Models of Effective Connectivity in Neural Systems

Klaas Enno Stephan and Karl J Friston

Wellcome Dept. of Imaging Neuroscience, Institute of Neurology, University College London, 12 Queen Square, WC1N 3BG, London, UK.

It is a longstanding scientific insight that understanding processes that result from the interaction of multiple elements require mathematical models of system dynamics (von Bertalanffy 1969). This notion is an increasingly important theme in neuroscience, particularly in neuroimaging, where causal mechanisms in neural systems are described in terms of effective connectivity. Here, we review established models of effective connectivity that are applied to data acquired with positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG) or magnetoencephalography (MEG). We start with an outline of general systems theory, a very general framework for formalizing the description of systems. This framework will guide the subsequent description of various establishd models of effective connectivity, including structural equation modeling (SEM), multivariate autoregressive modeling (MAR) and dynamic causal modeling (DCM). We focus particularly on DCM which distinguishes between neural state equations and a biophysical forward model that translates neural activity into a measured signal. After presenting some examples of applications of DCM to fMRI and EEG data, we conclude with some thoughts on pharmacological and clinical applications of models of effective connectivity.

1 General Systems Theory

The central goal of most scientific disciplines is to understand *systems*, i.e. ensembles of interacting elements. Today, this statement sounds almost trivial, yet in biology at least, the importance of the systems concept has been established only relatively recently. A key figure was Ludwig von Bertalanffy, a biologist and philosopher, who wrote a series of seminal articles in the first half of the 20^{th} century in which he argued that complex phenomena in biology (and indeed any other scientific field) invariably result from systems and could only be understood properly through a mathematical description of how system behavior emerged from the interactions of its constituent

elements. Demonstrating the existence of *system isomorphisms*, i.e. general mathematical descriptions that explained the dynamic behavior of very different kind of systems at different scales and across fields as diverse as physics, biology, economy and sociology, he introduced a very general framework that became known as *general system theory* (see the collection of essays in von Bertalanffy 1969). By the 1940s, the systems concept had experienced a scientific breakthrough in biology and led to the rise of *cybernetics*, "the science of control and communication in the animal and the machine" (Wiener 1948; Ashby 1956).

Today, biology uses the systems concept to address questions at all levels, from the molecular level to whole organisms and populations. The systems concept is now so omnipresent in biology that a recent special issue of the journal *Science* on systems biology renewed von Bertalanffy's (1969) previous diagnosis: "The [systems] concept has pervaded all fields of science and penetrated into popular thinking, jargon, and mass media" (Chong & Ray 2002).

But what exactly is a "system" and why is the systems concept so useful for framing scientific questions? A general, yet informal, definition is that a system is a set of elements which interact with each other in a spatially and temporally specific fashion. Before we attempt a formal definition of a system in the next section, let us remind ourselves that one of the classic scientific methods is to "analyze" a given phenomenon, i.e. to break it down into atomic units and processes that can be investigated independently of each other. This approach is appealing because it reduces a complex problem to a set of simpler problems, each of which can be addressed under conditions which can be controlled more easily for potentially confounding influences. For example, if one wanted to understand the physiological properties of a single neuron, one might decide to isolate it from its environment (e.g. let it grow in a dish) and then map its responses to currents injected into various parts of its dendritic tree. Unfortunately, this analytic approach cannot fully predict the neuron's behavior when it is part of a neural system, e.g. in the brain, and thus interacts with other neurons. When part of a system, the response of an individual neuron to a particular synaptic input (or injected current) u_1 depends on the spatial and temporal distribution of inputs $u_1 \ldots u_n$ that its dendritic tree receives from other neurons. If these additional inputs occur sufficiently close in time and space to u_1, they will affect the magnitude of the postsynaptic potential elicited by u_1, either linearly (by spatio-temporal summation) or nonlinearly (e.g. by changing the opening probability of voltage-gated channels) (Magee & Johnston 2005). In other words, the connectivity in the system mediates effects that cannot be predicted by studying a single neuron. Similar scenarios can be described for any other scientific field, for example biochemistry. Having studied a set of different biochemical processes in isolation, one would not necessarily be able to predict their collective dynamics. The problem is, as above, that different processes may interact, e.g. one process may change the substrate/product ratio of another process, or the efficacy of an enzyme that is relevant for a particular process may change due to the presence

of allosteric (in)activators that are produced by a second process or due to dynamic changes in gene expression mediated by a third process.

In summary, the general problem of analytical procedures in science is that they are blind to predicting the consequences arising from interactions between the elements in a system. Analytical procedures therefore need to be complemented with a theoretical framework that takes into account both the connectivity between the elements and external perturbations in order to achieve a mechanistic explanation of the dynamics of the system as a whole. This framework is provided by general system theory.

2 A General Form for System Models

Why is it useful at all to strive for formal mathematical definitions of systems? First, as described below, it allows one to pinpoint precisely what is meant by structure, function, and structure-function-relationships. Second, it allows one to predict system behavior for situations in which the system has not been observed before (see Bossel 1992 for an impressive collection of examples from biology). Third, it is the only way to fully understand how a system works and particularly, how system function could be restored if some of its components are rendered dysfunctional, e.g. by disease (Payne & Lomber 2001).

Here, we choose deterministic differential equations with time-invariant parameters as a mathematical framework; note that these are not the only possible mathematical representation of dynamic systems (see Bar-Yam 1997 for alternatives). The underlying concept, however, is quite universal: a *system* is defined by a set of elements with n time-variant properties altogether that interact with each other. Each time-variant property x_i $(1 \leq i \leq n)$ is called a *state variable*, and the n-vector $x(t)$ of all state variables in the system is called the *state vector* (or simply *state*) of the system at time t:

$$x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \tag{1}$$

Taking an ensemble of interacting neurons as an example, the system elements would correspond to the individual neurons, each of which is represented by one or several state variables. These state variables could refer to various neurophysiological properties, e.g. postsynaptic potentials, status of ion channels, etc. This touches on an important distinction: in *system construction* (e.g. in engineering), the relevant state variables and their mutual dependencies are usually known; in *system identification* (e.g. when trying to understand a biological system), however, they are not known. This means that we always require a *model* of the system that represents our current hypothesis about the structure of the system and how its function emerges from that structure (the structure-function relationship, SFR).

The crucial point is that the state variables interact with each other, i.e. the evolution of each state variable depends on at least one other state variable. This mutual functional dependence between the state variables of the system is expressed in a very natural fashion by a set of ordinary differential equations that operate on the state vector:

$$\frac{dx}{dt} = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{bmatrix} = F(x) \tag{2}$$

However, this description is not yet sufficient. First of all, the specific form of the dependencies f_i needs to be specified, i.e. the nature of the causal relations between state variables. This requires a set of parameters θ which determine the form and strength of influences between state variables. In neural systems, these parameters usually correspond to time constants or strengths of the connections between the system elements. And second, in the case of non-autonomous systems (and these are the ones of interest to biology) we need to consider the input into the system, e.g. sensory information entering the brain. We represent the set of all m known inputs by the m-vector function $u(t)$. Extending (2) accordingly leads to a general state equation for non-autonomous deterministic systems

$$\frac{dx}{dt} = F(x, u, \theta) \tag{3}$$

where θ is the parameter vector of the system. Such a model provides a causal description of how system dynamics results from system structure, because it describes (i) when and where external inputs enter the system and (ii) how the state changes induced by these inputs evolve in time depending on the system's structure. As explained below in more detail in Sect. 3, (3) therefore provides a general form for models of *effective connectivity* in neural systems, i.e. the causal influences that neural units exert over another (Friston 1994).

We have made two main assumptions to simplify the exposition. First, it is assumed that all processes in the system are deterministic and occur instantaneously. Random components (noise) and delays could be accounted for by using stochastic differential equations and delay differential equations, respectively. Second, we assume that we know the inputs that enter the system. This is a tenable assumption in neuroimaging because the inputs are experimentally controlled variables, e.g. changes in stimuli or instructions.[1]

[1] Note that using time-invariant dependencies f_i and parameters θ is neither an assumption nor a restriction. Although the mathematical form of f_i per se is static, the use of time-varying inputs u allows for dynamic changes in what components of f_i are "activated". For example, using box-car functions that are multiplied with the different terms of a polynomial function one can induce changes from linear to nonlinear behavior (and vice versa) over time. Also, there is no principled distinction between states and time-invariant parameters. Therefore, estimating time-varying parameters can be treated as a state estimation problem.

On the basis of the general system description provided by (3) we can now state accurately, given a particular system model, what we mean by structure, function, and structure-function relationships (see Stephan 2004 for more details):

- S*tructure* is defined by the time-invariant components of the system, i.e. the binary nature of θ (which connections exist and which do not; see (8)) and the mathematical form of the state variable dependencies f_i.
- *Function* refers to those time-variant components of the system model that are conditional on its structure, i.e. $x(t)$, but not $u(t)$.
- The *structure-function relationship* (SFR) is represented by F: integrating F in time determines the temporal evolution of the system state x from time $t=0$ up to a time point τ, given an initial state $x(0)$:

$$x(\tau) = x(0) + \int_0^\tau F(x, u, \theta)dt \tag{4}$$

In other words, given a particular temporal sequence of inputs $u(t)$, (4) provides a complete description of how the dynamics of the system (i.e. the trajectory of its state vector x in time) results from its structure and initial state.

3 Functional Integration and Effective Connectivity are Assessed through System Models

Modern cognitive neuroscience has adopted an explicit system perspective. A commonly accepted view is that the brain regions that constitute a given system are computationally specialized, but that the exact nature of their individual computations depends on context, e.g. the inputs from other regions. The aggregate behavior of the system depends on this *neural context*, the context-dependent interactions between the system components (McIntosh 2000; see also the chapter by Bressler & McIntosh in this volume). An equivalent perspective is provided by the twin concepts of functional specialization and functional integration (Friston 2002). *Functional specialization* assumes a local specialization for certain aspects of information processing but allows for the possibility that this specialization is anatomically segregated across different cortical areas. The majority of current functional neuroimaging experiments have adopted this view and interpret the areas that are activated by a certain task component as the elements of a distributed system. However, this explanation is incomplete as long as no insight is provided into how the locally specialized computations are bound together by context-dependent interactions among these areas, i.e. the *functional integration* within the system. This functional integration within distributed neural systems can be characterized in two ways, functional connectivity and effective connectivity.

Functional connectivity has been defined as the temporal correlation between regional time series (Friston 1994). Analyses of functional connectivity therefore do not incorporate any knowledge or assumptions about the structure and the SFR of the system of interest. Depending on the amount of knowledge about the system under investigation, this can either be a strength or a weakness. If the system is largely unknown, functional connectivity approaches are useful because they can be used in an exploratory fashion, either by computing functional connectivity maps with reference to a particular seed region (Horwitz et al. 1998; McIntosh et al. 2003; Stephan et al. 2001a) or using a variety of multivariate techniques that find sets of voxels whose time series represent distinct (orthogonal or independent) components of the covariance structure of the data (Friston & Büchel 2004; McIntosh & Lobaugh 2004). The information from these analyses can then be used to generate hypotheses about the system. Conversely, given sufficient information about the system structure and a specific hypothesis about the SFR of the system, models of effective connectivity are more powerful. Here, we only deal with models of effective connectivity. For analyses of functional connectivity, please see the chapters by Salvador et al., Bressler & McIntosh and Sporns & Tononi in this volume.

Effective connectivity has been defined by various authors, but in complementary ways. A general definition is that effective connectivity describes the causal influences that neural units exert over another (Friston 1994). Other authors have proposed that "effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" (Aertsen & Preißl 1991). Both definitions emphasize that determining effective connectivity requires a causal model of the interactions between the elements of the neural system of interest. Such a causal model has to take into account the external inputs that perturb the system and the anatomical connections by which neural units influence each other. In other words, any such model is a special case of the general system model as described in Sect. 2 and formalized by (3).

The equations presented in Sect. 2 are extremely general. To illustrate how the concept of effective connectivity emerges naturally from system models, we discuss the special case of a linear dynamic system. Although most natural phenomena are of a nonlinear nature, linear models play an important role in systems science because (i) they are analytically tractable, and (ii) given sufficiently long observation periods and non-negligible external input, their dynamics are largely independent of the initial state (Bossel 1992). Therefore nonlinear systems are usually investigated in restricted sub-spaces of interest, using linear models as local approximations. The following model of n inter-

acting brain regions is a simple linear case of (3) which uses a single state variable per region and m external inputs:

$$\begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_n}{dt} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nm} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \tag{5}$$

In this model the change of any given element depends on the state of the other system elements and on external inputs which affect it directly or indirectly. This system model can be written in compact matrix form as

$$F(x) = \frac{dx}{dt} = Ax + Cu \tag{6}$$

where the non-zero values of A and C represent the parameters of the system (i.e. θ in (3)) and the state of the system at time point τ can be obtained by integration (compare (4))

$$x(\tau) = e^{A\tau}x(0) + \int_0^\tau e^{A(\tau-t)}Cu(t)dt \tag{7}$$

where e^{At} is the matrix exponential (Bossel 1992). In this model, the system's behavior has two separable components: intrinsically sustained dynamics (parameterized by matrix A) and dynamics enforced by external inputs (parameterized by matrix C). The first term of (6) says that the change of the state variable x_i is a linear mixture of all state variables in the system, weighted by the parameters a_{ij}. By defining a particular parameter a_{ij} to be zero, we disallow for a direct effect of x_j on x_i (see Fig. 1 for an example). Conversely, any non-zero parameter a_{ij} represents a causal influence of the dynamics of x_j on that of x_i. The binarized parameter matrix $\tilde{A}$

$$\tilde{A} = \chi(A) = \begin{bmatrix} \chi(a_{11}) & \cdots & \chi(a_{1n}) \\ \vdots & \ddots & \vdots \\ \chi(a_{n1}) & \cdots & \chi(a_{nn}) \end{bmatrix},$$
$$\chi(a) = \begin{cases} 1 \text{ if } a \neq 0 \\ 0 \text{ if } a = 0 \end{cases} \tag{8}$$

represents the *structural connectivity* of the system model (see the chapter by Sporns & Tononi in this volume on how patterns of anatomical connections constrain effective connectivity and thus the dynamics of neural systems). The definition of the structural connectivity is usually guided by anatomical investigations in primates (Stephan et al. 2001b, Kötter 2004; see the chapter

by Paus in this volume for alternative approaches in humans). The values of A represent the influences of system elements over each other and thus correspond to the *effective connectivity* within the system. Finally, the values of the matrix C represent the magnitude of the direct effects that external (e.g. sensory) inputs have on system elements. By setting a particular parameter c_{ij} to be zero, we disallow for a direct effect of the external input u_j on x_i (see Fig. 1 for an example). A and C represent the system parameters (θ) that one needs to estimate when fitting this model to measured data. Simple linear models of this kind have found widespread application in various scientific disciplines (von Bertalanffy 1969). In Sect. 6, we will see that Dynamic Causal Modelling (DCM, Friston et al. 2003) extends the above formulation by bilinear terms that model context-dependencies of intrinsic connection strengths.

It should be noted that the framework outlined here is concerned with dynamic systems in continuous time and thus uses differential equations. The same basic ideas, however, can also be applied to dynamic systems in discrete time (using difference equations), as well as to "static" systems where the system is at equilibrium at each point of observation. The latter perspective, which is useful for regression-like equations, is used by classic system models for functional neuroimaging data, e.g. psycho-physiological interactions (PPI; Friston et al. 1997), structural equation modeling (SEM; McIntosh et al. 1994; Büchel & Friston 1997) or multivariate autoregressive models (MAR; Harrison et al. 2003; Göbel et al. 2003). These will be described in the following sections.

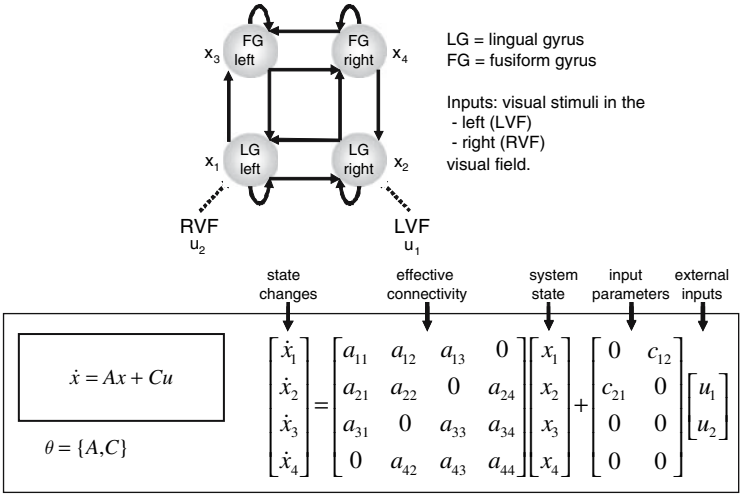

Fig. 1. A simple linear dynamic system as an example for a concrete implementation of (3), describing interactions between the lingual (LG) and the fusiform gyri (FG) in both hemispheres. The top panel shows the system structure and the sensory inputs (visual stimuli displayed in the left and right peripheral visual field) that perturb the system. The lower panel shows the state equation in matrix form

4 Psycho-Physiological Interactions (PPI)

PPI are one of the simplest models available to assess functional interactions in neuroimaging data (see Friston et al. 1997 for details). Given a chosen reference time series y_0 (obtained from a reference voxel or region), PPI computes whole-brain connectivity maps of this reference voxel with all other voxels y_i in the brain according to the regression-like equation

$$y_i = ay_0 + b(y_0 \times u) + cu + X\beta + e \tag{9}$$

Here, a is the strength of the intrinsic (context-independent) connectivity between y_0 and y_i. The bilinear term $y_0 \times u$ represents the interaction between physiological activity y_0 and a psychological variable u which can be construed as a contextual input into the system, modulating the connectivity between y_0 and y_i ($\times$ represents the Hadamard product, i.e. element-by element multiplication). The third term describes the strength c by which the input u determines activity in y_i directly, independent of y_0. Finally, β are parameters for effects of no interest X (e.g. confounds) and e is a Gaussian error term.

Notwithstanding the fact that this is a non-dynamic model, (9) contains the basic components of system descriptions as outlined in Sect. 2 and (3), and there is some similarity between its form and that of the state equation of DCM ((13), see below). However, since only pair-wise interactions are considered (i.e. separately between the reference voxel and all other brain voxels), this model is severely limited in its capacity to represent neural systems. This has also been highlighted in the initial description of PPIs (Friston et al. 1997). Although PPIs are not a proper system model, they have a useful role in exploring the functional interactions of a chosen region across the whole brain. This exploratory nature bears some similarity to analyses of functional connectivity. Unlike analyses of functional connectivity, however, PPIs model the contextual modulation of connectivity, and this modulation has a directional character, i.e. testing for a PPI from y_0 to y_i is not identical to testing for a PPI from y_i to y_0. This is because regressing $y_0 \times u$ on y_i is not equivalent to regressing $y_i \times u$ on y_0.

5 Structural Equation Modeling (SEM)

SEM has been an established statistical technique in the social sciences for several decades, but was only introduced to neuroimaging in the early 1990's by McIntosh & Gonzalez-Lima (1991). It is a multivariate, hypothesis-driven technique that is based on a structural model which represents the hypothesis about the causal relations between several variables (see McIntosh & Gonzalez-Lima 1994, Büchel & Friston 1997, Bullmore et al. 2000 and Penny et al. 2004a for methodological details). In the context of fMRI these variables are the measured BOLD (blood oxygen level dependent) time series

$y_1 \ldots y_n$ of n brain regions and the hypothetical causal relations are based on anatomically plausible connections between the regions. The strength of each connection $y_i \rightarrow y_j$ is specified by a so-called "path coefficient" which, by analogy to a partial regression coefficient, indicates how the variance of y_j depends on the variance of y_i if all other influences on y_j are held constant.

The statistical model of standard SEM implementations for neuroimaging data can be summarized by the equation

$$y = Ay + u \tag{10}$$

where y is a $n \times s$ matrix of n area-specific time series with s scans each, A is a $n \times n$ matrix of path coefficients (with zeros for non-existent connections), and u is a $n \times s$ matrix of zero mean Gaussian error terms, which are driving the modeled system ("innovations", see (11)). Note that the model on which SEM rests is a special case of the general equation for non-autonomous linear systems (with the exception that SEM is a static model and the inputs to the modeled system are random noise; compare (11) with (6)). Parameter estimation is achieved by minimizing the difference between the observed and the modeled covariance matrix Σ of the areas (Bollen 1989). For any given set of parameters, Σ can be computed by transforming (10):

$$\begin{aligned} y &= (I - A)^{-1}u \\ \Sigma &= yy^T \\ &= (I - A)^{-1}uu^T(I - A)^{-1^T} \end{aligned} \tag{11}$$

where I is the identity matrix and T denotes the transpose operator. The first line of 11 can be understood as a generative model of how system function results from the system's connectional structure: the measured time series y results by applying a function of the inter-regional connectivity matrix, i.e. $(I - A)^{-1}$, to the Gaussian innovations u.

In the special case of fMRI, the path coefficients of a SEM (i.e. the parameters in A) describe the effective connectivity of the system across the entire experimental session. What one would often prefer to know, however, is how the coupling between certain regions changes as a function of experimentally controlled context, e.g. differences in coupling between two different tasks. Notably, SEM does not account for temporal order: if all regional time series were permuted in the same fashion, the estimated parameters would not change. In case of blocked designs, this makes it possible to proceed as if one were dealing with PET data, i.e. to partition the time series into condition-specific sub-series and fit separate SEMs to them. These SEMs can then be compared statistically to test for condition-specific differences in effective connectivity (for examples, see Büchel et al. 1999; Honey et al. 2002). An alternative approach is to augment the model with bilinear terms (cf. (9)) which represent the modulation of a given connection by experimentally controlled variables (e.g. Büchel & Friston 1997; Rowe et al. 2002). In this case, only a single SEM is fitted to the entire time series.

One limitation of SEM is that one is restricted to use structural models of relatively low complexity since models with reciprocal connections and loops often become non-identifiable (see Bollen 1989 for details). There are heuristics for dealing with complex models that use multiple fitting steps in which different parameters are held constant while changing others (see McIntosh et al. 1994 for an example).

6 Multivariate Autoregressive Models (MAR)

In contrast to SEM, autoregressive models explicitly address the temporal aspect of causality in time series. They take into account the causal dependence of the present on the past: each data point of a regional time series is explained as a linear combination of past data points from the same region. MAR models extend this approach to n brain regions, modeling the n-vector of regional signals at time t (y_t) as a linear combination of p past data vectors whose contributions are weighted by the parameter matrices A_i:

$$y_t = \sum_{i=1}^{p} y_{t-i} A_i + u_t \tag{12}$$

MAR models thus represent directed influences among a set of regions whose causal interactions are inferred via their mutual predictability from past time points. Although MAR is an established statistical technique, specific implementations for neuroimaging were suggested only relatively recently. Harrison et al. (2003) suggested a MAR implementation that allowed for the inclusion of bilinear variables representing modulatory effects of contextual variables on connections and used a Bayesian parameter estimation scheme specifically developed for MAR models (Penny & Roberts 2002). This Bayesian scheme also determined the optimal model order, i.e. the number of past time points (p in (12)) to be considered by the model. A complementary MAR approach, based on the idea of "Granger causality" (Granger 1969), was proposed by Goebel et al. (2003). In this framework, given two time-series y_1 and y_2, y_1 is considered to be caused by y_2 if its dynamics can be predicted better using past values from y_1 and y_2 as opposed to using past values of y_1 alone.

7 Dynamic Causal Modeling (DCM)

An important limitation of the models discussed so far is that they operate at the level of the measured signals. Taking the example of fMRI, the model parameters are fitted to BOLD series which result from a haemodynamic convolution of the underlying neural activity. Any inference about inter-regional connectivity obtained by PPI, SEM or MAR is only an indirect one because

these models do not include the forward model linking neuronal activity to the measured haemodynamic data. In the case of EEG, this forward model means there is a big difference between signals measured at each electrode and the underlying neuronal activity: changes in neural activity in different brain regions lead to changes in electric potentials that superimpose linearly. The scalp electrodes therefore record a mixture, with unknown weightings, of potentials generated by a number of different sources.

The causal architecture of the system that we would like to identify is expressed at the level of neuronal dynamics. Therefore, to enable inferences about connectivity between neural units we need models that combine two things: (i) a parsimonious but neurobiologically plausible model of neural population dynamics, and (ii) a biophysically plausible forward model that describes the transformation from neural activity to the measured signal. Such models make it possible to fit jointly the parameters of the neural and of the forward model such that the predicted time series are optimally similar to the observed time series. In principle, any of the models described above could be combined with a modality-specific forward model, and indeed, MAR models have previously been combined with linear forward models to explain EEG data (Yamashita et al. 2004). So far, however, Dynamic Causal Modeling (DCM) is the only approach where the marriage between models of neural dynamics and biophysical forward models is a mandatory component. DCM has been implemented both for fMRI (Friston et al. 2003) and EEG/MEG data (David et al. 2006; Kiebel et al. 2006). These modality-specific implementations are briefly summarized in the remainder of this section (see Fig. 2 for a conceptual overview).

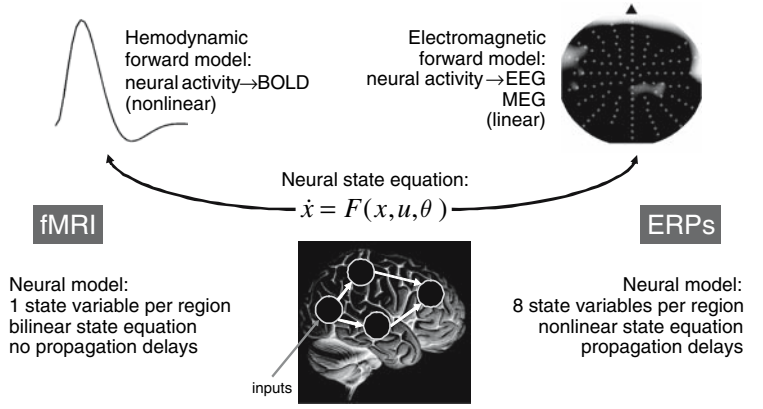

Fig. 2. A schematic overview that juxtaposes properties of DCM for fMRI and ERPs, respectively. It illustrates that DCM combines a model of neural population dynamics, following the generic form of (3), with a modality-specific biophysical forward model. Given appropriate formulations of the neural and the forward model, DCM can be applied to any kind of measurement modality

7.1 DCM for fMRI

DCM for fMRI uses a simple model of neural dynamics in a system of n interacting brain regions. It models the change of a neural state vector x in time, with each region in the system being represented by a single state variable, using the following bilinear differential equation:

$$\begin{aligned}\frac{dx}{dt} &= F\left(x, u, \theta^n\right) \\ &= \left(A + \sum_{j=1}^{m} u_j B^{(j)}\right) x + Cu \end{aligned} \tag{13}$$

Note that this neural state equation follows the general form for deterministic system models introduced by (3), i.e. the modeled state changes are a function of the system state itself, the inputs u and some parameters θ^n that define the functional architecture and interactions among brain regions at a neuronal level (n in θ^n is not an exponent but a superscript that denotes "neural"). The neural state variables represent a summary index of neural population dynamics in the respective regions. The neural dynamics are driven by experimentally controlled external inputs that can enter the model in two different ways: they can elicit responses through direct influences on specific regions (e.g. evoked responses in early sensory cortices; the C matrix) or they can modulate the coupling among regions (e.g. during learning or attention; the B matrices).

Equation (13) is a bilinear extension of (6) that was introduced earlier as an example of linear dynamic systems. Given this bilinear form, the neural parameters $\theta^n = \{A, B, C\}$ can be expressed as partial derivatives of F:

$$\begin{aligned} A &= \left.\frac{\partial F}{\partial x}\right|_{u=0} \\ B^{(j)} &= \frac{\partial^2 F}{\partial x \partial u_j} \\ C &= \left.\frac{\partial F}{\partial u}\right|_{x=0} \end{aligned} \tag{14}$$

The matrix A represents the effective connectivity among the regions in the absence of input, the matrices $B^{(j)}$ encode the change in effective connectivity induced by the j^{th} input u_j, and C embodies the strength of direct influences of inputs on neuronal activity (see Fig. 3 for a concrete example and compare it to Fig. 1).

DCM for fMRI combines this model of neural dynamics with an experimentally validated haemodynamic model that describes the transformation of neuronal activity into a BOLD response. This so-called "Balloon model" was initially formulated by Buxton et al. (1998) and later extended by

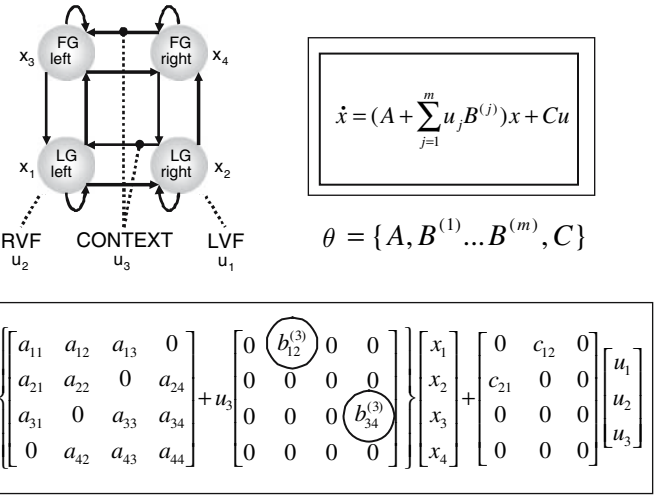

Fig. 3. A simple bilinear extension of the linear dynamic system shown in Fig. 1. This is an example for a concrete implementation of the neural state equation of DCM for fMRI. Note the role of the bilinear terms which model context-dependent (additive) changes of the strengths of the connections from the right to the left hemisphere (circled elements in the B matrix)

Friston et al. (2000). Briefly, it consists of a set of differential equations that describe the relations between four haemodynamic state variables, using five parameters (θ^h). More specifically, changes in neural activity elicit a vasodilatory signal that leads to increases in blood flow and subsequently to changes in blood volume and deoxyhemoglobin content. The predicted BOLD signal is a non-linear function of blood volume and deoxyhemoglobine content. Details of the haemodynamic model can be found in other publications (Friston et al. 2000; Stephan et al. 2004). Figure 4 provides a conceptual overview of DCM for fMRI.

The combined neural and haemodynamic parameter set $\theta = \{\theta^n, \theta^h\}$ is estimated from the measured BOLD data, using a fully Bayesian approach with empirical priors for the haemodynamic parameters and conservative shrinkage priors for the coupling parameters. Details of the parameter estimation scheme, which rests on a gradient ascent procedure embedded into an expectation maximization (EM) algorithm and uses a Laplace (i.e. Gaussian) approximation to the true posterior, can be found in Friston (2002). Eventually, the posterior distributions of the parameter estimates can be used to test hypotheses about connection strengths. Usually, these hypotheses concern context-dependent changes in coupling. If there is uncertainty about the connectional structure of the modeled system, or if one would like to compare competing hypotheses (represented by different DCMs), a Bayesian model selection procedure can be used to find the DCM that exhibits an optimal balance between model fit and model complexity (Penny et al. 2004b).

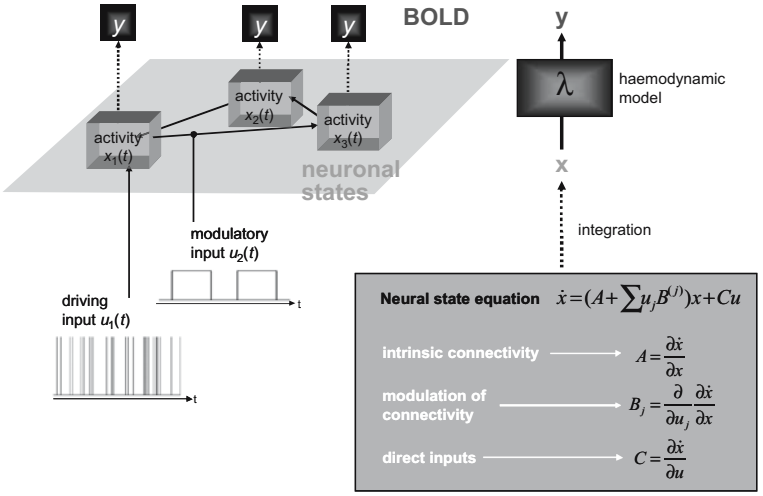

Fig. 4. Schematic summary of DCM for fMRI. The dynamics in a system of interacting neuronal populations (left panel), which are not directly observable by fMRI, are modeled using a bilinear state equation (right panel). Integrating the state equation gives predicted neural dynamics (x) which are transformed into predicted BOLD responses (y) by means of a haemodynamic forward model (λ). Neural and haemodynamic parameters are adjusted jointly such that the differences between predicted and measured BOLD series are minimized. The neural dynamics are determined by experimental manipulations that enter the model in the form of external inputs. Driving inputs (u_1; e.g. sensory stimuli) elicit local responses which are propagated through the system according to the intrinsic connections. The strengths of these connections can be changed by modulatory inputs (u_2; e.g. changes in task, attention, or due to learning). Note that in this figure the structure of the system and the scaling of the inputs have been chosen arbitrarily

7.2 DCM for Event-Related Potentials (ERPs)

ERPs as measured with EEG or MEG have been used for decades to study electrophysiological correlates of cognitive operations. Nevertheless, the neurobiological mechanisms that underlie their generation are still largely unknown. DCM for ERPs was developed as a biologically plausible model to understand how event-related responses result from the dynamics in coupled neural ensembles (David et al. 2006).

DCM for ERPs rests on a neural mass model, developed by David & Friston (2003) as an extension of the model by Jansen & Rit (1995), which uses established connectivity rules in hierarchical sensory systems (Felleman & Van Essen 1992) to assemble a network of coupled cortical sources. These rules characterize connections with respect to their laminar patterns of origin and termination and distinguish between (i) forward (or bottom-up) connections originating in agranular layers and terminating in layer 4, (ii) backward (or top-down) connections originating and terminating in agranular layers,

and (iii) lateral connections originating in agranular layers and targeting all layers. These long-range (extrinsic or inter-areal) cortico-cortical connections are excitatory, using glutamate as neurotransmitter, and arise from pyramidal cells.

Each region or source is modeled as a microcircuit following the model by David & Friston (2003). Three neuronal subpopulations are combined in this circuit and assigned to granular and supra-/infragranular layers. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons via intrinsic (intra-areal) connections. Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4 and in receipt of forward connections. Excitatory pyramidal cells and inhibitory interneurons are considered to occupy infra- and supragranular layers and receive backward and lateral inputs (see Fig. 5).

The neural state equations are summarized in Fig. 5. To perturb the system and model event-related responses, the network receives inputs via input connections. These connections are exactly the same as forward connections and deliver input u to the spiny stellate cells in layer 4. Input u represents afferent activity relayed by subcortical structures and are modelled as two parameterized components, a gamma density function (representing an event-related burst of input that is delayed and dispersed by subcortical synapses

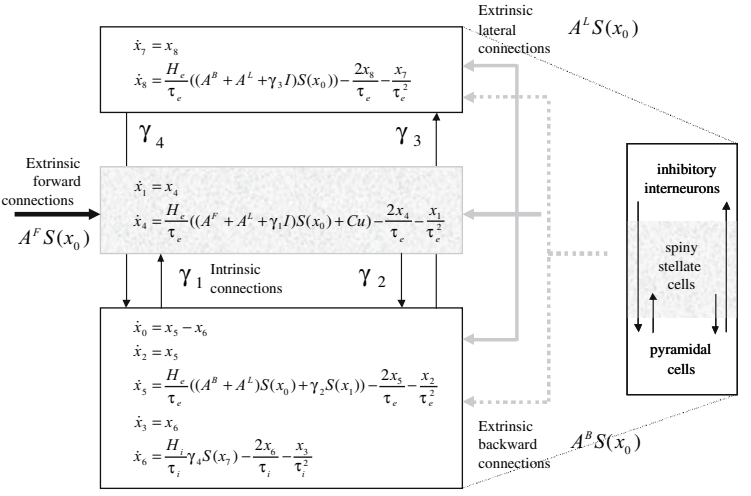

Fig. 5. Schematic of the neural model in DCM for ERPs. This schema shows the state equations describing the dynamics of a microcircuit representing an individual region (source). Each region contains three subpopulations (pyramidal, spiny stellate and inhibitory interneurons) that are linked by intrinsic connections and have been assigned to supragranular, granular and infragranular cortical layers. Different regions are coupled through extrinsic (long-range) excitatory connections that follow the laminar patterns of forward, backward and lateral connections, respectively

and axonal conduction) and a discrete cosine set (representing fluctuations in input over peristimulus time). The influence of this input on each source is controlled by a parameter vector C (see David et al. 2006 for details). Overall, the DCM is specified in terms of the state equations shown in Fig. 5 and a linear output equation

$$\begin{aligned} \frac{dx}{dt} &= f(x, u, \theta) \\ y &= Lx_0 + \varepsilon \end{aligned} \tag{15}$$

where x_0 represents the transmembrane potential of pyramidal cells and L is a lead field matrix coupling electrical sources to the EEG channels (Kiebel et al. 2006). In comparison to DCM for fMRI, the forward model is a simple linearity as opposed to the nonlinear haemodynamic model in DCM for fMRI. In contrast, the state equations of DCM for ERPs are much more complex and realistic (cf. Fig. 5). As an example, the state equation for the inhibitory subpopulation is

$$\begin{aligned} \frac{dx_7}{dt} &= x_8 \\ \frac{dx_8}{dt} &= \frac{H_e}{\tau_e}((A^B + A^L + \gamma_3 I)S(x_0)) - \frac{2x_8}{\tau_e} - \frac{x_7}{\tau_e^2} \end{aligned} \tag{16}$$

The parameter matrices A^F, A^B, A^L encode forward, backward and lateral connections respectively. Within each subpopulation, the dynamics of neural states are determined by two operators. The first transforms the average density of presynaptic inputs into the average postsynaptic membrane potential. This is modeled by a linear transformation with excitatory (e) and inhibitory (i) kernels parameterized by $H_{e,i}$ and $\tau_{e,i}$. $H_{e,i}$ control the maximum postsynaptic potential and $\tau_{e,i}$ represent lumped rate constants (i.e. lumped across dendritic spines and the dendritic tree). The second operator S transforms the average potential of each subpopulation into an average firing rate. This is assumed to be instantaneous and is a sigmoid function. Intra-areal interactions among the subpopulations depend on constants $\gamma_{1\ldots4}$ which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation. In (16), the top line expresses the rate of change of voltage as a function of current. The second line specifies how current changes as a function of voltage, current and presynaptic input from extrinsic and intrinsic sources. For simplification, our description here has omitted the fact that in DCM for ERPs all intra- and inter-areal connections have conduction delays. This requires the use of delay differential equations (see David et al. 2006 for details).

For estimating the parameters from empirical data, a fully Bayesian approach is used that is analogous to that used in DCM for fMRI and is described in detail by David et al. (2006). The posterior distributions of the parameter estimates can be used to test hypotheses about the modeled

processes, particularly differences in inter-areal connection strengths between different trial types. As in DCM for fMRI, Bayesian model selection can be used to optimize model structure or compare competing scientific hypotheses (Penny et al. 2004b).

8 Application of System Models in Functional Neuroimaging: Present and Future

Models of functional integration, which were originally developed for electrophysiological data from multi-unit recordings (Gerstein and Perkel 1968), are now taking an increasingly prominent role in functional neuroimaging. This is because the emphasis of the scientific questions in cognitive neuroscience is shifting from *where* particular processes are happening in the brain to *how* these processes are implemented. With increasing use, a word of caution may be appropriate here: Models of effective connectivity are not very useful without precise a priori hypotheses about specific mechanisms expressed at the level of inter-regional coupling. Simply describing *patterns* of connectivity that require post hoc interpretation does not lead to a mechanistic understanding of the system of interest. What is needed are parsimonious, well-motivated models that test precise hypotheses about *mechanisms*, either in terms of changes in particular connection strengths as a function of experimental condition, time (learning) or drug, or in terms of comparing alternative explanations by model selection (for examples, see Büchel & Friston 1997; Büchel et al. 1999; Honey et al. 2003; McIntosh et al. 1994, 1998; Rowe et al. 2002; Stephan et al. 2003, 2005; Toni et al. 2002). Figure 6 shows an example of such a model (Friston et al. 2003) where the parameters are mechanistically meaningful.

This search for mechanisms seems particularly promising for pharmacological questions. Since many drugs used in psychiatry and neurology change synaptic transmission and thus functional coupling between neurons, their therapeutic effects cannot be fully understood without models of drug-induced connectivity changes in particular neural systems. So far, only relatively few studies have studied pharmacologically induced changes in connectivity, ranging from simple analyses of functional connectivity (e.g. Stephan et al. 2001a) to proper system models (e.g. Honey et al. 2003). As highlighted in a recent review by Honey and Bullmore (2004), an exciting possibility for the future is to use system models at the early stage of drug development to screen for substances that induce desired changes of connectivity in neural systems of interest with a reasonably well understood physiology. The success of this approach will partially depend on developing models that include additional levels of biological detail (e.g. effects of different neurotransmitters and receptor types) while being parsimonious enough to ensure mathematical identifiability and physiological interpretability; see Breakspear et al. (2003),

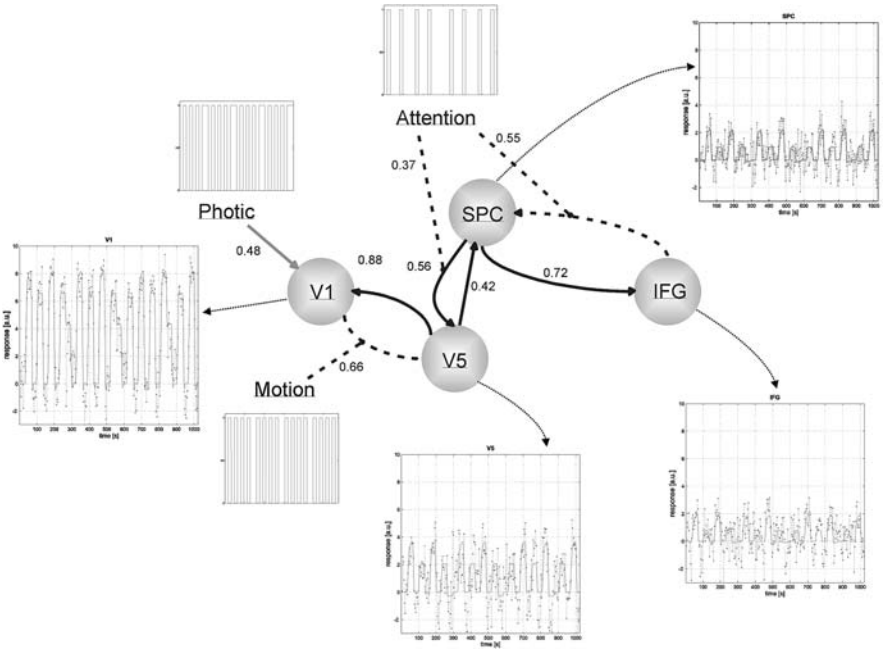

Fig. 6. DCM analysis of a single subject fMRI data from a study of attention to visual motion in which subjects viewed identical stimuli (radially moving dots) under different levels of attention to the stimuli (Büchel & Friston 1997). The model was introduced and described in detail by Friston et al. (2003). The figure is reproduced (with permission from Elsevier Ltd.) from Stephan et al. (2004). Only those conditional estimates are shown alongside their connections for which there was at least 90% confidence that they corresponded to neural transients with a half life shorter than 4 seconds. The temporal structure of the inputs is shown by box-car plots. Dashed arrows connecting regions represent significant bilinear affects in the absence of a significant intrinsic coupling. Fitted responses based upon the conditional estimates and the adjusted data are shown in the panels connected to the areas by dotted lines. The important parameters here are the bilinear ones. Note that while the intrinsic connectivity between areas V1 and V5 is non-significant and basically zero, motion stimuli drastically increase the strength of this connection, "gating" V1 input to V5. Top-down effects of attention are represented by the modulation of backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC) and from SPC to V5. See Penny et al. (2004b) and Stephan (2004) for a discussion how different neurophysiological mechanisms can be modeled with DCM

Harrison et al. (2005), Jirsa (2004) and Robinson et al. (2001) for examples that move in this direction.

Another important goal is to explore the utility of models of effective connectivity as diagnostic tools (Stephan 2004). This seems particularly attractive for psychiatric diseases whose phenotypes are often very heterogeneous and

where a lack of focal brain pathologies points to abnormal connectivity (dysconnectivity) as the cause of the illness. Given a pathophysiological theory of a specific disease, connectivity models might allow one to define an *endophenotype* of that disease, i.e. a biological marker at intermediate levels between genome and behaviour, which enables a more precise and physiologically motivated categorization of patients (Gottesman & Gould 2003). Such an approach has received particular attention in the field of schizophrenia research where a recent focus has been on abnormal synaptic plasticity leading to dysconnectivity in neural systems concerned with emotional and perceptual learning (Friston 1998; Stephan et al. 2006). A major challenge will be to establish neural systems models which are sensitive enough that their connectivity parameters can be used reliably for diagnostic classification and treatment response prediction of individual patients. Ideally, such models should be used in conjunction with paradigms that are minimally dependent on patient compliance and are not confounded by factors like attention or performance. Given established validity and sufficient sensitivity and specificity of such a model, one could use it in analogy to biochemical tests in internal medicine, i.e. to compare a particular model parameter (or combinations thereof) against a reference distribution derived from a healthy population (Stephan et al. 2006). Such procedures could help to decompose current psychiatric entities like schizophrenia into more well-defined subgroups characterized by common pathophysiological mechanisms and may facilitate the search for genetic underpinnings.

Acknowledgements

The authors are supported by the Wellcome Trust. We would like to thank our colleagues for stimulating discussions on modeling, particularly Lee Harrison, Will Penny, Olivier David, Stefan Kiebel, Randy McIntosh and Michael Breakspear.

References

Aertsen A, Preißl H (1991) Dynamics of activity and connectivity in physiological neuronal Networks. In: Schuster HG (ed.) Non Linear Dynamics and Neuronal Networks, VCH Publishers, New York, pp. 281–302

Ashby WR (1956) An introduction to cybernetics. Chapman & Hall, London

Bar-Yam Y (1997) Dynamics of complex systems. Addison-Wesley, Reading

Bollen KA (1989) Structural equations with latent variables. John Wiley, New York

Bossel H (1992) Modellbildung und Simulation. Vieweg, Braunschweig

Breakspear M, Terry JR, Friston KJ (2003) Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a biophysical model of neuronal dynamics. Network: Computation in Neural Systems 14, 703–732

Büchel C, Friston KJ (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cerebral Cortex 7, 768–778

Büchel C, Coull JT, Friston KJ (1999) The Predictive Value of Changes in Effective Connectivity for Human Learning. Science 283, 1538–1541

Bullmore ET, Horwitz B, Honey G, Brammer M, Williams S, Sharma T (2000) How good is good enough in path analysis of fMRI data? NeuroImage 11, 289–301

Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magnetic Resonance Medicine 39, 855–864

Chong L, Ray LB (2002) Whole-istic biology. Science 295, 1661

David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743–1755

David O, Kiebel SJ, Harrison LM, Mattout J, Kilner JM, Friston KJ (2006) Dynamic causal modeling of evoked responses in EEG and MEG. NeuroImage 30, 1255–1272

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex 1, 1–47

Friston KJ (1994) Functional and effective connectivity in neuroimaging: a synthesis. Human Brain Mapping 2, 56–78

Friston KJ, Büchel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218–229

Friston KJ (1998) The disconnection hypothesis. Schizophrenia Research 30, 115–125

Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477

Friston KJ (2002) Beyond phrenology: What can neuroimaging tell us abut distributed circuitry? Annual Reviews in Neuroscience 25, 221–250

Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. NeuroImage 19, 1273–1302

Friston KJ, Büchel C (2004) Functional connectivity: eigenimages and multivariate analyses. In: Frackowiack R et al. (ed.) Human Brain Function, 2nd edition, Elsevier, New York, pp. 999–1018

Gerstein GL, Perkel DH (1969) Simultaneously recorded trains of action potentials: Analysis and functional interpretation. Science 164, 828–830

Goebel R, Roebroeck A, Kim DS, Formisano E (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magnetic Resonance Imaging 21, 1251–1261

Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. American Journal of Psychiatry 160, 636–645

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438

Harrison LM, Penny W, Friston KJ (2003) Multivariate autoregressive modeling of fMRI time series. NeuroImage 19, 1477–1491

Harrison LM, David O, Friston KJ (2005) Stochastic models of neuronal dynamics. Philosophical Transactions of the Royal Society London B Biological Sciences 360, 1075–1091

Honey GD, Fu CHY, Kim J, Brammer MJ, Croudace TJ, Suckling J, Pich EM, Williams SCR, Bullmore ET (2002) Effects of verbal working memory load on corticocortical connectivity modeled by path analysis of functional magnetic resonance imaging data. NeuroImage 17, 573–582

Honey GD, Suckling J, Zelaya F, Long C, Routledge C, Jackson S, Ng V, Fletcher PC, Williams SCR, BrownJ, Bullmore ET (2003) Dopaminergic drug effects on physiological connectivity in a human cortico-striato-thalamic system. Brain 126, 1767–1281

Honey G, Bullmore E (2004) Human pharmacological MRI. Trends in Pharmacological Sciences 25, 366–374

Horwitz B, Rumsey JM & Donohue BC (1998) Functional connectivity of the angular gyrus in normal reading and dyslexia. Proceedings of the National Academy of Sciences USA 95, 8939–8944

Jirsa VK (2004) Connectivity and dynamics of neural information processing. Neuroinformatics 2, 183–204

Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead-field parameterization. NeuroImage 30, 1273–1284

Kötter R (2004) Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. Neuroinformatics 2, 127–144

Magee JC, Johnston D (2005) Plasticity of dendritic function. Current Opinion in Neurobiology 15, 334–342

McIntosh AR, Gonzalez-Lima F (1991) Metabolic activation of the rat visual system by patterned light and footshock. Brain Research 547, 295–302

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping 2, 2–22

McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1994) Network analysis of cortical visual pathways mapped with PET. Journal of Neuroscience 14, 655–666

McIntosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. Journal of Neurophysiology 80, 2790–2796

McIntosh AR (2000) Towards a network theory of cognition. Neural Networks 13, 861–870

McIntosh AR, Rajah MN, Lobaugh NJ (2003) Functional connectivity of the medial temporal lobe relates to learning and awareness. Journal of Neuroscience 23, 6520–6528

McIntosh AR, Lobaugh NJ (2004) Partial least squares analysis of neuroimaging data: applications and advances. NeuroImage 23, S250–S263

Payne BR, Lomber SG (2001) Reconstructing functional systems after lesions of cerebral cortex. Nature Reviews Neuroscience 2, 911–919

Penny WD, Roberts SJ (2002) Bayesian multivariate autoregressive models with structured priors. IEE Proceedings of Vision and Image Signal Processing 149, 33–41

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004a) Modeling functional integration: a comparison of structural equation and dynamic causal models. NeuroImage 23, S264–274

Penny WD, Stephan KE, Mechelli A, Friston KJ (2004b) Comparing dynamic causal models. NeuroImage 22, 1157–1172

Robinson PA, Rennie CJ, Wright JJ, Bahramali H, Gordon E, Rowe DL (2001) Prediction of electroencephalographic spectra from neurophysiology. Physical Reviews E 63, 021903

Rowe JB, Stephan KE, Friston KJ, Frackowiak RJ, Lees A, Passingham RE (2002) Attention to action in Parkinson's disease. Impaired effective connectivity among frontal cortical regions. Brain 125, 276–289

Stephan KE, Magnotta VA, White TJ, Arndt S, Flaum M, O'Leary DS & Andreasen, NC (2001a) Effects of Olanzapine on cerebellar functional connectivity in schizophrenia measured by fMRI during a simple motor task. Psychological Medicine 31, 1065–1078

Stephan KE, Kamper L, Bozkurt A, Burns GAPC, Young MP, Kötter R (2001b) Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). Philosophical Transactions of the Royal Society London B Biological Sciences 356, 1159–1186

Stephan KE, Marshall JC, Friston KJ, Rowe JB, Ritzl A, Zilles K, Fink GR (2003) Lateralized cognitive processes and lateralized task control in the human brain. Science 301, 384–386

Stephan KE (2004) On the role of general system theory for functional neuroimaging. Journal of Anatomy 205, 443–470

Stephan KE, Harrison LM, Penny WD, Friston KJ (2004) Biophysical models of fMRI responses. Current Opinion in Neurobiology 14, 629–635

Stephan KE, Penny WD, Marshall JC, Fink GR, Friston KJ (2005) Investigating the functional role of callosal connections with dynamic causal models. Annals of the New York Academy of Sciences 1064:16–36

Stephan KE, Baldeweg T, Friston KJ (2006) Synaptic plasticity and dysconnection in schizophrenia. Biological Psychiatry 59, 929–939

Toni I, Rowe JB, Stephan KE & Passingham RE (2002) Changes of cortico-striatal effective connectivity during visuo-motor learning. Cerebral Cortex 12: 1040–1047

von Bertalanffy L (1969) General System Theory. George Braziller, New York

Wiener N (1948) Cybernetics. Wiley, New York

Yamashita O, Galka A, Ozaki T, Biscay R, Valdes-Sosa P (2004) Recursive penalized least squares solution for dynamical inverse problems of EEG generation. Human Brain Mapping 21, 221–235