imaging has resulted in an explosion of the number of different methods for estimating the existence and strength of interregional interactions (Moeller and Strother, 1991; Horwitz et al., 1992; Alexander and Moeller, 1994; Friston, 1994; McIntosh and Gonzalez-Lima, 1994; Bullmore et al., 1996; McIntosh et al., 1996; McKeown et al., 1998; Büchel et al., 1999; Friston et al., 2003; Harrison et al., 2003; Sun et al., 2004; Valdés-Sosa, 2004; Roebroeck et al., 2005, 2011; Seth, 2005; Valdés-Sosa et al., 2005; Marrelec et al., 2006; Patel et al., 2006; Shimizu et al., 2006; Smith et al., 2006,

2010; Rajapakse and Zhou, 2007; Havlicek et al., 2010; Chen et al., 2011). Each connectivity method approximates interregional brain dynamics in a specific manner as determined by the statistical model behind the method. Though the term “connectivity” is consistently used to describe each of these methods, the underlying statistical models are often not consistent. Because of these differences, it is not clear which methods are directly comparable or even attempting to measure the same underlying type of approximate dynamics (Horwitz, 2003). How then should researchers evaluate different connectivity methods and select the method appropriate to their study? If one method identifies a pattern of connectivity in a dataset that another method fails to find does this indicate the first method is superior (cf. David et al., 2008)? If two methods identify a similar pattern of connectivity in the same dataset, does this indicate the connectivity is more likely to exist (cf. Penny et al., 2004)? Answers to these questions will require a deeper understanding of the relations between the mathematical foundations of different connectivity measures and direct comparative evaluations designed with these relations in mind.

The most direct comparison for connectivity methods is relative accuracy. That is, can the method identify interactions between functional units at the correct magnitude? This direct test requires knowledge of the ground truth; the nature, location, and magnitude of the actual interregional interactions. Unfortunately, this ground truth is simply unknown for *in vivo* data for several reasons. To avoid this difficulty, researchers often resort to testing methods using *in silico* simulated data where the ground truth is presumably known (e.g., Bullmore et al., 2001; Calhoun et al., 2002; Roebroeck et al., 2005; Smith et al., 2011a). While the types of simulated data used by researchers are often simplistic (though see Seth et al., 2013), simulations with known ground truth remain necessary to test the validity and effectiveness of connectivity methods.

Recently, an analysis of the accuracy of several disparate connectivity methods was conducted using simulated data (Smith et al., 2011a). Dynamic causal modeling (Friston et al., 2003; DCM) was used, in reverse, to generate simulated fMRI data from known patterns of interregional connectivity and known noisy inputs. The various methods were then applied to the simulated data and their ability to recover the known connectivity patterns was compared. A problem with this study as well as previous comparisons of multiple methods is that the ground truth used for evaluation considered only the location and magnitude of the simulated interregional interactions; the nature of the interactions contained in the simulated data and the nature of the interactions a method was intended to identify were not considered. Given the ambiguity of the concept of connectivity, it is not immediately clear that the connectivity simulated by one method can be reasonably used to comparatively evaluate multiple disparate connectivity methods. The true interregional dynamics of *in vivo* fMRI data are surely complex and different connectivity measures may identify different aspects of these dynamics. In contrast, simulated data is simplistic by design, containing only the type of interregional interactions that were induced by the researcher. If the simulated data does not contain the type of interaction that a method was developed to identify, it is trivial to show

that the method fails to identify the simulated connectivity. What is critical for method evaluation then is to simulate connectivity according to the type intended to be identified by the methods scrutinized. Therefore, only methods with compatible mathematical models of connectivity can be directly compared using the same simulated data.

The current study compares the accuracy of three different but fundamentally related methods, stochastic dynamic causal modeling for fMRI (sDCMf), switching linear dynamic systems for fMRI (sLDSf), and multivariate autoregressive models (MAR) using simulated data. First we present a unifying review of the foundations of several common connectivity methods. The methods are all placed into a similar multi-equation format that separates equations defining connectivity from those defining data observation. The underlying connectivity equation behind the three methods of interest is derived and a single equation is found to be applicable to all three. The form of the output equations for each method is defined and how these output equations may affect simulation accuracy is discussed. For completeness, parameter identification for each method is also briefly discussed. Finally, fMRI data from interconnected regions is simulated using the common equation and each method applied to attempt to recover the known connectivity. We conclude with a discussion of the strengths and failings of each method and suggest future directions for connectivity modeling.

UNIFYING CONNECTIVITY METHODS

First, a unifying discussion of common connectivity methods is presented to formally define what is meant by connectivity at the level of explicit equations. The discussion begins with the relatively simple equations behind so called “functional” connectivity measures. These simple equations are then expanded to form the basis of so called effective connectivity measures. The same equations are expanded again to produce the dynamic models of interest here. For clarity, constants in the equations necessary for centering or scaling have been omitted from the discussion unless necessary.

Assume a researcher is interested in approximating the interactions between signals arising from a set of brain regions of interest as measured by fMRI. Let x_t^i be a scalar variable representing the value of the signal of interest at spatial location i and temporal location t and let y_t^i be its measurement. To keep the discussion general, the specific nature of the signal of interest (e.g., local field potential or spatio-temporally integrated synaptic activity) will not be specified until later. Assume there are n spatial locations (brain regions) from which the signals of interest are measured and for which the connectivity is to be computed. Let x_t (resp. y_t) be a column vector in \mathcal{R}^n containing the signals (resp. measurements) from regions $[1 : n]$ at time t and \mathbf{X} (resp. \mathbf{Y}) be an n by T matrix collecting x_t (resp. y_t) for a set of temporally contiguous time points $[1\tau, 2\tau, 3\tau \dots T\tau]$ equally spaced by some distance τ in time.

RELATING OBSERVATIONS TO SIGNALS

The first step in generating any model is to state an observation equation that relates the signals of interest \mathbf{X} to the observed data \mathbf{Y} . A typical assumption is that the measurements are some

direct function of the signals but corrupted by some noise due to the act of measurement and other unknown factors. A simple, analytically tractable form of this relation would be a linear function with Gaussian errors¹ as shown in Equation 1.

$$y_t = Cx_t + \xi_t, \xi \sim \mathcal{N}(0, R) \quad (1)$$

Here ξ represents the corrupting noise with mean 0 and covariance \mathbf{R} . The observation matrix \mathbf{C} describes the extent to which the underlying signal at each spatial location is present at each measurement location. For some neuro-imaging methods where each measurement represents multiple underlying sources such as MEG, the matrix \mathbf{C} could represent a (simplistic) model of source mixing and bone/scalp conduction. In fMRI, however, the observations are spatially specific. This spatial localization requires that x^i not influence y^j for $i \neq j$; meaning \mathbf{C} must be diagonal. The simplest means to achieve this for the moment is to set \mathbf{C} to an identity matrix that can be ignored as in Equation 2. This simplified linear model will be expanded later in the discussion.

$$y_t = x_t + \xi_t, \xi \sim \mathcal{N}(0, R) \quad (2)$$

DEFINING STATIC FUNCTIONAL CONNECTIVITY

Different equations for the interactions among the signals yield different forms of connectivity. It is only by comparing and contrasting these equations that similarities and differences between various connectivity measures are realized. If a researcher assumes that the signals are independent through time (i.e., there is no function relating the signals at one time point to the signals at any other time point), they may choose from a family of so called static connectivity methods. The signals themselves are not known *a priori* and must be estimated from the data. Thus, the signals themselves are defined in terms of random variables. Assuming these variables are normally distributed and that the interactions among regions are linear results in the model shown in Equation 3.

$$x_t = G\varepsilon_t, \varepsilon \sim \mathcal{N}(0, Q) \quad (3)$$

Here \mathbf{G} contains weights for mixing the as-yet-unknown, non-localized components ε together to produce the localized signals of interest. Given Equations 2 and 3, an explicit form for the distributions of X and Y can be specified as in 4.

$$\begin{aligned} X &\sim \mathcal{N}(0, GQG^T) \\ Y &\sim \mathcal{N}(0, GQG^T + R) \end{aligned} \quad (4)$$

Obviously constraints must be placed on \mathbf{G} , \mathbf{Q} , and \mathbf{R} for the signal \mathbf{X} to be meaningful and for the model to be identified. A degenerate model (from the viewpoint of connectivity) could be obtained by simply setting \mathbf{R} the covariance matrix of the measurements \mathbf{Y} and setting \mathbf{G} to a zero matrix, resulting in a rather useless signal. Without loss of generality, \mathbf{Q} can be set to an identity matrix with the additional information subsumed by \mathbf{G} . This simplification makes the components ε

orthonormal (independent with unit variance). If the observation noise covariance, \mathbf{R} , is restricted to an identity matrix multiplied by an arbitrary scalar, σ , and the limit taken as σ goes to zero (i.e., zero observation noise), Equations 2, 3, 4 form the Karhunen–Loève transform from signal processing also known as Principal Components Analysis (PCA; Jolliffe, 1986; Diamantaras and Kung, 1996; Tipping and Bishop, 1999)². When PCA is used in (functional) connectivity analysis, typically only a few components with the greatest affect on the observed data variance are examined. Commonly used heuristics for selecting the number of components (e.g., 80 or 90% of the data variance; Jolliffe, 1986) are essentially assuming a non-zero bound for σ . If \mathbf{R} is merely restricted to an arbitrary diagonal matrix, Equations 2, 3, and 4 describe the exploratory factor analysis (FA) method. In either case, the signals can be identified by maximizing the posterior probability of \mathbf{x} given \mathbf{y} as in 5 with the appropriate constraints on \mathbf{R} (Roweis and Ghahramani, 1999). Note that as \mathbf{R} approaches zero the covariance

$$p(x|y) \sim \mathcal{N}\left(G^T (GG^T + R)^{-1} y, I - G^T (GG^T + R)^{-1} G\right) \quad (5)$$

of $p(x|y)$ goes to zero and the maximum likelihood solution for \mathbf{X} and \mathbf{G} are obtainable via iterative least squares (Wold et al., 1987; Diamantaras and Kung, 1996; Wold, 1996).

The normality assumption for the signals and measurements can be relaxed by adding a non-linear function to Equation 3 as given in Equation 6.

$$x_t = Gg(\varepsilon_t), \varepsilon \sim \mathcal{N}(0, I) \quad (6)$$

If \mathbf{R} is again restricted to an identity matrix multiplied by an arbitrary scalar, σ , and the limit taken as σ goes to zero, Equations 2 and 6 describe the popular Independent Components Analysis (ICA) model³. If some of the components are assumed to be “noise” their variance is returned to \mathbf{R} after identification of the zero noise system as in the PCA case above. It has been shown (Roweis and Ghahramani, 1999) that if the hyperbolic tangent function is used as the learning non-linearity in identifying the ICA demixing matrix, the non-linear function $g(\varepsilon)$ in Equation 6 is given by Equation 7. Thus, ICA can be seen as both

$$g(\varepsilon) = \ln \left(\tan \left(\frac{\pi}{4} \left(1 + \operatorname{erf} \left(\frac{\varepsilon}{\sqrt{2}} \right) \right) \right) \right) \quad (7)$$

a linear model with non-gaussian error or a non-linear model with Gaussian errors (Roweis and Ghahramani, 1999).

DEFINING STATIC EFFECTIVE CONNECTIVITY

The connectivity contained within the \mathbf{G} matrix of the PCA, FA, and ICA methods presented above are often referred to as

²The limit being taken as σ and thus R go to zero rather than simply setting R to zero is only necessary to preserve a proper probability model for PCA (Tipping and Bishop, 1999).

³This is, however, a considerable abuse of the terminology of ICA. Typically ICA would refer to ε as the underlying signals or sources and, since R is 0, would not distinguish between observations \mathbf{Y} and \mathbf{X} . Despite the odd terminology used the equations are correct.

¹We use the tilde \sim to indicate “distributed as” and $N(a, b)$ to indicate a normal distribution with mean a and covariance (inverse precision) b .

“functional.” The signals x^i and x^j are functionally connected if the elements of the matrix \mathbf{G} at $G^{(i,k)}$ and $G^{(j,k)}$ are both large in magnitude for at least one $k \in [1 : n]$. This would indicate the existence of a common component to the two signals of interest and therefore a non-zero correlation between them. There is, however, no inherent interaction, directionality, or “causality” between the signals in Equations 3 or 6 in their generic form. The signals x_t are functions of the components ε_t , not of each other.

Effective connectivity is typically taken to imply causality or at least directionality to the connectivity (Valdés-Sosa et al., 2011). Returning to the linear model of Equation 3 with $\mathbf{Q} = \mathbf{I}$, if \mathbf{R} is again restricted to a diagonal matrix, and \mathbf{G} is constrained *a priori* to have at most $n(n-1)/2$ non-zero values and be acyclic (satisfied if \mathbf{G} is lower-triangular), Equations 2,3, and 4 describe the confirmatory FA model, also known as Structural Equation Modeling (SEM)⁴. With appropriate priors, 2 through 4 can also describe Linear Gaussian Bayesian Networks (BN, Rajapakse and Zhou, 2007). To see how the lower triangular (or other acyclic) restriction implies directionality in Equation 3, consider the three element model shown in **Figure 1**. The matrix \mathbf{G} is shown in **Figure 1A** implying the functional connectivity (i.e., common components) shown in **Figure 1B**. Ignoring terms with zero coefficients, the restrictions described above result in the following set of equations for the \mathbf{G} matrix of **Figure 1A**.

$$\begin{aligned}x_t^i &= \varepsilon_t^i \\x_t^j &= \alpha \varepsilon_t^i + \varepsilon_t^j \\x_t^k &= \beta \varepsilon_t^i + \gamma \varepsilon_t^j + \varepsilon_t^k\end{aligned}\quad (8)$$

Since the signal x^i is localized in space and x^i and ε^i are equivalent, the component ε^i is localized as well. Thus, we are justified in rewriting the signal in region j , x_t^j , as a function of the region x_t^i and its residual error ε_t^j with α indicating the strength of the connectivity from region i to region j .

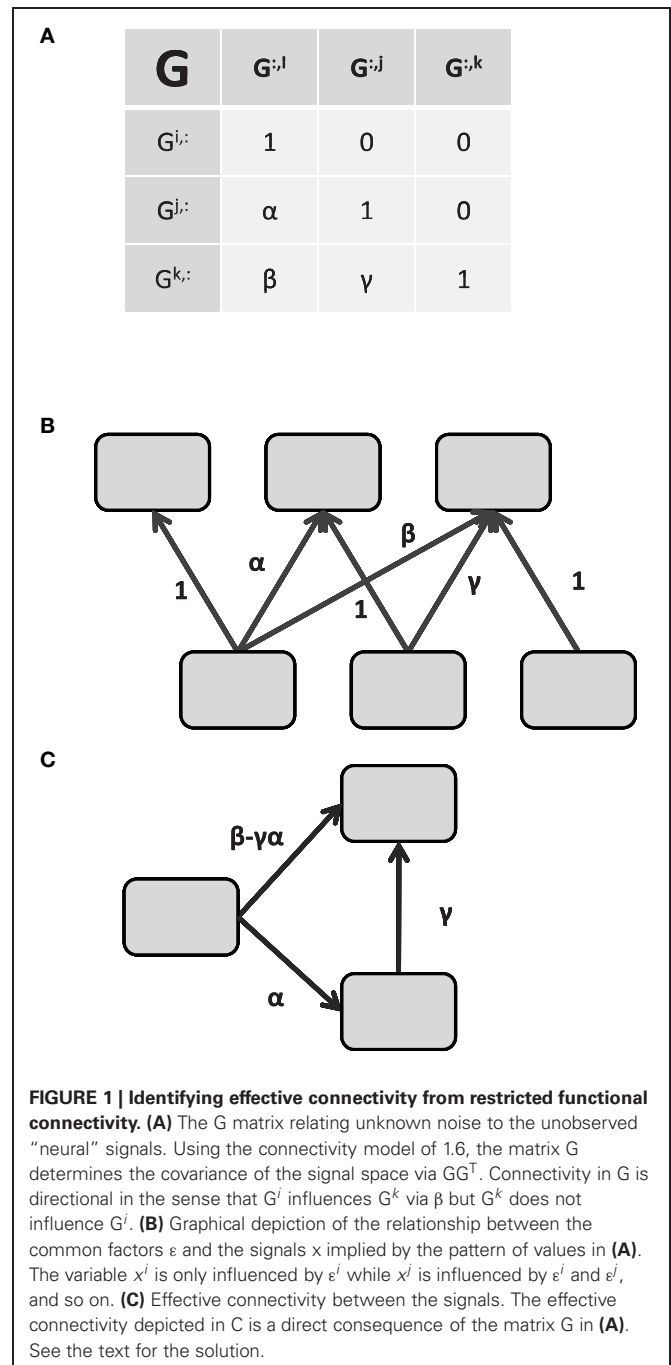
$$x_t^j = \alpha x_t^i + \varepsilon_t^j \quad (9)$$

Similarly, x_t^k may be rewritten as a function of x_t^i and x_t^j .

$$\begin{aligned}x_t^k &= \beta x_t^i + \gamma(x_t^j - \alpha x_t^i) + \varepsilon_t^k \\x_t^k &= (\beta - \gamma\alpha) x_t^i + \gamma x_t^j + \varepsilon_t^k\end{aligned}\quad (10)$$

Therefore, the directional connectivity graph in **Figure 1C** is a direct result of the Equations in 8. Returning to the non-linear model in Equation 6 and again restricting \mathbf{G} to be

⁴These assumptions lead to recursive SEMs (i.e., a SEM that is described by a directed acyclic graph) that are uniquely identifiable via ordinary least squares. Some cyclic (i.e., non-recursive) SEMs can be identified using instrumental variables or other approaches, but identification can be considered only on a case by case basis. Without instrumental variables, often only a set of multiple equivalent non-recursive (cyclic) models can be identified without resorting to stochastic sampling methods (Richardson, 1996; Storkey et al., 2007; Lacerda et al., 2008).



lower-triangular, the functional connectivity of the ICA method becomes the effective connectivity of the Linear Non-Gaussian Acyclic method (LiNGAM, Shimizu et al., 2006) in a manner directly analogous to the relation between exploratory FA and SEM. Thus, the type of connectivity contained in the static effective connectivity methods SEM and LiNGAM is not inherently different from their static functional connectivity counterparts as they rely on the same underlying equations; they are merely restricted subsets of the same connectivity model. In light of this similarity, and to keep consonance with terminology from other fields, we suggest the use of the term “confirmatory connectivity”

and “exploratory connectivity” in place of effective and functional respectively for these static methods.

DEFINING DYNAMIC CONNECTIVITY IN DISCRETE TIME

SEM, BN, and LiNGAM are static connectivity methods in that they require the assumption of no temporal structure to the data (though see e.g., Fisher, 1970 for a dynamic interpretation of SEM). That is, any permutation of the temporal ordering of \mathbf{Y} will produce identical values in \mathbf{G} for these methods. Though a useful fiction, this assumption is untenable when the observations are closely spaced in time as is the case for fMRI. If a linear temporal dependency of x_t on its past values (e.g., $x_{t-\tau}$) is included in Equation 3, the multivariate autoregressive (MAR⁵) model is recovered.

$$x_t = \mathbf{A}x_{t-\tau} + G\varepsilon_t, \varepsilon \sim \mathbb{N}(0, I) \quad (11)$$

Note that the measurement function given by Equation 2 is still being used with Equation 11 for the MAR model⁶. Note as well that there are now two “connectivity” matrices. As before, \mathbf{G} describes the connectivity that is instantaneous relative to the sampling rate, or occurring “within” the sample interval. This connectivity can be thought of as residual static exploratory (functional) connectivity or, if restricted as in SEM, residual static confirmatory (effective) connectivity. In addition, the new matrix \mathbf{A} describes connectivity across time. The connectivity in \mathbf{A} is directional and causal. However, if \mathbf{G} is not strictly diagonal, the elements of \mathbf{A} as identified in 11 must be interpreted with some caution as they do not account for all the interregional interactions (Sims, 1980; Sims and Zha, 1998; Lütkepohl, 2007; Rubio-Ramírez et al., 2010; Chen et al., 2011; Smith et al., 2011b). To identify the unique effect of an externally induced change in one variable on the system described by 11 requires an additional restrictions on the within time connectivity \mathbf{G} as seen before with SEM (Rubio-Ramírez et al., 2010). If \mathbf{G} is invertible structured such that it can be written in lower-triangular form by reordering the variables, the within time connectivity, \mathbf{S}_0 , is given by \mathbf{G}^{-1} while the cross time connectivity, \mathbf{S}_1 , is given by $\mathbf{A}\mathbf{G}^{-1}$ (Sims, 1980; Rubio-Ramírez et al., 2010). While most MAR modeling for fMRI has been conducted using the framework of Granger Causality⁷ and has considered only the through time connectivity in the \mathbf{A} matrix (cf. Seth, 2005) analysis of both \mathbf{A} and \mathbf{G} may be more useful to better understand connectivity at the relatively low sampling rate of fMRI (Chen et al., 2011; Smith et al., 2011b).

For completeness Equation 12 gives a generic form of the MAR connectivity equation that includes external inputs, v_t ,

and possible temporal variability for the connectivity and noise covariance. The t subscript on the matrices indicates that these values may change with time

$$x_t = A_t x_{t-\tau} + B_t v_t + G_t \varepsilon_t, \varepsilon \sim \mathbb{N}(0, I) \quad (12)$$

(e.g., with changing task condition). This generic connectivity equation contains stationary SEM ($A_t = 0$, $B_t = \mathbf{B}$, $G_t = \mathbf{G}$) and stationary MAR ($A_t = \mathbf{A}$, $B_t = \mathbf{B}$, $G_t = \mathbf{G}$) as special limited cases. Obviously, the MAR connectivity equation should not be used to simulate data for SEM with a non-zero \mathbf{A} matrix (see Fisher, 1970 for a possible exception to this rule).

The full temporal variability in 12 is difficult to solve; identifying a unique full \mathbf{A} matrix for each time-point is a severely under-determined problem. Methods for identifying \mathbf{X} and each unique \mathbf{A}_t do exist; for example by assuming a model for \mathbf{A}_t given $\mathbf{A}_{t-\tau}$ (Wan and van der Merwe, 2001; Havlicek et al., 2011; Smith et al., 2011b). However, typically \mathbf{A}_t is more constrained. If the matrices in 12 are assumed constant over some interval and interval boundaries are known, standard methods for identifying MAR model parameters may be used within the intervals. Bilinear models, which assume \mathbf{A}_t is a function of a constant matrix \mathbf{A} altered by a known external variable are another simple method to identify a constrained \mathbf{A}_t (cf. Penny et al., 2005; Makni et al., 2007; Ryalı et al., 2011). The bilinear model is easily presented in the form of 12 by setting A_t to $A + \sum_j H_j u_t^j$ where \mathbf{H} is the (non-temporally varying) bilinear term and u_t is the known input influencing this term (Smith et al., 2010)⁸. If u_t is simply an indexing variable, the result is similar to the constant interval method. However, the bilinear model does not account for non-stationarity in the error and other inputs, holding \mathbf{G}_t and \mathbf{B}_t to constant matrices. If a small number of distinct \mathbf{A} , \mathbf{B} , and \mathbf{G} matrices are assumed to exist (e.g., $A^i, B^i, G^i, i \in [1, 2, \dots, p]$) and the probability of their application at a time t is conditioned on an additional variable s with the same temporal index (e.g., $p(A_t = A^i | s_t = s^i)$), the switching linear dynamic system (sLDSf) model is recovered (Smith et al., 2010). If the index variable s is known, the parameters of the sLDSf model can be recovered in a similar manner as in the constant interval case. If the index variable s is unknown, the problem is more difficult, but solvable (Kim, 1994; Murphy, 1998; Barber, 2006; Fox et al., 2010). The connectivity of sLDSf can be as general as Equation 12 assuming the limited set of $\mathbf{A}^i, \mathbf{B}^i$, and \mathbf{G}^i matrices in the sLDSf model form a basis spanning the space of possible $\mathbf{A}_t, \mathbf{B}_t$, and \mathbf{G}_t matrices such that any \mathbf{A}_t matrix (resp. $\mathbf{B}_t, \mathbf{G}_t$) can be represented as a weighted combination of the given \mathbf{A}^i (resp. $\mathbf{B}^i, \mathbf{G}^i$) matrices (Smith et al., 2011b).

DEFINING DYNAMIC CONNECTIVITY IN CONTINUOUS TIME

The MAR and sLDSf methods are defined in discrete time. That is, the connectivity equations describe the evolution of values of the signals x_t over a fixed set of time-points $t \in [1\tau 2\tau 3\tau \dots]$. These equations can be transformed, without error, such that they describe the differences between temporally contiguous values of

⁵Also referred to as vector autoregressive (VAR) and, if instantaneous relationships between the signals are considered, structural vector autoregressive (SVAR) models.

⁶The \mathbf{R} matrix could also be set to zero to ensure a pure MAR model; otherwise 1.2 and 1.10 describe a similar but more general state space model (Lütkepohl, 2007). Any MAR model can be described in state space form, but not all state space models are MAR models.

⁷Strictly speaking a MAR analysis is not identical to a Granger Causality analysis though the terms are often used interchangeably. See REFS for more details.

⁸Most descriptions of bilinear models for connectivity use \mathbf{B} to indicate the bilinear term that is labeled \mathbf{H} here.

the signals x_t (i.e., $x_{nt} - x_{(n-1)\tau}$). This is the so called “delta” form of the MAR model and is shown in Equation 13.

$$(x_{t+\tau} - x_t)/\tau = \tilde{A}_t x_t + \tilde{B}_t v_t + \tilde{G}_t \varepsilon_t \quad (13)$$

For the technically minded, the conversion from 12 to 13 is given in **Appendix B**. The left-hand side of Equation 13 can be seen as an Euler approximation to a derivative, thus as the sampling rate τ approaches 0, Equation 13 approaches the derivative of x_t as in 14.

$$\dot{x}_t = \tilde{A}_t x_t + \tilde{B}_t v_t + \dot{\eta}_t \quad (14)$$

Because it is expressed in terms of a derivative, if the value of x is known at some time point (e.g., x_0), 14 can be used to identify a value of x_t for any value of $t > 0$. Thus, 14 is referred to as continuous time. The continuous time matrices in 14 have been marked with the tilde to distinguish them from the discrete time versions in 12. The vector $\dot{\eta}_t$ is a white noise disturbance which must technically have infinite variance and its full treatment is beyond the scope of the current discussion. Because of this, Equation 14 is more often expressed via a stochastic integral of the integrated version of $\dot{\eta}_t$ as

$$dx = \tilde{A}_t x_t dt + \tilde{B}_t v_t dt + d\eta_t, E[[d\eta]] E[[d\eta]]^T = \tilde{Q} dt \quad (15)$$

in 15 where $E[x]$ indicates the expected value of x and $\tilde{Q} dt$ can be considered an incremental covariance matrix. If the noise process in 14 is removed, the result is a deterministic differential equation similar to that used to describe connectivity in the original deterministic Dynamic Causal Model (dDCM, Friston et al., 2003). All that is needed to recover dDCM from 14 with $\dot{\eta}_t = 0$ is to replace \tilde{A}_t with $\tilde{A} + \sum_j \tilde{H}_j u_t^j$ to form the bilinear model as done in the previous section. The bilinear continuous time neural connectivity of sDCMf can be recovered from 15 in a similar manner and is given in Equation 16.

$$\begin{aligned} dx &= \left(\tilde{A} + \sum_j \tilde{H}_j u_t^j \right) x_t dt + \tilde{B} v_t dt + d\eta_t, E[[d\eta]] E[[d\eta]]^T \\ &= \tilde{Q} dt \end{aligned} \quad (16)$$

The letters used as matrix labels in 16 have been changed from those of Friston et al. (2003) to remain consistent with the equations above.

The relation between continuous time connectivity expressed in 16 and the discrete time connectivity of 12 has been well studied (cf. Åström, 1970; Bar-Shalom et al., 2001; Ljung and Wills, 2010). Assuming the system matrices (\tilde{A} , \tilde{B} , and \tilde{Q}) and the input v_t are constant over the interval $[t - \tau : t]$ then the exact system matrices for an equivalent discrete time system can be derived from the continuous time versions as follows:

$$\begin{aligned} A_t &= e^{\tau \tilde{A}_t} \\ B_t &= \tilde{A}_t^{-1} \left(e^{\tau \tilde{A}_t} - I \right) \tilde{B}_t \end{aligned} \quad (17)$$

where e^A indicates the matrix exponential of A and A^{-1} indicates the inverse of A . The matrix B_t can still be computed from \tilde{B}_t

even if \tilde{A}_t is singular. The covariance of the stochastic input for the discrete time system can also be derived from the stochastic continuous time system via the Equations in 18.

$$\begin{aligned} \varepsilon_t &= \int_{s=t-\tau}^{s=t} e^{\tilde{A}_t(t-s)} d\eta_s \\ Q &= \int_{s=0}^{s=\tau} e^{\tilde{A}_s} \tilde{Q} e^{\tilde{A}_s^T} ds \end{aligned} \quad (18)$$

Thus, the corresponding discrete time Q is the solution to the continuous Lyapunov equation:

$$\tilde{A}_t Q + Q \tilde{A}_t^T + \tilde{Q} - e^{\tau \tilde{A}_t} \tilde{Q} e^{\tau \tilde{A}_t^T} = 0 \quad (19)$$

To reduce this Q to an identity matrix as assumed in 12 and recover G , recall that if a matrix W is the Cholesky factor of a symmetric matrix Z then:

$$\begin{aligned} W &= chol(Z) \\ Z &= W W^T \end{aligned} \quad (20)$$

Examining Equations 3 and 4 we see that if the covariance of ε_t is Q , then the covariance of $G\varepsilon_t$ is GQG^T . Thus, the matrix G from 12, assuming identity covariance of ε , is derived from the matrix \tilde{Q} in 15 by solving for Q in 19 and then taking the Cholesky factor of the resulting matrix⁹.

It is important to stress that the mapping from continuous time to discrete time is subjective. That is, assuming an appropriate sampling interval, there is one and only one discrete time system that corresponds to a given continuous time system. The matrix exponential function in Equation 17 is unique as is the solution to 19¹⁰. Assuming the continuous time matrices contain no frequencies higher than the half the sampling rate, the converse is also true. The inverse of the matrix exponential, the matrix log, may or may not have a unique solution (Culver, 1966), but other methods exist for discrete-to-continuous conversion (Raol et al., 1987; Åström and Wittenmark, 1996; Franklin et al., 1997). Unfortunately, if the sampling period is insufficient, there will be many continuous time systems that could correspond to a given discrete time system.

To summarize, Equation 15 and its equivalent discrete time counterpart in Equation 12, as derived by Equations 17 through 20, represent generic dynamic effective connectivity equations that contain dDCM, sDCMf, sLDSf, MAR, SEM, FA, and PCA as special cases. However, not all of the special cases are compatible. Assuming appropriate care is taken with scaling, definition, and conversion of the system matrices, sDCMf, sLDSf, and MAR can all share the single connectivity Equation 12. However, dDCM is not compatible with sLDSf, MAR or any of the static connectivity

⁹If there are known zeros in G , it can be solved via maximum likelihood methods rather than Cholesky factors.

¹⁰While the discrete time error covariance matrix Q is unique, the representation of the same system using the G matrix may not be. There may be additional G matrices for which $GIG^T = Q$. This is not a concern if a method solves for Q rather than G .

measures. The dDCM form of 15 assumes \tilde{Q}_t (and thus G_t) is zero and all connectivity is described by $\tilde{A} + \sum_j \tilde{H}_j u_t^j$ while the static connectivity measures such as SEM assume A_t (and thus \tilde{A}_t) is zero and all connectivity is described by G_t . Therefore, using dDCM to generate simulated fMRI data for testing static connectivity methods (e.g., SEM, FA, PCA, ICA, correlations, partial correlations, etc.), because of its lack of compatibility with these measures, is inappropriate.

DIVERGING OBSERVATION MODELS

The discussion above focused on identifying the connectivity equations; the observation equation was ignored beyond the assumption that y^i was dependent only on x^i and not x^j for $i \neq j$ and assuming various restrictions on the observation noise covariance \mathbf{R} . Though 12 represent the common connectivity equation for sDCMf, sLDSf, and MAR, the methods do not share common observation equations. The observation equations for each of the three methods are discussed below in turn.

MAR

The MAR observation equation is the simplest of the three methods discussed here. The observation equation assumed by a MAR analysis of fMRI data is essentially Equation 2. However, 2 can be further simplified in the MAR case by setting ξ and \mathbf{R} to 0 and placing any observation noise variance into the variance of ϵ in Equation 12 and scaling the system matrices accordingly. This would make \mathbf{Q} an arbitrary diagonal matrix. It could be argued that this simplistic observation equation is a poor model for mapping from predominantly neurally based variables (e.g., local field potentials) to the hemodynamic variable observed via fMRI (Smith et al., 2011a). The BOLD signal obtained in fMRI is an epiphenomenon arising from neural, glial, and vascular sources (Logothetis et al., 2001; Logothetis and Pfeuffer, 2004; Logothetis, 2008). It has been assumed that BOLD signals in one region do not causally influence BOLD signals in other regions. This would seem to invalidate the MAR observation equation since the signals, \mathbf{X} , would be defined and interact in “hemodynamic” space. However, the BOLD response can be accurately modeled as a causal low pass filter (Henson, 2003). While the exact shape of this filter may vary from subject to subject and from region to region within subjects, there is a common generic shape of the BOLD filter (Aguirre et al., 1998; Handwerker et al., 2004, 2012). The impulse response of this filter can be approximated by a Poisson function or mixture of two Beta functions (Friston et al., 1994). To the extent that the transfer function of the system described by Equation 12 contains no frequencies higher than those passed by the BOLD filter, passage of \mathbf{X} through this filter would result in a relatively undistorted temporal shift of the true signal (Barnet and Seth, 2011; Seth et al., 2013). Unfortunately, a change from one set of system matrices to another in the connectivity equation would likely induce a high frequency component in \mathbf{X} . This suggests that MAR is likely better suited to the so-called resting state fMRI or slow block design experiments than to event related or other rapidly changing designs. However, the output equation of MAR requires no additional parameters to be estimated and the MAR model can be identified in closed form.

sLDSF

The observation equations used in sLDSf are based on the linear convolution model of Penny et al. (2005). The sLDSf observation equations are presented in Equations 21 and 22.

$$y_t = \beta \phi z_t + D v_t + \zeta_t, \zeta \sim \mathbb{N}(0, R) \quad (21)$$

$$z_t = [x_t, x_{t-\tau}, x_{t-2\tau}, x_{t-3\tau}, \dots, x_{t-(h-1)\tau}] \quad (22)$$

The variable z_t contains h errorless lagged copies of the signals x from $x_{t-(h-1)\tau}$ to x_t . The observation, y_t , is an instantaneous linear function of z_t any additional observation level input v_t (e.g., movement related variables) and noise ζ_t with a diagonal covariance matrix $\mathbf{R}^{ij} = 0$ for $i \neq j$. The matrix Φ is an *a priori* known set of basis vectors that span the likely variability in the hemodynamic impulse response function (hIRF) such as a canonical hemodynamic response and its derivatives with respect to time and dispersion (Penny et al., 2005). The matrix β contains regionally specific weights for these bases to generate a unique hIRF $\beta^i \Phi$. The linear output $\beta^i \Phi z_t^i$ is thus equivalent to convolving each signal with a regionally specific hemodynamic response (Smith et al., 2010). The sLDSf output equations with three basis vectors require the estimation of three additional parameters per region.

sDCMF

The output equations for sDCMf are the most complex of the three methods. DCMf uses a non-linear dynamic system to model the biophysical states that engender the BOLD signal (Friston et al., 2003). The DCMf output equations embody the Balloon–Windkessel model (Buxton et al., 1998; Mandeville et al., 1999; Friston et al., 2003; Riera et al., 2004; Buxton, 2012). The regionally localized signal x_i causes an increase in an equally localized vasodilatory signal. Inflow responds to this signal and affects blood volume as well as feeding back to the vasodilatory signal. Change in deoxyhemoglobin content is a function of flow, volume, oxygen extraction as well as its past. This complex interaction can be summarized with Equations 23 through 26 where s is the vasodilatory signal, f is inflow, v is blood volume, q is:

$$\frac{ds}{dt} = x - \kappa s - \gamma(f - 1) \quad (23)$$

$$\frac{df}{dt} = s \quad (24)$$

$$\tau \frac{dv}{dt} = f - v^{1/\alpha} \quad (25)$$

$$\tau \frac{dq}{dt} = f(1 - (1 - \rho)^{1/f} / \rho - v^{1/\alpha} / v) \quad (26)$$

deoxyhemoglobin content, alpha is Grubb’s exponent and rho is resting oxygen extraction fraction. The observed BOLD signal is then taken to be a static non-linear function of volume and deoxyhemoglobin with additional biophysical parameters as in Equation 27. In all, the output

$$y = 0.02(9\rho - \rho^2 + 1.8 - 2q/v - 2\rho v - 0.2v) \quad (27)$$

equations of sDCMf require five additional parameters to be estimated per region.

PARAMETER ESTIMATION

The parameter estimation methods of MAR, sLDSf, and sDCMf are briefly reviewed. For more detailed descriptions see (Murphy, 1998; Neumaier and Schneider, 2001; Lütkepohl, 2007; Daunizeau et al., 2009; Friston et al., 2010; Li et al., 2011; Smith et al., 2010).

MAR

Coefficients of MAR models can be obtained through linear regression of past on future values using least-squares methods (Lütkepohl, 2007). To avoid numerical issues related to forming and inverting the covariance of the predictor matrix, we implemented the regression via QR factorization (Stewart, 1987). For each simulation, all time points within a condition were selected and the first four observations per block were discarded. Inputs were shifted forward in time by 4 s to accommodate the delay in the peak hemodynamic response¹¹. The inputs and observations were combined to form a single design matrix and used to predict the observations 1 s after the times included.

sLDSf

The parameters of the sLDSf model are identified using an iterative Bayesian Expectation Maximization (EM, Dempster et al., 1977) algorithm written by the first author in a combination of Matlab and ANSI C. In brief, starting estimates of the system matrices based on a MAR model are used to estimate the signals given the observed data using Kalman filtering/smoothing (Bar-Shalom et al., 2001; Haykin, 2002; Lütkepohl, 2007; Smith et al., 2010). The likelihood of the data given the model is also calculated. This is the expectation step of the EM algorithm. The likelihood is maximized given these signals by updating the system matrices. The matrices are updated by setting the derivatives of the likelihood function with respect to each of the system matrices to zero and solving (Shumway and Stoffer, 1982; Ghahramani and Hinton, 1996; Murphy, 1998; Penny et al., 2005). These updated system matrices are then used to estimate new signals. This iterative process is repeated until the change in likelihood at each step drops below a specified threshold.

The likelihood of the data given the model was modified to include priors on the system matrices (cf. Makni et al., 2007; Ryali et al., 2011). The parameter identification method implemented here used the Empirical Bayes framework (Bishop, 2006). A multivariate normal-inverse gamma prior distribution was used for β and the diagonal elements of \mathbf{R} (Bishop, 2006). The prior mean of β was set to [1,0,0] at each region for the canonical hIRF, its temporal derivative, and its dispersion derivative, respectively. The necessary hyperparameters were estimated from the data (Bishop, 2006). A matrix-normal-Wishart density was used for the conjugate prior for \mathbf{A} , \mathbf{B} , and \mathbf{Q} given \mathbf{X} and \mathbf{u} (Minka, 2000) and shown

in 28. The prior mean of \mathbf{A} was set to the

$$p\left([A, B] \begin{bmatrix} X \\ u \end{bmatrix}\right) \sim \mathbb{N}W^{-1} \\ \times \left(\begin{bmatrix} 0.5 \\ TR \end{bmatrix} I, 0, \sum_j \alpha_j \begin{bmatrix} X \\ u \end{bmatrix} S_j \begin{bmatrix} X \\ u \end{bmatrix}^T, zI, z \right) \quad (28)$$

identity matrix multiplied by 0.5/TR and the prior mean of \mathbf{B} was set to zero. Two separate variances were estimated for the values in \mathbf{A} and \mathbf{B} ; one for the elements of \mathbf{A} and one for the elements of \mathbf{B} . The matrix S_j in 28 is an appropriate selection matrix to allow for the separate variances and z is the estimated degrees of freedom. The necessary hyperparameters were estimated from the data (Minka, 2000).

sDCMf

Parameter identification for sDCMf was performed with SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/4750/2012-05-24/14:05:41Z>). This version uses the generalized filtering method described by Friston et al. (2010) and Li et al. (2011). Parameter identification involves approximating the conditional density on the unknown parameters and signals and maximizing an approximate bound on the log evidence using a variational Bayesian method. The function `spm_dcm_estimate` included in SPM12 was called via a separate script with the following parameter settings: `stochastic = 1`; `center = 1`; `endogenous = 0`; `delays = 0`; `dt = 1` s. In addition, if for a given simulation any element of the simulated \mathbf{A} matrix was zero for all conditions in that simulation, sDCMf, was informed to set this element to zero.

IMPLICATIONS

The commonality of the connectivity equations for sDCMf, sLDSf, and MAR suggest the following steps to simulate *in silico* interregional effective connectivity for these methods. First, the continuous time matrices $\hat{A}_t \hat{B}_t \hat{Q}_t$ and the vector v_t are defined in the simulation for all possible values of t . Second, a small step size, τ , is chosen such that $\hat{A}_t \hat{B}_t \hat{Q}_t$ and v_t are all constant on the interval $t-\tau$ to t . The discrete time matrices A_t , B_t , and Q_t and/or G_t for step size τ are identified using 17 and solving 19. The signal data, \mathbf{X} , is then simulated using these discrete time matrices at a step size τ with suitable random noise passed through G . We stress that this is not equivalent to integrating a deterministic version of 15 using the Euler or Runge–Kutta method with a temporal step of τ and then adding noise with covariance \tilde{Q} after each step. The matrix \tilde{Q} is not the appropriate noise covariance for such discrete time steps because the noise also needs to be integrated over this period (Ljung and Wills, 2010). Simulated fMRI data can then be generated using an appropriate function of the simulated signal that is consistent with real fMRI data but is not the actual output function of any of these methods. Parameters of sDCMf, sLDSf, and MAR models are identified for the simulated fMRI data. The true system matrices and the identified sDCMf system matrices are then converted, without error, to a step size common to the sLDSf and MAR (typically the simulated TR). Once all system matrices are expressed in a common format they can be directly compared.

¹¹The MAR regressions were also attempted using the inputs at the true time convolved with a canonical hIRF. This resulted in poorer performance as did deconvolution of the observations using a canonical hIRF (division in frequency space followed by a low-pass filter). Results of these inferior methods are not reported here.

To produce simulated fMRI data from the signal data created as above, one could apply method specific output equations to generate method specific data. However, because the MAR output equation does explicitly not model the BOLD response this would not be a realistic test of the utility of MAR for fMRI. To avoid bias, a different output model was chosen for the simulations. The mixture of two betas model implemented in the function `spm_hrf()` distributed with SPM8 ([www.fil.ion.ucl.ac.uk/spm/](#)) was used to create hIRFs that would then be convolved with the signals. The `spm_hrf()` function essentially forms two time series from beta probability distributions and subtracts the second from the first. The purpose of the second is to model the post-stimulus undershoot seen in BOLD data. The function has eight parameters: step time, delay of response peak from onset, delay of undershoot peak from onset, dispersion of response, dispersion of undershoot, ratio of response to undershoot, onset length, and length of kernel. To create variable hIRFs, three parameters to the `spm_hrf()` function were allowed to vary at random from typical values; delay of response peak from onset, delay of undershoot peak from onset, and ratio of response to undershoot.

The resulting variable hIRFs, while distinct from the exact sDCMf or sLDSf output equations, should be well modeled by both methods. Though DCMf can model non-linear aspects of the hemodynamic response, these non-linearities are only important when using an estimated hemodynamic response from a short stimulus to predict the response to a longer stimulus or when considering only overly simplistic models of the neural response (Yesilyurt et al., 2008; Marxen et al., 2012). The Balloon–Windkessel model as implemented in DCMf and the linear convolution model produce equivalent outputs when inputs are slow and no distinction is typically seen between them (Seth et al., 2013). To ensure both DCMf and SLDSf were able to mimic the simulation hIRFs, both the basis function method of SLDSf and the non-linear system method of DCMf were directly tested. To test the SLDSf basis method, fifty random hIRFs were generated and optimal weights on the three bases were identified using ordinary least squares. The mean R^2 for the basis function method was 0.9966 and the minimum R^2 was 0.9809. For the non-linear system method of DCMf, the best fitting set of parameters (in terms of squared error) for the hemodynamic model were identified for the same 50 random hIRFs using the Levenburg–Marquardt method as implemented in the Matlab function `lsqnonlin()`. Both the mean and minimum r^2 for the non-linear system method were greater than 0.9999. While a paired t -test indicated that the superior performance of the non-linear system method over the basis function method was statistically significant [$t_{(49)} = 5.96, p < 0.001$] the basis function method fits were extraordinarily accurate and in practice the difference is likely meaningless.

METHODS

All simulations were performed in the Matlab (Mathworks Inc., Natic MA) 2012a environment. Matlab functions for generating the simulations are included as supplementary material. Briefly, for each individual simulation with n regions, an n -dimensional Hurwitz stable continuous time transition matrix was created for each of p conditions. For a continuous time transition matrix to

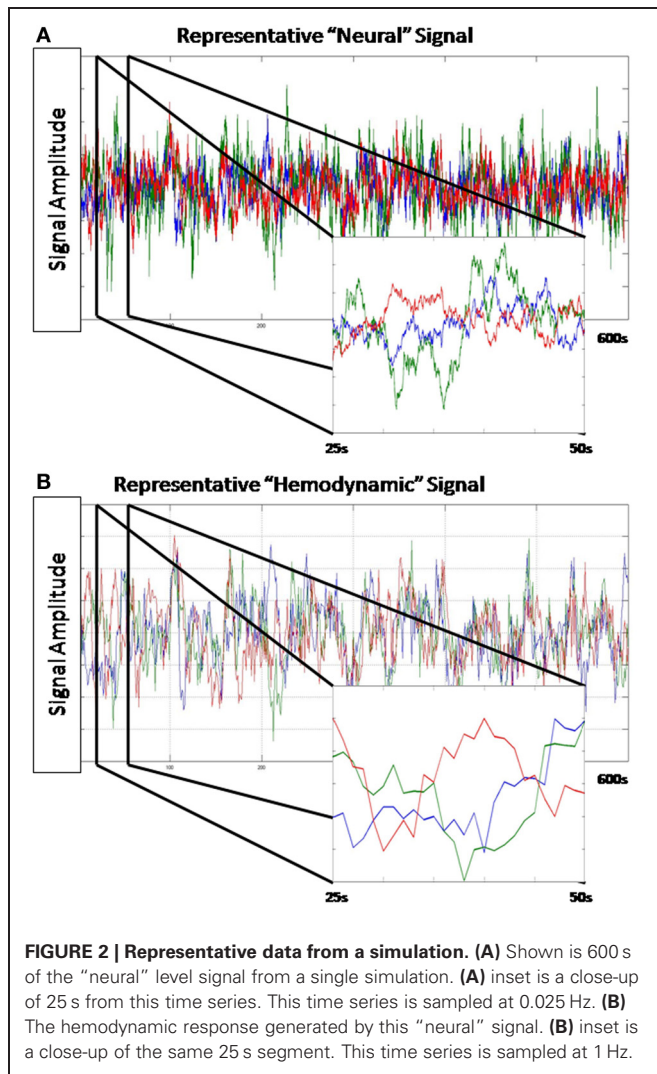
be Hurwitz stable, all of its eigenvalues must be in the open left half plane of the complex plane (i.e., negative real part). Stable transition matrices were created by drawing n real random values from a gamma distribution with shape parameter 3 and scale parameter 0.275 and negating them. These n negative values were treated as eigenvalues and left-multiplied, right-divided by an orthonormal n by n eigenvector matrix created using the `randn()` function and orthonormalized using the `orth()` function. The resulting symmetric matrix was rotated using an additional random n by n matrix. Entries in the resulting matrix with absolute value less than 0.1 were set to zero. The continued continuous time stability of the resulting matrix was tested by examining the eigenvalues of the constructed matrix. If a non-stable matrix was created, the process was repeated. The instantaneous error covariance matrix was created using n random positive eigenvalues from the range zero to one generated using the `rand()` function and a second orthonormal n by n eigenvector matrix without further rotation to maintain symmetry. The matrix describing the effect of external input on the state space was generated using the `randn()` function and all entries with absolute value less than 0.1 were set to zero. The resulting matrices were converted to their discrete time counterparts (A, B, and G) using Equations 14 to 17 with a step size of 25 ms. Discrete time versions of the true matrices were also created with a step size of 1000 ms for evaluation purposes.

The simulation consisted of a 10 min (600 s) time series having 25 blocks with 24 s per block. The ordering of the blocks in each simulation was created as follows. For simulations consisting of two conditions (i.e., $p = 2$), an ABAB design was used. For three or more conditions (i.e., $p > 2$) a Markov transition matrix was created with zero probability of remaining within condition an equal probability of jumping to any other condition. A random ordering of 25 blocks was generated using this Markov matrix and the `rand()` function.

Inputs to the state space were generated by randomly setting 60 of the 25 ms steps to 1 for each of two inputs. Discrete time noise was created from an orthonormal distribution for every 25 ms of a 600 s period. The signal data was then created at each 25 ms step using the condition series, input, noise, and system matrices according to Equation 11.

A hIRF was created for each of the n regions at the 25 ms step size. The mixture of two betas model implemented in the function `spm_hrf()` distributed with SPM8 ([www.fil.ion.ucl.ac.uk/spm/](#)) was used to create the hIRFs with the following parameter settings: $TR = 0.025$, delay of response peak from onset = $3.5\text{ s} + [0-3]\text{ s}$, delay of undershoot peak from onset = $12\text{ s} + [0-3]\text{ s}$, dispersion of response = 1, dispersion of undershoot = 1, ratio of response to undershoot = $11 + [0-3]$, length of kernel = 20 s. For each region, the signal from that region was convolved with the region's impulse response and then down-sampled by selecting every 40th value to create a 1 s step size. Orthogonal white noise was then added with a SNR of 6.0206 db. The resulting observations were mean centered and scaled to unit variance. A sample time series is presented in **Figure 2**.

For testing the methods, the number of regions $\in \{3, 5\}$ and conditions $\in \{2, 3\}$ in the simulated dataset were varied, and 50 different simulations were created for each combination of



parameter settings. Each of the methods was then applied to the simulated data. Analysis was conducted on a custom workstation (AVADirect, www.avadirect.com, Cleveland, OH) with two Xeon X5690 processors and 48GiB RAM running RedHat Enterprise Linux Release 6.1, Kernel 2.6.32-131.13.1.el6.x86_64. Identification of MAR models was essentially instantaneous for all simulations. Identification of sLDSf models took approximately 36 s for the most complex simulations while identification of sDCMf models took approximately 6495 s for the same simulations. Because of the time cost for estimating the sDCMf models, more complex simulations (e.g., more than 5 regions or more than 3 conditions) were not examined.

To directly compare the three methods, the continuous time transition matrices identified by sDCMf were converted to their discrete time counterparts at a 1 s sampling rate as were the original true continuous time simulation matrices. The sLDSf and MAR transition matrices were identified at this sampling rate and thus did not require further manipulation. Four measures were chosen to assess the relative accuracy of the methods. First, the r^2 for the comparison between the true and estimated A matrices

each method on each simulation according to Equation 29.

$$r^2 = 1 - \sum_{\text{elements}} (A^{\text{true}} - A^{\text{est}})^2 / (A^{\text{true}} - \bar{A}^{\text{true}})^2 \quad (29)$$

Note that while 29 is the correct formula, because the result is a subtraction of a squared value from one, there is a possibility that the resulting r^2 will be negative if the prediction of A is worse than simply guessing the mean of A. For sDCMf and sLDSf, the scaling of the signals of interest, and thus the A matrices, is arbitrary. Considering the sLDSf method equations, given an arbitrary invertible matrix T of the correct dimensions, a model with system matrices A and C is equivalent to one with matrices AT and CT⁻¹. Thus, any scaling that preserves the one-to-one spatial relationship between the signals within and observations of the regions is acceptable. To account for such scaling differences, two additional accuracy measures were calculated as in 30 and 31. In 31 κ is a scalar that

$$r^2 = 1 - \sum_{\text{elements}} (A^{\text{true}} - \kappa A^{\text{est}})^2 / (A^{\text{true}} - \bar{A}^{\text{true}})^2 \quad (30)$$

$$r^2 = 1 - \sum_{\text{elements}} (A^{\text{true}} - \kappa A^{\text{est}})^2 / (A^{\text{true}} - \bar{A}^{\text{true}})^2 \quad (31)$$

minimizes the squared difference between the matrices as identified via least-squares. In 3.3, K is a diagonal matrix where each element minimizes the squared difference between corresponding columns of the two matrices.

The simulation matrices were selected to be stable and thus are constrained to have a certain form. Both sDCMf and sLDSf use priors on the A matrices to bias parameter identification toward stable matrices. Thus the “chance” r^2 for the three accuracy measures above given this knowledge is greater than zero. To avoid this bias a final accuracy measure was calculated. Because more than one true transition matrix was created for each simulation (one per condition) the element-wise difference between the true matrices in each simulation was computed (i.e., the change in connectivity from one condition to another). The same difference was computed for the matrices estimated by the connectivity methods. The r^2 between the true difference and the estimated difference was calculated by converting the matrix differences to vectors and correlating them. The expected value of the correlation between the difference matrices is zero whether or not knowledge of the stability of the matrices is used.

In addition, a “chance” accuracy level was determined for the three matrix value accuracy measures. Each simulated transition matrix was used as an “estimate” of all other matrices of the same size with column-wise, global, and no scaling as above. These random estimates were then scaled as appropriate for each of the three measures considered involving the accuracy for a single matrix. This chance measure gives a baseline level of accuracy for a hypothetical method with perfect knowledge of how the system matrices were created. We note that this is more information than any of the methods here had access to and as such may not be a actual baseline but should still be a useful comparison.

RESULTS

Median r^2 values of the four accuracy measures for each of the three connectivity methods are given in **Tables 1–4**. In general all methods performed above chance levels for all comparisons. All sign tests, corrected for multiple comparisons, contrasting the three methods against the chance distributions for all simulation types were significant. Thus, all three methods were able to identify some interregional interactions from the simulated data. All three methods performed best when column-wise scaling was allowed followed by global scaling, then no scaling. All methods were least accurate when examining the differences between system matrices though these accuracies were still superior to chance (zero) levels. Using paired sign tests (Matlab function `signtest`) to compare median r^2 levels, the sLDSf method achieved superior accuracy to both sDCMf and MAR in all conditions. When column-wise scaling was permitted (**Table 1**), there were no significant differences in performance between the sDCMf and MAR methods. When only global scaling was permitted (**Table 2**), sDCMf outperformed MAR for the five-region simulations though there were no significant performance differences between these two methods for the three-region simulations. When no rescaling was permitted (**Table 3**), sDCMf performance was statistically better than MAR for all simulations except the simplest (two condition, three region) simulations where performance was not reliably different. When considering differences between matrices within simulations, sDCMf performance was statistically better than MAR for all combinations of number of conditions and regions.

Table 1 | Median accuracy of estimated transition matrices in r^2 .

Method	Two conditions	Three conditions	Number of regions
COLUMN-WISE SCALING OF ESTIMATED MATRICES			
sDCMf	0.7631	0.7887	3
	0.6816	0.7059	5
SLDSf	0.8778	0.8879	3
	0.8423	0.8149	5
MAR	0.8158	0.7702	3
	0.6904	0.6888	5

Table 2 | Median accuracy of estimated transition matrices in r^2 .

Method	Two conditions	Three conditions	Number of regions
GLOBAL SCALING OF ESTIMATED MATRICES			
sDCMf	0.6957	0.6817	3
	0.6140	0.6243	5
SLDSf	0.7710	0.7501	3
	0.7316	0.7412	5
MAR	0.7036	0.6404	3
	0.5902	0.6205	5

DISCUSSION

The current study presents a systematic and valid test between several connectivity methods including the popular sDCMf and MAR. The nature of connectivity estimated by the methods was described at the level of equations between signals of interest. A single equation was found that could accommodate the connectivity type estimated by all three methods. This allowed the three methods to be compared fairly using the same simulated datasets.

While none of the methods performed perfectly, all performed above chance levels. When differential scaling of the signals was accounted for, all methods performed well with most variances-accounted-for exceeding 70%. The more realistic global scaling results were also adequate with all methods performing at approximately $r^2 > 0.6$. Unfortunately true scaling information would not be available for *in vivo* data. Without scaling the performance of sDCMf and MAR, and to a lesser extent sLDSf, deteriorated. For sDCMf this was most likely due to the use of its own internal scaling which possibly differed from the scaling of the simulation. For MAR this was most likely due to the method's combination of observation noise and signal noise into a single component affecting the signals thus shrinking the connectivity parameters.

As noted however, if prior knowledge of the stability of the transition matrices is used, the expected accuracy of a random matrix is non-zero. The sDCMf and sLDSf methods were able to use this information in the form of priors during estimation. The MAR method applied here used no prior information regarding matrix stability and in addition, required fewer parameters. The reduced performance of MAR relative to sDCMf and sLDSf should be evaluated in light of these facts. Indeed, when column-wise scaling was permitted, performance of MAR was on par

Table 3 | Median accuracy of estimated transition matrices in r^2 .

Method	Two conditions	Three conditions	Number of regions
NO SCALING OF ESTIMATED MATRICES			
sDCMf	0.5144	0.5179	3
	0.5084	0.5248	5
SLDSf	0.7161	0.7247	3
	0.6852	0.6948	5
MAR	0.5238	0.4523	3
	0.4758	0.4677	5

Table 4 | Median accuracy of estimated differences in transition matrices in r^2 .

Method	Two conditions	Three conditions	Number of regions
sDCMf	0.3480	0.2436	3
	0.2263	0.1517	5
SLDSf	0.5868	0.4420	3
	0.3971	0.3706	5
MAR	0.2061	0.1052	3
	0.1274	0.1089	5

with that of sDCMf which used both prior knowledge and multiple additional parameters with a non-linear output function. A Bayesian MAR with appropriate priors would likely improve the performance of the method (Bishop, 2006). However, the QR based regression used here is similar to that used in popular MAR statistical packages (e.g., Schneider and Neumaier, 2001; He et al., 2011; Seth, 2010).

The correlation between the true and estimated task differences, rather than the individual matrices themselves, represents the strongest test of the methods. Prior knowledge of the form of a stable matrix would not improve performance on these differences since deviation from this form (i.e., the condition differences) is random. The sLDSf method performed better than the other methods recovering a median variance accounted for of almost 60% for the simplest condition. However, the remaining accuracies for the other conditions and those for the other methods were fairly poor, dropping to as low as 11% variance-accounted-for by the MAR method. The simulations here used a fairly realistic signal to noise level around 6 db. This corresponds to standard deviations of 0.8 and 0.2 for the signal and noise, respectively, when combined into the observations passed to the connectivity methods. This corresponds to levels seen in somewhat above average quality fMRI data. Higher SNR would likely increase performance of all methods. Given the time commitment required by sDCMf, testing the additional factor of SNR would be difficult. The effect of observation noise magnitude as well as noise type (e.g., white versus colored) on these connectivity methods remains an important area of further study.

The sDCMf and sLDSf methods are quite similar. However, there are three important distinctions between them. First is the inclusion of a probability model for the task in sLDSf that allows for estimation of the task sequence as well as signal states in additional datasets (Smith et al., 2010). Second, sLDSf uses a linear observation equation while sDCMf uses a non-linear system. Third, in estimating parameters, sLDSf uses state estimates conditioned on the full data set; that is, it estimates a probability density for the neural state conditioned on all of the available observed data and uses these estimates to identify parameters. In contrast, sDCMf conditions the probability of the signal state only on the observed data from the beginning of the series up to 25 s after the given time point (Daunizeau et al., 2009). This is because the non-linearity in the output equations makes conditioning on the full series difficult (though see Wan and van der Merwe, 2001). The task probability model contained in sLDSf was not used in the current simulations were it was assumed that the task sequence was known. The superior performance then of sLDSf relative to sDCMf was likely due to the full data conditioning and the relative ease of fitting parameters of linear versus non-linear functions. It is well known that “smoothed” estimates that are conditioned on the full data set are superior to “filtered” estimates that consider only the data up to the current time point. This was shown in an fMRI context using a model similar to sLDSf (Penny et al., 2005). Though the non-linear system output equation used in sDCMf is more flexible and can mimic a greater variety of hIRFs than the linear equation, identifying these parameters requires more complex algorithms. As shown above, the non-linear system model, though based on biophysical realism, is only slightly

more accurate than the linear basis method at estimating hIRFs when no non-linearities are present in the signal. For block experimental designs or well spaced event related designs, the cost of this non-linear system model in terms of connectivity parameter accuracy likely exceeds any benefit due to superior estimation of the hIRF. The benefit of the non-linearity would be restricted to situations where stimuli and trials were of variable length and included very short stimuli or where hemodynamic parameters are estimated using data with trials or inputs less than 1–2 s then applied to data with a much longer trials or inputs (Friston et al., 2000). In these cases possible non-linearity in the hIRFs would result in an incorrect response by a linear model. Neither of these situations was applicable in the simulations here thus the non-linearity was not necessary.

All of the tested methods were far from perfect at identifying connectivity parameters from the data sets examined here. While higher SNR simulated data would likely increase performance, all of the presented methods suffer from the same flaws that are present in essentially all models of fMRI connectivity. The directed transfer functions of the simulated connectivity matrices considered here contained power at frequencies higher than the Nyquist of typical fMRI sampling rates. Information regarding this portion of the simulated connectivity would be absent in simulated fMRI at realistic sampling rates (cf. Seth et al., 2013). While aliasing of these higher frequency signals is not an issue given that the hIRF passes little below 0.25 Hz (Henson, 2003), the slowness and dispersion of the hemodynamic response creates its own problems. **Figure 3A** shows the mean (\pm one standard deviation) coherence between the true signals of interest and those estimated by sLDSf for the two condition, three region simulations. The location of the task transition effect in frequency space is identified in black. Clearly, there is little relation between the true and estimated signals at frequencies higher than 0.2 Hz. Compare **Figure 3A** with **Figure 3B** which shows the mean (\pm one standard deviation) coherence between the true signals of interest and true simulated fMRI data from the same set of data. The difference between the two coherence estimates is shown in **Figure 3C** where negative values indicate higher coherence for the sLDSf estimate. The coherence between the true signals and the sLDSf estimated signals is slightly better than the simulated fMRI for lower frequency portions of the data. However, due to the hemodynamic filtering, neither the simulated fMRI data nor the sLDSf estimate of the signal accurately approximates the high frequency portions of the true signal of interest.

Unfortunately, the higher frequency components of the true signal are simply not contained in the measured fMRI data when examined as a time series. Though a proper experimental design can be used to identify sub-second differences in BOLD signal onsets, data sampled at 1 Hz. cannot contain information regarding signals with frequencies higher than 0.5 Hz. None of the methods examined here can be expected to recover information about a signal that is not there to recover. None of the methods examined here will truly identify interactions occurring between regions a few hundreds of milliseconds apart unless the interaction produces longer lasting consequences. Temporal directionality of individual neural “spikes” 100 ms. Apart in time cannot be seen using fMRI data at conventional sampling rates.

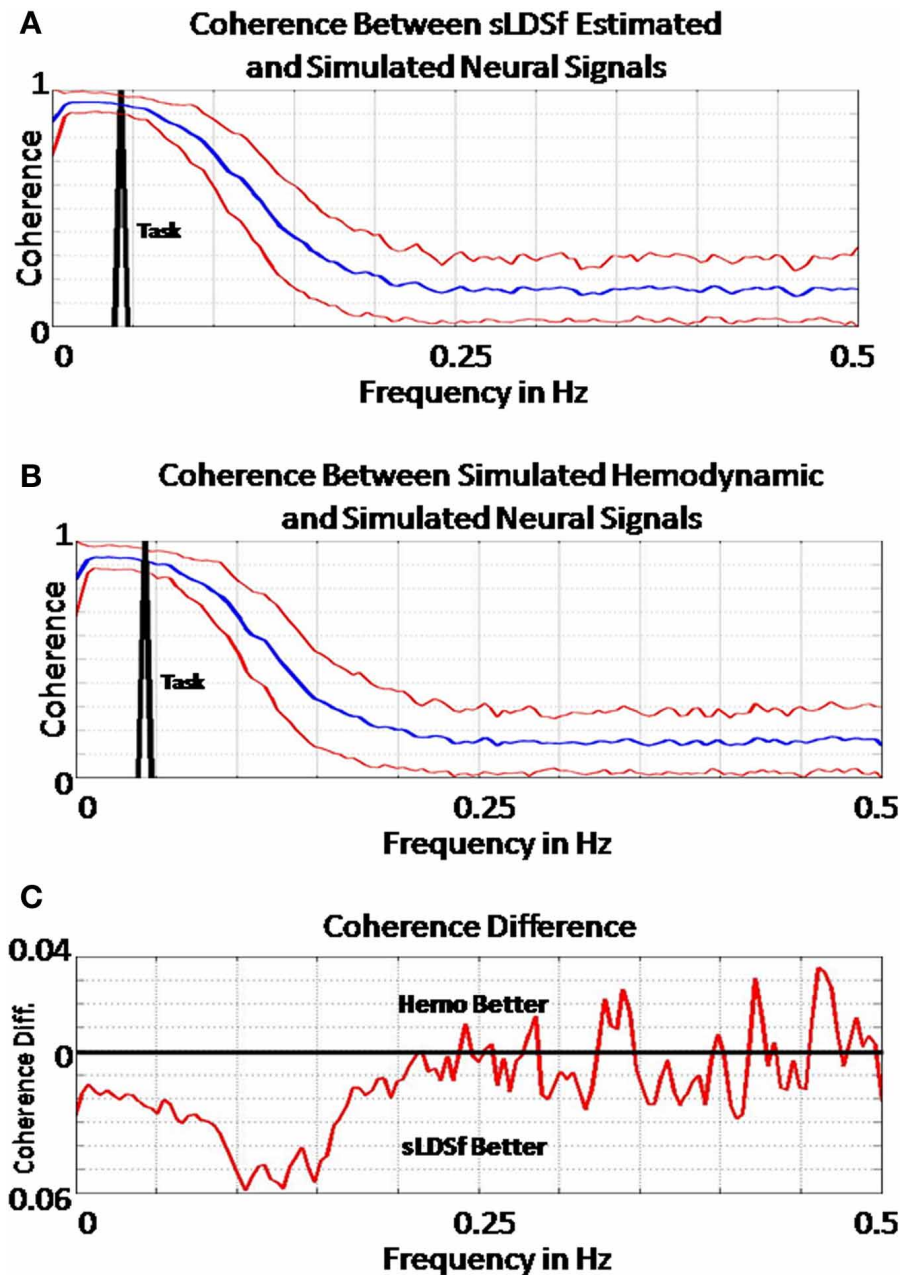


FIGURE 3 | (A) Mean (± 1.96 SD) coherence estimates between the simulated “neural” signals and those estimated via sLDSf across all two task block simulations. The frequency of task block alternation is shown in black. Coherence between the signals becomes negligible for frequencies above 0.25 Hz. **(B)** A Mean (± 1.96 SD) coherence estimates between the simulated hemodynamic and simulated “neural” signals across all two task block simulations. The frequency of task block alternation is again shown in black. As with the estimated data, coherence between the signals becomes

negligible for frequencies above 0.25 Hz. **(C)** Mean difference between the coherence estimates for the estimated “neural” signal and the simulated hemodynamic measurements. The zero difference line is shown in black. Values above this line are frequencies where the simulated hemodynamic signal had greater coherence with the simulated “neural” signal than did the estimated “neural” signal. The estimated “neural” signals have marginally better coherence with the true simulated neural signals until approximately 0.2 Hz after which the coherence differences are essentially random.

However, lower frequency fluctuations in the power of high frequency signals may be observed and are likely more relevant to connectivity analysis in fMRI (cf., Martínez-Monts et al., 2004). MAR models and linear dynamic systems can be indented within limited frequency bands (Wills et al., 2009). Such

frequency restrictions could force the methods to ignore higher frequency components of the signal that could not possibly be passed through the hemodynamic filter. This would have the added benefit of encouraging researchers who use connectivity methods to consider the meaning of the identified connectivity

parameters from a signal processing stand-point rather than simply considering the numeric value of the parameters in isolation (Smith et al., 2011b).

Based on the overview presented here, we make the following recommendations for fMRI connectivity simulation studies. While simulation based studies of connectivity methods are necessary, they are not trivial to perform. Some parts of some models are directly comparable after appropriate conversion. However, some models will always remain distinct. For example, deterministic DCM cannot be used to simulate data for SEM; one considers the task effect on internal dynamics (changes in A) in the absence of state noise while the other considers the task effect on state noise patterns (changes in G) in the absence of internal dynamics. Comparisons of methods that ignore this issue as well as issues concerning the distinction between continuous and discrete time can lead to false conclusions (cf. Smith et al., 2011a). In addition, simulation studies that fail to consider signal processing realities relevant to fMRI (for example by simulating then attempting to recover directionality of individual neural spikes separated by a few tens of milliseconds) can also lead to incorrect conclusions. For all simulations then, care must be taken to identify the nature of the connectivity simulated by a model as well as the extent to which that type of connectivity is present in the actual data simulated.

Based on the results presented here, we make the following recommendations for fMRI connectivity analyses. For exploratory analyses, unless non-linearities in the hIRFs are expected to be relevant, sDCMf is over-parameterized, slow, and less accurate and is thus not recommended. SLDSf outperformed MAR models but at a cost of additional parameters. For small models (i.e., less than 10 regions) the additional parameters of SLDSf relative to MAR may not be of concern and SLDSf is recommended. However, for larger models where the number of parameters would approach the degrees of freedom of the data, MAR with appropriate priors may be preferred. Because of the speed of estimating parameters in MAR models, this preference extends to full brain, seed region analyses where hundreds of thousands of models are

identified such as Granger Causal Mapping (Roebroeck et al., 2005). For confirmatory models where certain connections are assumed to take known values (e.g., 0), estimating a model in continuous time may be preferable. The conversion from continuous to discrete time using Equation 17 for models with extensive feedback can cause connections “known” to be zero in continuous time to be non-zero when considered in discrete time. In such cases, there will be multiple continuous time models that are all equally compatible with the discrete time data and a priori knowledge must be used to select one over the others. To account for known parameter values defined in continuous time, the models can be identified in continuous time (as in sDCMf) or identified in discrete time and a second optimization applied to convert this model to continuous time appropriate known zeros (Ljung and Wills, 2010). It remains to be seen if there is a preference for either of these methods in fMRI.

CONCLUSION

Connectivity, though widely used, has not always been well-defined. Comparing methods that use divergent underlying connectivity equations using a single simulated data set can be problematic. Here we compared the performance of three methods with a single common connectivity equation. Performance of all methods was better than chance though no method could capture high frequency components of the data. The results here, as well as basic signal processing theory, suggest that fMRI researchers should avoid searching for interregional interactions at frequencies higher than those passed by the hemodynamic filter. Researchers should also avoid interpreting fMRI connectivity results in terms of single event communications between regions on timescales of tens or even hundreds of milliseconds.

ACKNOWLEDGMENTS

This work was supported by the intramural program at the National Institute on Deafness and Communication Disorders at the National Institutes of Health.

REFERENCES

- Aguirre, G. K., Zarahn, E., and D'Esposito, M. (1998). The variability of human BOLD hemodynamic responses. *Neuroimage* 8, 360–369.
- Alexander, G. E., and Moeller, J. R. (1994). Application of the scaled subprofile model to functional imaging in neuropsychiatric disorders: a principal component approach to modeling brain function and disease. *Hum. Brain Mapp.* 2, 79–94.
- Åström, J. K. (1970). *Introduction to Stochastic Control Theory*. New York, NY: Academic Press.
- Åström, J. K., and Wittenmark, B. (1996). *Computer-Controlled Systems: Theory and Design*. Upper Saddle River, NJ: Prentice-Hall.
- Barber, D. (2006). Expectation correction for smoothed inference in switching linear dynamical systems. *J. Mach. Learn. Res.* 7, 2515–2540.
- Barnet, L. C., and Seth, A. K. (2011). Behaviour of Granger causality under filtering: theoretical invariance and practical application. *J. Neurosci. Methods* 201, 404–419.
- Bar-Shalom, Y., Li, X. R., and Kirubarajan, T. (2001). *Estimation with Applications to Tracking and Navigation*. New York, NY: John Wiley and Sons.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Vol. 4. New York, NY: Springer.
- Bokde, A. L. W., Ewers, M., and Hampel, H. (2009). Assessing neuronal networks: understanding Alzheimer's disease. *Prog. Neurobiol.* 89, 125–133.
- Büchel, C., Coull, J. T., and Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science* 283, 1538–1541.
- Bullmore, E. T., Long, C., Suckling, J., Fadili, J., Calvert, G., Zelaya, F., et al. (2001). Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Hum. Brain Mapp.* 12, 61–78.
- Bullmore, E. T., Rabe-Hesketh, S., Morris, R. G., Williams, S. C. R., Gregory, L., Gray, J. A., et al. (1996). Functional magnetic resonance image analysis of a large-scale neurocognitive network. *Neuroimage* 4, 16–23.
- Buxton, R. B. (2012). Dynamic models of BOLD contrast. *Neuroimage* 62, 953–961.
- Buxton, R. B., Wong, E. C., and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the Balloon model. *Magn. Reson. Med.* 39, 855–864.
- Calhoun, V. D., Adali, T., Pearlson, G. D., van Zijl, P. C. M., and Pekar, J. J. (2002). Independent component analysis of fMRI data in the complex domain. *Magn. Reson. Med.* 48, 180–192.
- Chen, G., Saad, Z. S., Glen, D. R., Hamilton, J. P., Thomason, M. E., Gotlib, I. H., et al. (2011). Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis. *Comput. Biol. Med.* 41, 1142–1155.
- Culver, W. J. (1966). On the existence and uniqueness of the real

- logarithm of a matrix. *Proc. Am. Math. Soc.* 17, 1146–1151.
- Daunizeau, J., Friston, K. J., and Kiebel, S. J. (2009). Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. *Physica D* 238, 2089–2118.
- David, O., Guillemain, I., Saitlet, S., Rey, S., Deransart, C., Segebarth, C., and Depaulis, A. (2008). Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS Biol.* 6:e315. doi:10.1371/journal.pbio.0060315
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. Ser. B* 39, 1–38.
- Diamantaras, K. I., and Kung, S. Y. (1996). *Principal Component Neural Networks: Theory and Applications*. New York, NY: John Wiley and Sons.
- Fisher, F. (1970). A correspondence principle for simultaneous equation models. *Econometrica* 28, 79–92.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2010). Bayesian nonparametric methods for learning Markov switching processes. *IEEE Signal Process. Mag.*, 27, 43–54.
- Franklin, G. F., Powell, D. J., and Workman, M. L. (1997). *Digital Control of Dynamic Systems*. Upper Saddle River, NJ: Prentice Hall.
- Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Hum. Brain Mapp.* 2, 56–78.
- Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302.
- Friston, K. J., Jezzard, P., and Turner, R. (1994). Analysis of functional MRI time series. *Hum. Brain Mapp.* 1, 153–171.
- Friston, K. J., Mechelli, A., Turner, R., and Price, C. J. (2000). Nonlinear responses in fMRI: the balloon model Volterra kernels, and other hemodynamics. *Neuroimage* 4, 466–477.
- Friston, K. J., Stephan, K. E., Li, B., and Daunizeau, J. (2010). Generalized filtering. *Math. Probl. Eng.* 621670. doi:10.1155/2010/621670
- Ghahramani, Z., and Hinton, G. E. (1996). *Parameter Estimation for Linear Dynamical Systems*. Technical Report CRG-TR-96–92. Toronto, ON: University of Toronto.
- Harrison, L., Penny, W. D., and Friston, K. (2003). Multivariate autoregressive modeling of fMRI time series. *Neuroimage* 19, 1477–1491.
- Haken, H. (1996). *Principles of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior, and Cognition*. Berlin: Springer.
- Handwerker, D. A., Gonzalez-Castillo, J., D'Esposito, M., and Bandettini, P. A. (2012). The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *Neuroimage* 62, 1017–1023.
- Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651.
- Havlicek, M., Friston, J. K., Jan, J., Brazdil, M., and Calhoun, V. D. (2011). Dynamic modeling of neural responses in fMRI using cubature Kalman filtering. *Neuroimage* 56, 2109–2128.
- Havlicek, M., Jan, J., Brazdil, M., and Calhoun, V. D. (2010). Dynamic Granger causality based on Kalman filter for evaluation of functional network connectivity in fMRI data. *Neuroimage* 53, 65–77.
- Haykin, S. (2002). *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice-Hall.
- He, B., Dai, Y., Astolfi, L., Babiloni, F., Yuan, H., and Yang, L. (2011). eConnectome: a MATLAB toolbox for mapping and imaging of brain functional connectivity. *J. Neurosci. Methods* 195, 261–269.
- Henson, R. (2003). “Analysis of fMRI timeseries: linear time-invariant model, event related fMRI and optimal experimental design,” in *Human Brain Function*, eds R. S. J. Frackowiak, K. J. Friston, C. Frith, R. Dolan, C. J. Price, S. Zeki, J. Ashburner, and W. D. Penny (Academic Press).
- Horwitz, B. (2003). The elusive concept of brain connectivity. *Neuroimage* 19, 466–470.
- Horwitz, B., Grady, C. I., Haxby, J. V., Ungerleider, L. G., Schapiro, M. B., Mishkin, M., et al. (1992). Functional associations among posterior extrastriate brain region during object and spatial vision. *J. Cogn. Neurosci.* 4, 311–322.
- Horwitz, B., and Rowe, J. B. (2011). Functional biomarkers for neurodegenerative disorders based on the network paradigm. *Prog. Neurobiol.* 95, 505–509.
- Jolliffe, I. T. (1986). *Principal Component Analysis*. New York, NY: Springer-Verlag.
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior*. Cambridge, MA: MIT Press.
- Kim, C. J. (1994). Dynamic linear models with Markov-switching. *J. Econom.* 60, 1–22.
- Lacerda, G., Spirtes, P., Ramsey, J., and Hoyer, P. (2008). “Discovering cyclic causal models by independent components analysis,” in *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence* (Helsinki).
- Li, B., Daunizeau, J., Stephan, K. E., Penny, W., Hu, D., and Friston, K. J. (2011). Generalised filtering and stochastic DCM for fMRI. *Neuroimage* 58, 442–457.
- Ljung, L., and Wills, A. G. (2010). Issues in sampling and estimating continuous-time models with stochastic disturbances. *Automatica* 46, 925–931.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature* 453, 869–878.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.
- Logothetis, N. K., and Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magn. Reson. Imaging* 22, 1517–1531.
- Lütkepohl, H. (2007). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.
- Makni, S., Beckmann, C., Smith, S., and Woolrich, M. (2007). Bayesian deconvolution of fMRI data using bilinear dynamical systems. *Neuroimage* 42, 1381–1396.
- Mandeville, J. B., Marota, J. J., Ayata, C., Zarachuk, G., Moskowitz, M. A., Rosen, B., et al. (1999). Evidence of a cerebrovascular postarteriole windkessel with delayed compliance. *J. Cereb. Blood Flow Metab.* 19, 679–689.
- Marrelec, G., Krainik, A., Duffau, H., Péligrini-Issac, M., Lehericy, S., Doyon, J., et al. (2006). Partial correlation for functional brain interactivity investigation in functional MRI. *Neuroimage* 32, 228–237.
- Martínez-Monts, E., Valdés-Sosa, P. A., Miwakeichi, F., Goldman, R. I., and Cohen, M. S. (2004). Concurrent EEG/fMRI analysis by multiway Partial Least Squares. *Neuroimage* 22, 1023–1034.
- Marxen, M., Cassidy, R. J., Dawsom, T. L., Ross, B., and Graham, S. J. (2012). Transient and sustained components of the sensorimotor BOLD response in fMRI. *Magn. Reson. Imaging* 30, 837–847.
- McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory* 7, 523–548.
- McIntosh, A. R. (2001). Towards a network theory of cognition. *Neural Netw.* 13, 861–876.
- McIntosh, A. R., Bookstein, F. L., Haxby, J. V., and Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage* 3, 143–157.
- McIntosh, A. R., and Gonzalez-Lima, F. (1994). Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* 2, 2–22.
- McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., et al. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188.
- Mechelli, A., Penny, W. D., Price, C. J., Gitelman, D. R., and Friston, K. J. (2002). Effective connectivity and intersubject variability: using a multisubject network to test differences and commonalities. *Neuroimage* 17, 1459–1469.
- Minka, T. (2000). *Bayesian Linear Regression*. Technical report. Cambridge MA: MIT.
- Moeller, J. R., and Strother, S. C. (1991). A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *J. Cereb. Blood Flow Metab.* 11, A121–A135.
- Murphy, K. (1998). *Learning Switching Kalman Filter Models*. Cambridge MA: Compaq Cambridge Research Lab Tech Report 98-10.
- Neumaier, A., and Schneider, T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.* 27, 27–57.
- Patel, R. S., Bowman, F. D., and Rilling, J. K. (2006). A Bayesian approach to determining connectivity of the human brain. *Hum. Brain Mapp.* 27, 267–276.
- Penny, W. D., Ghahramani, Z., and Friston, K. J. (2005). Bilinear dynamical systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 983–993.
- Penny, W. D., Stephan, K. E., Mechelli, A., and Friston, K. J. (2004). Modeling functional integration: a comparison of structural equation and dynamic causal models. *Neuroimage* 23(Suppl. 1), S264–S274.
- Rajapakse, J. C., and Zhou, J. (2007). Learning effective brain connectivity with dynamic Bayesian networks. *Neuroimage* 37, 749–760.
- Raol, J. R., Puthenpura, S., and Sinha, N. K. (1987). Algorithms for transformation of multivariable discrete-time models to

- continuous-time models. *Adv. Model. Simul.* 6, 52–62.
- Richardson, T. (1996). “A discovery algorithm for directed cyclic graphs,” in *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence* (Portland, OR).
- Riera, J. J., Watanabe, J., Kazuki, I., Naoki, M., Aubert, E., Ozaki, T., et al. (2004). A state space model of the hemodynamic approach: non-linear filtering of BOLD signals. *Neuroimage* 25, 547–567.
- Roebroeck, A., Formisano, E., and Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* 25, 230–242.
- Roebroeck, A., Formisano, E., and Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58, 296–302.
- Rowe, J. B. (2010). Connectivity analysis is essential to understand neurological disorders. *Front. Syst. Neurosci.* 4:144. doi: 10.3389/fnsys.2010.00144
- Roweis, S., and Ghahramani, Z. (1999). A unifying review of linear Gaussian models. *Neural Comput.* 11, 305–346.
- Rubio-Ramírez, J. F., Waggoner, D. F., and Zha, T. (2010). Structural vector autoregressions: theory of identification and algorithms for inference. *Rev. Econ. Stud.* 77, 665–696.
- Ryali, S., Supekar, K., Chen, T., and Menon, V. (2011). Multivariate dynamical systems models for estimating causal interactions in fMRI. *Neuroimage* 54, 807–823.
- Schneider, T., and Neumaier, A. (2001). Algorithm 808: ARfit-A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw.* 27, 58–65.
- Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. *Network* 16, 35–54.
- Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *J. Neurosci. Methods* 186, 262–273.
- Seth, A. K., Chorley, P., and Barnett, L. C. (2013). Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *Neuroimage* 65, 540–555.
- Shimizu, S., Hoyer, P., Hyvärinen, A., and Kerminen, A. (2006). A linear non-gaussian acyclic model for causal discovery. *J. Mach. Learn. Res.* 7, 2003–2030.
- Shumway, R. H., and Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* 3, 253–264.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* 48, 1–48.
- Sims, C. A., and Zha, T. (1998). Bayesian methods in dynamic multivariate models. *Int. Econ. Rev.* 39, 949–968.
- Smith, J. F., Chen, K., Johnson, S., Morrone-Strupinsky, J., Reiman, E. M., Nelson, A., et al. (2006). Network analysis of single subject fMRI during a finger opposition task. *Neuroimage* 32, 325–332.
- Smith, J. F., Pillai, A. S., Chen, K., and Horwitz, B. (2010). Identification and validation of effective connectivity networks in functional magnetic resonance imaging using switching linear dynamic systems. *Neuroimage* 52, 1027–1040.
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011a). Network modeling methods for FMRI. *Neuroimage* 54, 875–891.
- Smith, J. F., Pillai, A., Chen, K., and Horwitz, B. (2011b). Effective connectivity modeling for fMRI: Six issues and possible solutions using linear dynamic systems. *Front. Syst. Neurosci.* 5:104. doi: 10.3389/fnsys.2011.00104
- Stewart, G. W. (1987). Collinearity and least squares regression. *Stat. Sci.* 2, 68–84.
- Storkey, A. J., Simonotto, E., Whalley, H., Lawrie, S., Murray, L., and McGonigle, D. (2007). “Learning structural equation models for fmri,” in *Advances in Neural Information Processing Systems, Vol. 19*, eds B. Schölkopf, J. Platt, and T. Hoffman (Cambridge, MA: MIT Press), 1329–1336.
- Sun, F. T., Miller, L. M., and D’Esposito, M. (2004). Measuring interregional functional connectivity using coherence and partial coherence analyses of fMRI data. *Neuroimage* 21, 647–658.
- Tipping, M. E., and Bishop, C. M. (1999). Probabilistic principal component analysis. *J. R. Statist. Soc. Ser. B* 61, 611–622.
- Valdés-Sosa, P. A. (2004). Spatio-temporal autoregressive models defined over brain manifolds. *Neuroinformatics* 2, 239–250.
- Valdés-Sosa, P. A., Roebroeck, A., Daunizeau, J., and Friston, K. (2011). Effective connectivity: influence, causality and biophysical modelling. *Neuroimage* 58, 339–361.
- Valdés-Sosa, P. A., Sanchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Garcia, L., et al. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 969–981.
- Wan, E. A., and van der Merwe, R. (2001). “The unscented Kalman filter,” in *Kalman Filtering and Neural Networks*, ed S. Haykin (New York, NY: Wiley Publishing), 221–280.
- Whalley, H. C., Steele, J. D., Mukherjee, P., Romaniuk, L., McIntosh, A. M., Hall, J., et al. (2009). Connecting the brain and new drug targets for schizophrenia. *Curr. Pharm. Des.* 15, 2615–2631.
- Wills, A., Ninness, B., and Gibson, S. (2009). Maximum likelihood estimation of state space models from frequency domain data. *IEEE Trans. Automat. Contr.* 54, 19–33.
- Wold, H. (1996). “Estimation of principal components and related models by iterative least squares,” in *Multivariate Analysis*, ed P. R. Krishnaiah (New York, NY: Academic Press), 391–420.
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometr. Intell. Lab. 2*, 37–52.
- Yesilyurt, B., Ugurbil, K., and Uludag, K. (2008). Dynamics and nonlinearities of the BOLD response at very short stimulus durations. *Magn. Reson. Imaging* 26, 853–862.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 February 2013; paper pending published: 14 March 2013; accepted: 16 April 2013; published online: 14 May 2013.

Citation: Smith JF, Chen K, Pillai AS and Horwitz B (2013) Identifying effective connectivity parameters in simulated fMRI: a direct comparison of switching linear dynamic system, stochastic dynamic causal, and multivariate autoregressive models. *Front. Neurosci.* 7:70. doi: 10.3389/fnins.2013.00070
This article was submitted to *Frontiers in Brain Imaging Methods*, a specialty of *Frontiers in Neuroscience*.

Copyright © 2013 Smith, Chen, Pillai and Horwitz. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

The discrete time dynamic system can be described using the temporal shift-operator, q as in A1:

$$qx_t = Ax_t + Bu_t + \varepsilon_t \quad (\text{A1})$$

where,

$$qx_t \equiv x_{t+1} \quad (\text{A2})$$

The delta-operator, δ can be defined in terms of the shift-operator:

$$\begin{aligned} \delta &\triangleq \frac{q-1}{\Delta} \\ \delta x_t &= \frac{(q-1)x_t}{\Delta} \\ \delta x_t &= \frac{x_{t+1} - x_t}{\Delta} \end{aligned} \quad (\text{A3})$$

where Δ is the step size. The same discrete time dynamic system can also be described using the delta operator:

$$\delta x_t = Ex_t + Fu_t + \xi \quad (\text{A4})$$

Collecting terms of x_t yields:

$$\delta x_t - Ex_t = Fu_t + \xi \quad (\text{A5})$$

Using the definition of the delta-operator and the shift-operator yields,

$$\begin{aligned} \frac{(q-1)x_t}{\Delta} - Ex_t &= Fu_t + \xi \\ \frac{qx_t - x_t}{\Delta} - Ex_t &= Fu_t + \xi \\ \Delta^{-1}x_{t+1} - \Delta^{-1}x_t - Ex_t &= Fu_t + \xi \\ x_{t+1} - (I + E\Delta)x_t &= \Delta(Fu_t + \xi) \\ x_{t+1} &= (I + E\Delta)x_t + \Delta Fu_t + \Delta \xi \end{aligned} \quad (\text{A6})$$

Thus,

$$\begin{aligned} A &= (I + E\Delta) \\ B &= F\Delta \end{aligned} \quad (\text{A7})$$

Equation A6 shows the utility of the delta-operator form. If the sampling rate is very fast relative to the dynamics of the system, $E\Delta$ approaches zero and A approaches I .