

Dynamical systems

K. Friston

INTRODUCTION

This appendix reviews models of dynamical systems. The review is framed in terms of analyses of functional and effective connectivity, in which we focus on the nature and form of the models and less on estimation or inference issues. The aim is to relate various models and to make their underlying assumptions transparent.

As we have seen in the preceding chapters, there are a number of models for estimating effective connectivity using neuroimaging time-series. By definition, effective connectivity depends on a causal model, through which it is defined operationally (Friston, 1995). This appendix reviews the principal models that could be adopted and how they relate to each other. We consider dynamic causal models (DCM), generalized convolution models (GCM), coherence analyses, structural equation models (SEM), state-space models (SSM) and multivariate autoregression-moving average (ARMA) models. In brief, we will show that they are all special cases of each other and try to emphasize their points of contact. However, some fundamental distinctions arise that guide the selection of the appropriate models in different situations. We now review these distinctions.

Coupling among inputs, outputs or hidden states?

The first distinction rests upon whether the model is used to explain the coupling between the inputs and outputs, among different outputs or among the system's states (e.g. neuronal activity in different ensembles). In terms of models, this distinction is between input-output models, e.g. multiple-input-single-output models (MISO) or multiple-input-multiple-output models (MIMO) and

explicit input-state-output models. Usually, the input-output approach is concerned with the non-linear transformation of inputs, enacted by a system, to produce its outputs. This is like trying to establish a statistical dependence of the outputs on the inputs, without any comment on the mechanisms mediating this dependency. In some instances (e.g. ARMA and coherence analyses), dependences among different outputs are characterized (cf. functional connectivity).

Conversely, the input-state-output approach is generally concerned with characterizing the coupling among hidden variables that represent the states of the system. These states are observed vicariously through the outputs (Figure A2.1). Inferring the coupling among states induces the need for a causal model of how states affect each other and form outputs (cf. effective connectivity). Examples of input-output models include the Volterra formulation and generalized coherence analyses in the spectral domain. An example of a model that tries to estimate coupling among hidden states is DCM. In short, input-output models of coupling can proceed without reference to the hidden states. Conversely, interactions among hidden states require indirect access to the states through some model of the causal architecture of the system. In the next section, we start by reviewing input-output models and then turn to input-state-output models.

Deterministic or stochastic inputs?

The second key distinction is when the input is known (e.g. DCM) and when it is not (e.g. ARMA and SEM). This distinction depends on whether the inputs enter as known and/or deterministic quantities (e.g. experimentally designed causes of evoked responses) or whether we know (or can assume) something about the statistics

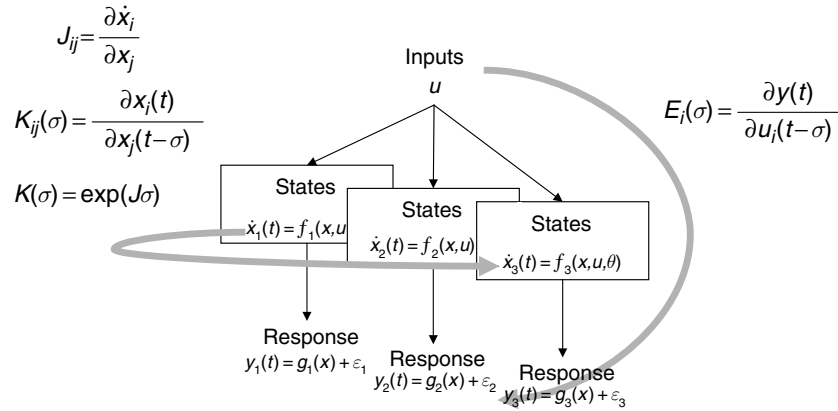


FIGURE A2.1 Schematic depicting the difference between analyses that address input-output behaviours and those that refer explicitly to interactions among coupled states.

of the input (i.e. its statistics up to second or higher orders). Most models of the stochastic variety assume the inputs are Gaussian, IID (independent and identically distributed) and stationary. Some stochastic models (e.g. coherence) use local stationarity assumptions to estimate high-order moments from observable but noisy inputs. For example, polyspectral analysis represents an intermediate case in which the inputs are observed but only their statistics are used.

Stationarity assumptions in stochastic models are critical because they preclude full analyses of evoked neuronal responses or transients that, by their nature, are non-stationary. On the other hand, there are situations where the input is not observable or under experimental control. In these cases, approaches like ARMA and SEM can be used if the inputs can be regarded as stationary. The distinction between deterministic and stochastic inputs is critical in the sense that it would be inappropriate to adopt one class of model in a context that calls for the other.

Connections or dependencies?

The final distinction is in terms of what is being estimated or inferred. Recall that functional connectivity is defined by the presence of statistical dependences among remote neurophysiological measurements. Conversely, effective connectivity is a parameter of a model that specifies the causal influences among states. It is useful to distinguish *inferences* about statistical dependencies and *estimation* of effective connectivity in terms of the distinction between functional and effective connectivity. Examples of approaches that try to establish statistical dependences include coherence analyses and ARMA. This is because these techniques do not presume a model of how hidden states interact to produce responses. They

are interested only in establishing dependences among outputs over different frequencies or time lags. Although ARMA may employ some model to assess dependences, this is a model of dependences among outputs. There is no assertion that outputs *cause* outputs. Conversely, SEM and DCM try to estimate the model parameters and constitute analyses of effective connectivity proper.

EFFECTIVE CONNECTIVITY

Effective connectivity is the influence that one system exerts over another at a unit or ensemble level. This should be contrasted with functional connectivity, which implies a statistical dependence between two systems that could be mediated in any number of ways. Operationally, effective connectivity can be expressed as the response induced in an ensemble, unit or node by input from others, in terms of partial derivatives of the target activity x_i , with respect to the source activities. First- and second-order connections are then:

$$K_{ij}(\sigma_1) = \frac{\partial x_i(t)}{\partial x_j(t - \sigma_1)}$$

$$K_{ijk}(\sigma_1, \sigma_2) = \frac{\partial^2 x_i(t)}{\partial x_j(t - \sigma_1) \partial x_k(t - \sigma_2)}, \dots \quad \text{A2.1}$$

First-order connectivity embodies the response evoked by a change in input at $t - \sigma_1$. In other words, it is a time-dependent measure of *driving* efficacy. Second-order connectivity reflects the *modulatory* influence of the input at $t - \sigma_1$ on the response evoked at $t - \sigma_2$. And so on for higher orders. Note that, in this general formulation, effective connectivity is a function of inputs over

the recent past.¹ Furthermore, implicit in Eqn. A2.1 is the fact that effective connectivity is causal, unless σ_1 is allowed to be negative. It is useful to introduce the dynamic equivalent, in which the response of the target is expressed in terms of *changes in activity*:

$$J_{ij} = \frac{\partial \dot{x}_i}{\partial x_j} \quad \frac{\partial J_{ij}}{\partial x_k} = \frac{\partial^2 \dot{x}_i}{\partial x_j \partial x_k} \dots \quad \text{A2.2}$$

where $\dot{x}_i = \partial x_i / \partial t$. In this dynamic form, all influences are causal and instantaneous. In this appendix, we will call the $K(\sigma)$ effective connections and J coupling. We will see later that they are related by $K(\sigma) = \exp(J\sigma)$ and that effective connectivity can be formulated in terms of Volterra kernels. Before considering specific models of effective connectivity, we will review briefly their basis (see also Chapter 38).

Dynamical systems

A plausible model of neuronal systems is a non-linear dynamical model that corresponds to an analytic multiple-input-multiple-output (MIMO) system. The state and output equations of a analytic dynamical system are:

$$\begin{aligned} \dot{x}(t) &= f(x, u, \theta) \\ y(t) &= g(x) + \varepsilon \end{aligned} \quad \text{A2.3}$$

Typically the inputs $u(t)$ correspond to designed experimental effects (e.g. stimulus functions in functional magnetic resonance imaging, fMRI), or represent stochastic fluctuations or system perturbations. Stochastic observation error $\varepsilon \sim N(0, \Sigma)$ enters linearly in this model. For simplicity, the expressions below deal with single-input-single-output (SISO) systems, and will be generalized later. The measured response y is some non-linear function of the states of the system x . These state variables are usually unobserved or hidden (e.g. the configuration of all ion channels, the depolarization of every dendritic compartment etc.). The parameters of the state equation embody effective connectivity, either in terms of mediating the coupling between inputs and outputs (MISO models of a single region) or through the coupling among state variables (MIMO models of multiple regions). The objective is to estimate and make inferences (usually Bayesian) about these parameters, given the outputs and possibly the inputs.

¹ In contrast, functional connectivity is model-free and simply reflects the mutual information $I(x_i, x_j)$. In this appendix we are concerned only with models of effective connectivity.

Sometimes this requires one to specify the form of the state equation. A ubiquitous and useful form is the bilinear approximation; expanding around x_0 :

$$\begin{aligned} \dot{x}(t) &\approx Ax + uBx + Cu \\ y &= Lx \\ A &= \frac{\partial f}{\partial x}, \quad B = \frac{\partial^2 f}{\partial x \partial u}, \quad C = \frac{\partial f}{\partial u}, \quad L = \frac{\partial g}{\partial x} \end{aligned} \quad \text{A2.4}$$

For simplicity, we have assumed $x_0 = 0$ and $f(0) = g(0) = 0$. This bilinear model is sometimes expressed in a more compact form by augmenting the states with a constant:

$$\begin{aligned} \dot{X} &= (M + uN)X \\ y &= HX \\ X &= \begin{bmatrix} 1 \\ x \end{bmatrix} \quad M = \begin{bmatrix} 0 & 0 \\ f & A \end{bmatrix} \quad N = \begin{bmatrix} 0 & 0 \\ C & B \end{bmatrix} \quad H = [g \ L] \end{aligned} \quad \text{A2.5}$$

(see Friston, 2002). Here the coupling parameters comprise the matrices $\theta = A, B, C, L$. We will use the bilinear parameterization when dealing with MIMO models and their derivatives below. We will first deal with MISO models, with and without deterministic inputs.

INPUT-OUTPUT MODELS

Models for deterministic inputs – The Volterra formulation

In this section, we review the Volterra formulation of dynamical systems. This formulation is important because it allows the input-output behaviour of a system to be characterized in terms of kernels that can be estimated without knowing the states of the system.

The Fliess fundamental formula (Fliess *et al.*, 1983) describes the causal relationship between the outputs and the history of the inputs. This relationship conforms to a Volterra series which expresses the output as a generalized convolution of the input, critically without reference to the states. This series is simply a functional Taylor expansion of the outputs with respect to the inputs (Bendat, 1990). The reason it is a *functional* expansion is that the inputs are a function of time:

$$\begin{aligned} y(t) &= h(u, \theta) + \varepsilon \\ h(u, \theta) &= \sum_i \int_0^t \dots \int_0^t \kappa_i(\sigma_1, \dots, \sigma_i) u(t - \sigma_1), \dots, \\ &\quad u(t - \sigma_i) d\sigma_1, \dots, d\sigma_i \\ \kappa_i(\sigma_1, \dots, \sigma_i) &= \frac{\partial^i y(t)}{\partial u(t - \sigma_1), \dots, \partial u(t - \sigma_i)} \end{aligned} \quad \text{A2.6}$$

where $\kappa_i(\sigma_1, \dots, \sigma_i)$ is the i -th order kernel. In Eqn. **A2.6**, the integrals are over the past or history of the inputs. This renders the model causal. In some situations an acausal model may be appropriate (e.g. in which the kernels have non-zero values for future inputs; see Friston and Büchel, 2000). One important thing about the Volterra expansion is that it is linear in the unknowns, enabling relatively simple unbiased estimates of the kernels. In other words, Eqn. **A2.6** can be treated as a general linear observation model enabling all the usual estimation and inference procedures (see Chapter 38 for an example). Volterra series are generally thought of as a high-order or generalized non-linear convolution of the inputs to provide an output. To ensure the kernels can be estimated efficiently, they can be expanded in terms of some appropriate basis functions $q_j^i(\sigma_1, \dots, \sigma_i)$ to give the general linear model:

$$\begin{aligned} y(t) &= \sum_{ij} \beta_j^i h_j^i(u) + \varepsilon \\ h_j^i(u) &= \int_0^t \dots \int_0^t q_j^i(\sigma_1, \dots, \sigma_i) u(t - \sigma_1), \dots, \\ &\quad u(t - \sigma_i) d\sigma_1, \dots, d\sigma_i \\ \kappa_i(\sigma_1, \dots, \sigma_i) &= \sum_j \beta_j^i q_j^i(\sigma_1, \dots, \sigma_i) \end{aligned} \quad \text{A2.7}$$

The Volterra formulation is useful as a way of characterizing the influence of inputs on the responses of a region. The kernels can be regarded as a re-parameterization of the bilinear form in Eqn. **A2.4** that encodes the impulse response to input. The kernels for the states are:

$$\begin{aligned} \kappa_0 &= X(0) \\ \kappa_1(\sigma_1) &= e^{\sigma_1 M} N e^{-\sigma_1 M} X(0) \\ \kappa_2(\sigma_1, \sigma_2) &= e^{\sigma_2 M} N e^{(\sigma_1 - \sigma_2) M} N e^{-\sigma_1 M} X(0) \\ \kappa_2(\sigma_1, \sigma_2, \sigma_3) &= \dots \end{aligned} \quad \text{A2.8}$$

The kernels associated with the output follow from the chain rule:

$$\begin{aligned} h_0 &= H \kappa_0 \\ h_1(\sigma_1) &= H \kappa_1(\sigma_1) \\ h_2(\sigma_1, \sigma_2) &= H \kappa_2(\sigma_1, \sigma_2) + \kappa_1(\sigma_1)^T \partial H / \partial X \kappa_1(\sigma_2) \\ h_2(\sigma_1, \sigma_2, \sigma_3) &= \dots \end{aligned} \quad \text{A2.9}$$

(see Friston, 2002 for details). If the system is fully non-linear, then the kernels can be considered local approximations. If the system is bilinear they are globally exact. It is important to remember that the estimation of the kernels does not assume any form for the state

equation and completely eschews the states. This is the power and weakness of Volterra-based analyses.

The Volterra formulation can be used directly in the assessment of effective connectivity if we assume the measured response of one region constitutes the input to another, i.e. $u_i(x) = y_j(t)$. In this case, the Volterra kernels have a special interpretation; they are synonymous with effective connectivity. From Eqn. **A2.6**, the first-order kernels are:

$$\kappa_1(\sigma_1)_{ij} = \frac{\partial y_i(t)}{\partial y_j(t - \sigma_1)} = K_{ij}(\sigma_1) \quad \text{A2.10}$$

Extensions to multiple inputs (MISO) models are trivial and allow for high-order interactions among inputs to a single region to be characterized. This approach was used in Friston and Büchel (2000) to examine parietal modulation of V2 inputs to V5, by making inferences about the appropriate second-order kernel. The advantage of the Volterra approach is that non-linearities can be modelled and estimated in the context of highly non-linear transformations within a region and yet model inversion proceeds in a standard linear setting. However, one has to assume that the inputs conform to measured responses elsewhere in the brain. This may be tenable for some electrophysiological data, but the haemodynamic responses measured by fMRI make this a more questionable approach. Furthermore, there is no causal model of the interactions among areas that would otherwise offer useful constraints on the inversion. The direct application of Volterra estimation, in this fashion, simply examines each node, one at a time, assuming the activities of other nodes are veridical measurements of the inputs to the node in question. In summary, although the Volterra kernels are useful characterizations of the input-output behaviour of single nodes, they are not constrained by any model of interactions among regions. Before turning to DCMs that embody these interactions, we will deal with the SISO situation in which the input is treated as stochastic.

Models for stochastic inputs – coherence and polyspectral analysis

In this section, we deal with systems in which the input is stochastic. The aim is to estimate the kernels (or their spectral equivalents) given only statistics about the joint distribution of the inputs and outputs. When the inputs are unknown, one generally makes an assumption about their distributional properties and assumes [local] stationariness. Alternatively, the inputs may be measurable but too noisy to serve as inputs in a Volterra expansion.

In this case, they can be used to estimate the input and output densities in terms of high-order cumulants or polyspectral density. The n -th order cumulant of the input is:

$$c_u\{\sigma_1, \dots, \sigma_{n-1}\} = \langle u(t)u(t-\sigma_1), \dots, u(t-\sigma_{n-1}) \rangle \quad \text{A2.11}$$

where we have assumed here, and throughout, that the expectation $E(u(t)) = 0$. It can be seen that cumulants are a generalization of autocovariance functions. The second-order cumulant is simply the autocovariance function of lag and summarizes the stationary second-order behaviour of the input. Cumulants allow one to formulate the Volterra expansion in terms of the second-order statistics of input and outputs. For example:

$$\begin{aligned} c_{yu}\{\sigma_a\} &= \langle y(t)u(t-\sigma_a) \rangle \\ &= \sum_i \int_0^t \dots \int_0^t \kappa_i(\sigma_1, \dots, \sigma_i) \langle u(t-\sigma_a)u(t-\sigma_1) \dots \\ &\quad u(t-\sigma_i) \rangle d\sigma_1 \dots d\sigma_i \\ &= \sum_i \int_0^t \dots \int_0^t \kappa_i(\sigma_1, \dots, \sigma_i) \times \\ &\quad c_u\{\sigma_a - \sigma_1, \dots, \sigma - \sigma_i\} d\sigma_1 \dots d\sigma_i \end{aligned} \quad \text{A2.12}$$

This equation says that the cross-covariance between the output and the input can be decomposed into components that are formed by convolving the i -th order kernel with the input's $i+1$ -th cumulant. The important thing about this is that all cumulants, greater than second order, of Gaussian processes are zero. This means that if we can assume the input is Gaussian then:

$$c_{yu}\{\sigma_a\} = \int_0^t \kappa_i(\sigma_1) c_u\{\sigma_a - \sigma_1\} d\sigma_1 \quad \text{A2.13}$$

In other words, the cross-covariance between the input and output is simply the autocovariance function of the inputs convolved with the first-order kernel. Although it is possible to formulate the covariance between inputs and outputs in terms of cumulants, the more conventional formulation is in frequency space using polyspectra. The n -th polyspectrum is the Fourier transform of the corresponding cumulant:

$$\begin{aligned} g_u(\omega_1, \dots, \omega_{n-1}) &= \left(\frac{1}{2\pi}\right)^{n-1} \int \dots \\ &\int c_u\{\sigma_1, \dots, \sigma_{n-1}\} e^{-j(\omega\sigma_1, \dots, \omega\sigma_{n-1})} d\sigma_1, \dots, d\sigma_{n-1} \end{aligned} \quad \text{A2.14}$$

Again, polyspectra are simply a generalization of spectral densities. For example, the second polyspectrum is

spectral density and the third polyspectrum is bi-spectral density. It can be seen that these relationships are generalizations of the Wiener-Khinchine theorem, relating the autocovariance function and spectral density through the Fourier transform. Introducing the spectral density representation:

$$u(t) = \int s_u(\omega) e^{-j\omega t} d\omega \quad \text{A2.15}$$

we can now rewrite the Volterra expansion as:

$$\begin{aligned} h(u, \theta) &= \sum_i \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} e^{j(\omega_1 + \dots + \omega_i)t} \\ &\times \Gamma_1(\omega_1, \dots, \omega_i) s_u(\omega_1), \dots, s_u(\omega_i) d\omega_1, \dots, d\omega_i \end{aligned} \quad \text{A2.16}$$

where the functions

$$\begin{aligned} \Gamma_1(\omega_1) &= \int_0^{\infty} e^{-j\omega_1 \sigma_1} \kappa_1(\sigma_1) d\sigma_1 \\ \Gamma_2(\omega_1, \omega_2) &= \int_0^{\infty} \int_0^{\infty} e^{-j(\omega_1 \sigma_1 + \omega_2 \sigma_2)} \kappa_2(\sigma_1, \sigma_2) d\sigma_1 d\sigma_2 \\ &\dots \end{aligned}$$

are the Fourier transforms of the kernels. These functions are called *generalized transfer functions* and mediate the expression of frequencies in the output given those in the input. Critically, the influence of higher order kernels, or equivalently generalized transfer functions means that a given frequency in the input can induce a *different* frequency in the output. A simple example of this would be squaring a sine wave input to produce an output of twice the frequency. In the Volterra approach, the kernels were identified in the time domain using the inputs and outputs directly. In this section, system identification means estimating their Fourier transforms (i.e. the transfer functions) using second and higher order statistics of the inputs and outputs. Generalized transfer functions are usually estimated through estimates of polyspectra. For example, the spectral form for Eqn. A2.13, and its high-order counterparts are:

$$\begin{aligned} g_{uy}(-\omega_1) &= \Gamma_1(\omega_1) g_u(\omega_1) \\ g_{uy}(-\omega_1, -\omega_2) &= 2\Gamma_2(\omega_1, \omega_2) g_u(\omega_1) g_u(\omega_2) \\ &\vdots \\ g_{u\dots y}(-\omega_1, \dots, -\omega_n) &= n! \Gamma_n(\omega_1, \dots, \omega_n) g_u(\omega_1) \dots g_u(\omega_n) \end{aligned} \quad \text{A2.17}$$

Given estimates of the requisite [cross]-polyspectra, these equalities can be used to provide estimates of the transfer

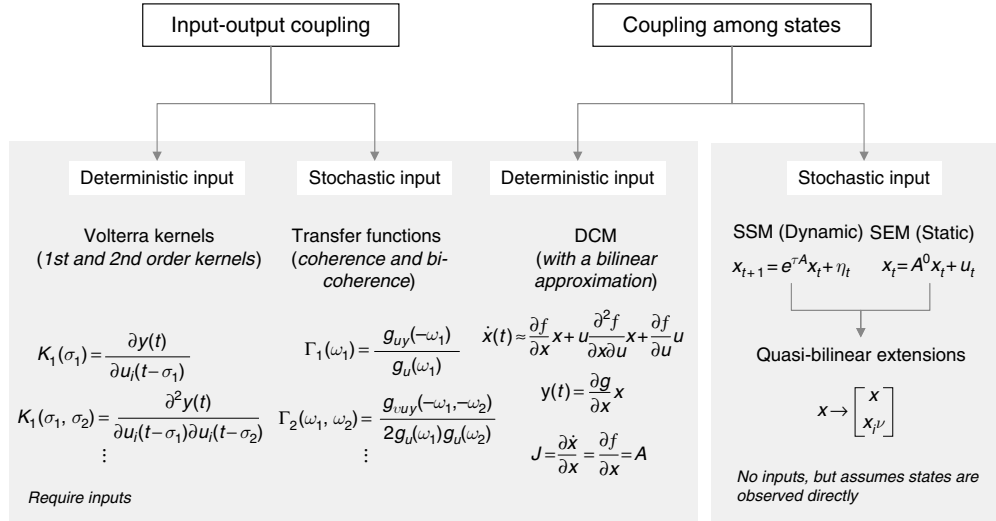


FIGURE A2.2 Overview of the models considered in this chapter. They have been organized to reflect whether they require known inputs or not and whether the model is a time-series model or not.

functions (see Figure A2.2). These equalities hold when the Volterra expansion contains just the n -th order term and are a generalization of the classical results for the transfer function of a linear system (i.e. the first equality in Eqn. A2.17). The importance of these results, in terms of effective connectivity, is the implicit meaning conferred on *coherence* and *bi-coherence* analyses. Coherence is simply the second-order cross spectrum $g_{uy}(\omega)$ between the input and output and is related to first-order effects (i.e. the first-order kernel or transfer function) through Eqn. A2.17. Coherence is therefore a surrogate for first-order or linear connectivity. Bi-coherence or the cross-bi-spectrum $g_{uy}(\omega_1, \omega_2)$ is the third-order cross-polyspectrum and implies a non-zero second-order kernel or transfer function. Bi-spectral analysis was used (in a simplified form) to demonstrate non-linear coupling between parietal and frontal regions using magnetoencephalography (MEG) in Chapter 39. In this example, cross-bi-spectra were estimated, in a simple fashion, using time-frequency analyses.

Summary

In summary, Volterra kernels (generalized transfer functions) characterize the input-output behaviour of a system. The n -th order kernel is equivalent to n -th order effective connectivity when the inputs and outputs conform to processes that mediate interactions among neuronal systems. If the inputs and outputs are known, or can be measured precisely, the estimation of the kernels is straightforward. In situations where inputs and outputs are observed less precisely, kernels can be

estimated indirectly through their generalized transfer functions using cross-polyspectra. The robustness of kernel estimation, conferred by expansion in terms of temporal basis functions, is recapitulated in the frequency domain by smoothness constraints during estimation of the polyspectra. The spectral approach is limited because it assumes the system contains only the kernel of the order estimated and stationariness. The intuition behind the first limitation relates to the distinction between parameter estimation and variance partitioning in standard regression analyses. Although it is perfectly possible to estimate the parameters of a regression model given a set of non-orthogonal explanatory variables, it is not possible uniquely to partition variance in the output caused by these explanatory variables.

INPUT-STATE-OUTPUT MODELS

In this section, we address models for multiple interconnected nodes (e.g. brain regions) where one can measure their responses to input that may or may not be known. Although it is possible to extend the techniques of the previous sections to cover MIMO systems, the ensuing inferences about the influence of input to one node on the response of another are not sufficiently specified to constitute an analysis of effective connectivity. This is because these influences may be mediated in many ways and are not parameterized in terms of the effective connectivity among the nodes themselves. A

parameterization that encodes this inter-node coupling is therefore required. All the models discussed below assume some form or model for the interactions among the state variables in one or more nodes and attempt to estimate the parameters of this model, sometimes without observing the states themselves.

Models for known inputs – dynamic causal modelling

The most direct and generic approach is to estimate the parameters of Eqn. A2.3 directly and, if necessary, use them to compute effective connectivity as described in Eqn. A2.1 and Eqn. A2.2. Although there are many forms one could adopt for Eqn. A2.3, we will focus on the bilinear approximation, which is possibly the most parsimonious but useful non-linear approximation available. Furthermore, as shown below, the bilinear approximation re-parameterizes the state equations of the model directly in terms of effective connectivity. Dynamic causal modelling does not necessarily entail the use of a bilinear model. Indeed, DCMs can be specified to any degree of biological complexity and realism supported by the data. There are examples in this book where the parameters of the state equation are already effective connectivity or coupling parameters, for example, the extrinsic and intrinsic connections in neural-mass models of event-related potentials (ERPs) (see Chapter 42). However, we focus on bilinear approximations here because they represent the simplest form to which all DCMs can be reduced. This reduction allows analytic derivation of kernels and other computations, like integrating the state equation, to proceed in an efficient fashion.

Each region may comprise several state variables whose causal interdependencies are summarized by the bilinear form in Eqn. A2.4. Here the coupling parameters of the state equation are the matrices M and N . For a given set of inputs or experimental context, the bilinear approximation to any state equation is:

$$\begin{aligned}\dot{X}(t) &= JX(t) \\ X(t + \sigma) &= e^{J\sigma} X(t) \\ J &= M + \sum_i N_i u_i\end{aligned}\tag{A2.18}$$

Notice that there are now as many N matrices as there are inputs. The bilinear form reduces the model to first-order connections that can be modulated by the inputs.

In MIMO models, the coupling is among the states such that first-order effective connections are simply:

$$\begin{aligned}J &= \frac{\partial \dot{X}}{\partial X} \\ K &= \frac{\partial X(t)}{\partial X(t - \sigma)} = e^{J\sigma}\end{aligned}\tag{A2.19}$$

Note that these are context-sensitive in the sense that the Jacobian J is a function of experimental context or inputs $u(t) = u_1(t), \dots, u_m(t)$. A useful way to think about the bilinear matrices is to regard them as the intrinsic or latent dynamic coupling, in the absence of input, and changes induced by each input (see Chapter 41 for a fuller description):

$$\begin{aligned}J(0) = M &= \begin{bmatrix} 0 & 0 \\ f(0) & A \end{bmatrix} \\ \frac{\partial J}{\partial u_i} = N_i &= \begin{bmatrix} 0 & 0 \\ C_i & B_i \end{bmatrix}\end{aligned}\tag{A2.20}$$

The latent coupling among the states is A . Often, one is more interested in B_i as embodying changes in this coupling induced by different cognitive set, time or drugs. Note that C_i enters as the input-dependent component of coupling to the constant term. Clearly, it would be possible to introduce other high-order terms to model interactions among the states, but we will restrict ourselves to bilinear models for simplicity.

Dynamic causal modelling has two parts: first, specification of the state and output equations of an ensemble of region-specific state variables. If necessary, a bilinear approximation to any state equation reduces the model to first-order coupling and bilinear terms that represent the modulation of that coupling by inputs. Second, inversion of the DCM allows one to estimate and make inferences about inter-regional connections and the effect of experimental manipulations on those connections.

As mentioned above, the state equations do not have to conform to the bilinear form. This is important because the priors may be specified more naturally in terms of the original biophysical parameters of the DCM, as opposed to the bilinear form. A nice example of this is the use of log-normal priors to enforce positivity constraints on the rate constants of ERP models in Chapter 42. Furthermore, the choice of the state variables clearly has to reflect their role in mediating the effect of inputs on responses and the interactions among areas. In the simplest case, the state variables could be reduced to mean synaptic activity per region, plus any biophysical state variables needed to determine the output (e.g. the states of haemodynamic models for fMRI). Implicit in choosing such state variables is the assumption that they model all the dynamics to the level of detail required. Mean field models and

neural-mass models may be useful here in motivating the number of state variables and the associated state equations (see Chapter 31). Operationally, issues of parameterization and number of state variables can be resolved with Bayesian model selection and is directed principally by the nature of the data.

Summary

In summary, DCM is the most general and direct approach to identifying the effective connectivity among the states of MIMO systems. The identification of DCMs usually proceeds using Bayesian inversion to estimate the posterior mode or most likely parameters of the model given the data. The state equations can be arbitrarily complicated and non-linear, however, there will be an optimal level of model complexity that is supported by the data (and identified using Bayesian model selection). The simplest model is probably a bilinear approximation to causal influences among state variables. This serves to minimize the complexity of the model by parameterizing the model in terms of first-order coupling and its changes with input (the bilinear terms). In the next section, we deal with the situations in which the input is unknown. This precludes DCM with deterministic systems, because the likelihood of the responses cannot be computed unless we know what caused them.

Models for stochastic inputs – SEM and regression models

When the inputs are unknown, and the statistics of the outputs are considered to second order, one is effectively restricted to linear or first-order models of effective connectivity. Although it is possible to deal with discrete-time bilinear models, with white noise inputs, they have the same covariance structure as ARMA (autoregressive moving average) models of the same order (Priestley, 1988: 66). This means that to distinguish between linear and non-linear models, one would need to study moments higher than second order (cf. the third-order cumulants in bi-coherence analyses). Consequently, we will focus on linear models of effective connectivity, under white stationary inputs. There are two important classes of model here: structural equation models and ARMA models. Both are finite parameter linear models that are distinguished by their dependency on dynamics. In SEM, the interactions are assumed to be instantaneous, whereas in ARMA the dynamics are retained.

An SEM can be derived from any DCM by assuming the inputs vary slowly in relation to neuronal and

haemodynamics. This is appropriate for positron emission tomography (PET) experiments and possibly some epoch-related fMRI designs, but not for event-related designs in ERP or fMRI. Note that this assumption pertains to the inputs or experimental design, not to the time constants of the outputs. In principle, it would be possible to apply DCM to a PET study.

Consider a linear approximation to any DCM where we can observe the states precisely and there was only one state variable per region:

$$\begin{aligned}\dot{x} &= f(x, u) \\ &= Ax + u = (A^0 - 1)x + u \\ y &= g(x) = x\end{aligned}\tag{A2.21}$$

Here, we have discounted observation error but allow stochastic inputs $u \sim N(0, Q)$. To make the connection to the SEM more explicit, we have expanded the connectivity matrix into off-diagonal connections and a leading diagonal matrix, modelling unit decay $A = A^0 - 1$. For simplicity, we have absorbed C into the covariance structure of the inputs Q . As the inputs are changing slowly relative to the dynamics, the change in states will be zero at the point of observation and we obtain the regression model used by SEM:

$$\begin{aligned}\dot{x} &= 0 \Rightarrow \\ (1 - A^0)x &= u \\ x &= (1 - A^0)^{-1}u\end{aligned}\tag{A2.22}$$

(see Chapter 38). The more conventional motivation for Eqn. A2.22 is to start with an instantaneous regression equation $x = A^0x + u$ that is formally identical to the second line above. Although this regression model obscures the connection with dynamic formulations, it is important to consider because it is the basis of commonly employed methods for estimating effective connectivity in neuroimaging to data. These are simple regression models and SEM.

Simple regression models

$x = A^0x + u$ can be treated as a general linear model by focusing on one region at a time, for example the first, to give (cf. Eqn. 38.11 in Chapter 38):

$$x_1 = [x_2, \dots, x_n] \begin{bmatrix} A_{12} \\ \vdots \\ A_{1n} \end{bmatrix} + u_1\tag{A2.23}$$

The elements of A can then be solved in a least squares sense by minimizing the norm of the unknown stochastic inputs u for that region (i.e. minimizing the unexplained variance of the target region given the states of

the remainder). This approach was proposed in Friston (1995) and has the advantage of providing precise estimates of connectivity with high degrees of freedom. However, these maximum likelihood estimators assume, rather implausibly, that the inputs are orthogonal to the states and, more importantly, do not ensure the inputs to different regions conform to the known covariance Q . Furthermore, there is no particular reason that the input variance should be minimized just because it is unknown. Structural equation modelling overcomes these limitations at the cost of degrees of freedom for efficient estimation

Structural equation modelling

In SEM, estimates of A^0 minimize the difference (KL divergence) between the observed covariance among the [observable] states and that implied by the model and assumptions about the inputs.

$$\begin{aligned} \langle xx^T \rangle &= \langle (1 - A^0)^{-1} uu^T (1 - A^0)^{-1T} \rangle \\ &= (1 - A^0)^{-1} Q (1 - A^0)^{-1T} \end{aligned} \quad \text{A2.24}$$

This is critical because the connectivity estimates implicitly minimize the discrepancy between the observed and implied covariances among the states induced by stochastic inputs. This is in contradistinction to the instantaneous regression approach (above) or ARMA analyses (below) in which the estimates simply minimize unexplained variance on a region-by-region basis. It should be noted that SEM can be extended to embrace dynamics by temporal embedding. However, these models then become formally the same as autoregressive-moving average models, which are considered below. Estimation of the effective connectivity in SEM, in the context of designed experiments (i.e. in neuroimaging) is rather poorly motivated. This is because one throws away all the information about the temporal pattern of designed inputs and uses only $Q = \langle uu^T \rangle$. In many applications of SEM, the inputs are discarded and Q is assumed to be a leading diagonal or identity matrix.

Quasi-bilinear models – psychophysiological interaction and moderator variables

There is a useful extension to the regression model implicit in Eqn. A2.22 that includes bilinear terms formed from known inputs that are distinct from stochastic inputs inducing [co]variance in the states. Let these known inputs be denoted by v . These usually represent some manipulated experimental context, such as cognitive set (e.g. attention) or time. These deterministic inputs

are also known as moderator variables in SEM. The underlying quasi-bilinear DCM, for one such input, is:

$$\dot{x} = (A^0 - 1)x + Bvx + u \quad \text{A2.25}$$

Again, assuming the system has settled at the point of observation:

$$\begin{aligned} \dot{x} &= 0 \\ (1 - A^0 - Bv)x &= u \\ x &= A^0 x + Bvx + u \end{aligned} \quad \text{A2.26}$$

This regression equation can be used to form least squares estimates as in Eqn. A2.23, in which case the additional bilinear regressors vx are known as *psychophysiological interaction* (PPI) terms (for obvious reasons). The corresponding SEM or path analysis usually proceeds by creating extra ‘virtual’ regions whose dynamics correspond to the bilinear terms. This is motivated by rewriting the last expression in Eqn. A2.26 as:

$$\begin{bmatrix} x \\ vx \end{bmatrix} = \begin{bmatrix} A^0 & B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ vx \end{bmatrix} + \begin{bmatrix} u \\ 0 \end{bmatrix} \quad \text{A2.27}$$

It is important to note that psychophysiological interactions and moderator variables in SEM are exactly the same thing and both speak of the importance of bilinear terms in causal models. Their relative success in the neuroimaging literature is probably due to the fact that they model changes in effective connectivity that are generally much more interesting than the connection strengths *per se*. Examples are changes induced by attentional modulation, changes during procedural learning and changes mediated pharmacologically. In other words, bilinear components afford ways of characterizing *plasticity* and, as such, play a key role in methods for functional integration. It is for this reason we focused on bilinear approximations as a minimal DCM in the previous section.

Summary

In summary, SEM is a simple and pragmatic approach to effective connectivity when dynamical aspects can be discounted, a linear model is sufficient, the state variables can be measured precisely and the input is unknown but stochastic and stationary. These assumptions are imposed by ignorance about the inputs. Some of these represent rather severe restrictions that limit the utility of SEM in relation to DCM or state-space models considered next. The most profound criticism of linear regression and SEM in imaging neuroscience is that they are models for interacting brain systems in the context of unknown

input. The whole point of designed experiments is that the inputs are known and under experimental control. This renders the utility of SEM for designed experiments somewhat questionable.

MULTIVARIATE ARMA MODELS

ARMA (autoregressive-moving average) models can be represented as *state-space* (or Markov) models that provide a compact description of any finite parameter linear model. From this state-space representation, multivariate autoregressive (MAR) models can be derived and estimated using a variety of well-established techniques (see Chapter 40). We will focus on how the state-space representation of linear models can be derived from the dynamic formulation and the assumptions required in this derivation. Many treatments of dynamic systems consider the dynamic formulation in terms of a state-equation, a continuous state-space model (SSM). We preserve the distinction because there is an important asymmetry in the sense that one can always derive a discrete SSM from a DCM. However, there is no necessary mapping from an SSM to a DCM. This is critical for causal inference because only DCMs are causal in the control theory sense (see below).

Assume a linear DCM in which inputs comprise stationary white Weiner processes $u \sim N(0, Q)$ that are offered to each region in equal strength (i.e. $C = 1$). This renders Eqn. A2.3 a linear stochastic differential equation (SDE):

$$\begin{aligned}\dot{x} &= Ax + u \\ y &= Lx\end{aligned}\tag{A2.28}$$

The value of x at some future lag comprises a deterministic and a stochastic component η that obtains by regarding the effects of the input as an accumulation of local linear perturbations:

$$\begin{aligned}x(t + \tau) &= e^{\tau A} x(t) + \eta \\ \eta &= \int_0^\tau e^{\sigma A} u(t + \sigma) d\sigma\end{aligned}\tag{A2.29}$$

Notice that the stochastic part can be regarded as convolving the random state-fluctuations with the system's first-order kernel. Using the assumption that the input is uncorrelated, the covariance of the stochastic part is:

$$W = \langle \eta \eta^T \rangle = \int_0^\tau e^{\sigma A} Q e^{\sigma A^T} d\sigma\tag{A2.30}$$

It can be seen that when the lag is small in relation to the Lyapunov exponents, $\text{eig}(A)$ we get $e^{\sigma A} \approx 1$ and $W \approx Q\tau$. By incorporating the output transformation and observation error, we can augment this model to furnish a state-space model with system matrix $F = e^{\tau A}$, input matrix $G = \sqrt{W}$ and observation matrix L :

$$\begin{aligned}x_t &= Fx_{t-1} + Gz_t \\ y_t &= Lx_t + \varepsilon_t\end{aligned}\tag{A2.31}$$

where z is an innovation representing dynamically transformed stochastic input. If we knew L and were interested in inferring on the hidden states, we would normally turn to Bayesian filtering (e.g. Kalman filtering) as described in Appendix 5. However, we will assume that we are more interested in inference about the coupling implied by the system matrix. In this case, we can reformulate the state-space model and treat it as an ARMA model.

Critically, every state-space model has an ARMA representation and vice versa. For example, if $L = 1$, we can eliminate the hidden states to give:

$$y_t - Fy_{t-1} = Gz_t + \varepsilon_t - F\varepsilon_{t-1}\tag{A2.32}$$

This is simply an ARMA(1,2) model that can be inverted using the usual procedures (see Chapter 40). The autoregressive part is on the left and the moving average of the innovations is on the right. Critically, Eqn. A2.32 formulates the dynamics in terms of, and only of, the response variable and random terms. Although it is always possible to derive an ARMA representation from a DCM (through the state-space representation), the reverse mapping is not necessarily defined. Having said this, ARMA models can be useful in establishing the presence of coupling even if the exact form of the coupling is not specified (cf. Volterra characterizations).

In summary, discrete-time linear models of effective connectivity can be reduced to multivariate AR (or, more generally ARMA) models, whose coefficients can be estimated given only the states (or outputs) by assuming the inputs are white and Gaussian. They therefore operate under similar assumptions as SEM but are time-series models.

A note on causality

There are many schemes for inverting state-space models of the sort in Eqn. A2.31. Inference on the system matrix could be considered in the light of functional connectivity, however, the regression coefficients are not really measures of effective connectivity. This is because there is no necessary mapping to the parameters of a DCM. In other words, although one can always map from the parameters of a causal model to its state-space

representation $F \leftarrow e^{\tau A}$, the inverse mapping does not necessarily exist (a simple intuition here is that the log of a negative number is not real).

An interesting aspect of inference on the system matrix (i.e. regression coefficients) is the use of model comparison to compare the regression of one channel on another. Because these coefficients encode statistical dependence at different temporal lags, this model comparison is often framed in causal terms, by appeal to temporal precedence (e.g. Granger causality). However, for many, this rhetoric represents a category error because the regression coefficients cannot have the attribute 'causal'. This is because causal is an attribute of the state equation that implies $A = \partial f / \partial x$ is real. This Jacobian is not defined in the state-space or ARMA representations because the mapping $\frac{1}{\tau} \ln(F) \rightarrow \partial f / \partial x$ does not necessarily exist. It is quite possible to infer Granger causality that is acausal when a cause is observed after its effect. fMRI presents a good example of an acausal system, because of the delay imposed on the expression of neuronal dynamics (which are causal) at the level of haemodynamics (which are not). For example, one region, with a long haemodynamic latency, could cause a neuronal response in another that was expressed, haemodynamically, before the source. This example demonstrates that one cannot estimate effective connectivity or coupling using just the outputs of a system (e.g. observed fMRI responses).

CONCLUSION

We have reviewed a series of models, all of which can be formulated as special cases of DCMs. Two fundamental

distinctions organize these models. The first is whether they model the coupling of inputs to outputs or whether they model interactions among hidden states. The second distinction (see Figure A2.2) is that between models that require the inputs to be known, as in designed experiments and those where the input is not under experimental control but can be assumed to be well behaved. With only information about the density of the inputs (or the joint density of the inputs and outputs) the models of connectivity that can be used are limited; unless one uses moments greater than second-order, only linear models can be estimated.

Many methods for non-linear system identification and causal modelling have been developed in situations where the system input is not under experimental control and, in the case of SEM, for static data. Volterra kernels and DCMs may be especially useful in neuroimaging because we deal explicitly with time-series data generated by designed experiments.

REFERENCES

- Bendat JS (1990) *Nonlinear system analysis and identification from random data*. John Wiley and Sons, New York
- Fliess M, Lamnabhi M, Lamnabhi-Lagarigue F (1983) An algebraic approach to nonlinear functional expansions. *IEEE Trans Circ Sys* **30**: 554–70
- Friston KJ (1995) Functional and effective connectivity in neuroimaging: a synthesis. *Hum Brain Mapp* **2**: 56–78
- Friston KJ (2002) Bayesian estimation of dynamical systems: an application to fMRI. *NeuroImage* **16**: 465–83
- Friston KJ, Büchel C (2000) Attentional modulation of V5 in humans. *Proc Natl Acad Sci USA* **97**: 7591–96
- Priestley MB (1988) *Non-linear and non-stationary time series analysis*. Academic Press, London